Terminology in the IDPro Body of Knowledge

Heather Flanagan, editor - ©2023 IDPro

Editor's Note: *This is a consolidated list of the terminology highlighted in each of the articles published in the Body of Knowledge (BoK). It is not, by any means, a definitive or even broadly supported set of definitions; the context an individual reader brings to the table will influence how accurate the terminology for their use case. We offer the consolidated list here as a touchpoint for discussion. Article authors are encouraged to review and use existing definitions before offering new ones for terms already described in the BoK.*

*You are encouraged to also read the article, "*Words of Identity*" by Espen Bago, for a cautious view of how, despite efforts like this Terminology document, words in the IAM space are often ambiguous.*

Please consider offering feedback to the articles that use these terms via the IDPro GitHub repository: https://github.com/IDPros/bok.

Term	**Definition**	**Source**
Abstraction	the practice of identifying and isolating repeated aspects of operations or business logic so that they can be maintained in one place and referenced in many places.	Introduction to Policy-Based Access Controls (v2)
Access Certification	Certification is the ongoing review of who has which accesses (i.e., the business process to verify that access rights are correct).	Introduction to Identity - Part 1: Admin-time (v2), Techniques To Approach Least Privilege
Access Control	Controlling who can have access to data, systems, services, resources, locations. The 'Who' can be a user, a device or thing, a service	Introduction to Access Control

Access Control	Various methods to limit access to data, systems, services, resources, locations by a user, a device or thing, or a service.	IAM Reference Architecture
Access Control Lists	Access Control Lists are definitions around who or what are allowed or denied access to a resource. For example, a file share may have an Access Control List that allows Marketing Department users to read and write, IT Department users to read-only, and denies all other users' access.	Authentication and Authorization
Access Control System	a structure that manages and helps enforce decisions about access within an organization.	Introduction to Policy-Based Access Controls (v2)
Access Governance	The assurance that all access has been given based on the correct decision criteria and parameters	Introduction to Access Control
Access Governance	Access Governance provides oversight and control over access rights implemented in multiple local or shared authorization systems. These rights may be controlled in a variety of ways, starting with the existence and validity of the digital identity. Other controls include various mechanisms such as policies, the mapping of roles, permissions, and identities. The abbreviation used is for Identity Governance and Administration and is commonly used in the commercial sector. This roughly corresponds to the Access Certification section of the first-class component Governance Systems in the FICAM model. IGA is not specifically addressed in the ISO/IEC model.	IAM Reference Architecture

Access Management	Use of identity information to provide access control to protected resources such as computer systems, databases, or physical spaces.	Introduction to IAM Architecture
Access Management	The process and techniques used to control access to resources. This capability works together with identity management and the Relying Party to achieve this goal. The model shows access management as a conceptual grouping consisting of the Access Governance function and the shared authorization component. However, access management impacts local authorization as well (through the governance function).	IAM Reference Architecture
Access Policy	Definition of the rules to allow or disallow access to secured objects.	Introduction to Access Control
Access Requester	The person, process, system, or thing that seeks to access a protected resource.	Introduction to Access Control
Access Supplier	The component granting access to data, systems, services after the access policy requirements (set in the Policy Administration Point) have been met by the Access Requester.	Introduction to Access Control
Access Token	The OAuth2 token that allows a client to get access to a protected resource	An Introduction to OAuth2.0
Account Owner	An entity that “owns” or claims responsibility for an account. Generally, an account is issued in the name of the owner(s) or their delegate(s) in the case of enterprises.	Account Recovery (v2)
Account Recovery	The process of returning account access to an account owner when they lose, forget, or cannot otherwise produce the account’s nominal	Account Recovery (v2)

	credentials. This may be accomplished in person, remote, or in a hybrid format.	
Account Recovery	The process of updating a user's credentials within a scenario where the user cannot validate those credentials	Managing Identity in Customer Service Operations
Account Takeover	Account takeover is a form of identity theft and fraud, where a malicious third party successfully gains access to a user's account credentials.	Account Recovery (v2), Designing MFA for Humans, Techniques To Approach Least Privilege
Accountability	The obligation of a person to accept the results of one's actions, be they positive or negative. This person is probably also a species of an owner.	Introduction to Access Control
Action	a protected operation available for a resource, such as "view", "edit", or "submit".	Introduction to Policy-Based Access Controls (v2)
Adaptive Authentication	Adaptive authentication aims to determine and enforce the authentication level required at any time during a user session - when the session is commenced, during the session when access requirements force a re-evaluation, or when the session token expires. The factors to be used in achieving that authentication level are determined dynamically based on the access control policy governing the resources being accessed, and a variety of environmental conditions and risk factors in effect at that time for that user.	Designing MFA for Humans

Agent (also "Customer Service Agent")	The person responsible for communicating with and solving problems on behalf of customers or end-users.	Account Recovery (v2), Managing Identity in Customer Service Operations
Agile Project Management	A framework that uses a continuous, iterative process to deliver a defined piece of functionality, typically a component of a product or service. Scrum is a popular framework (https://www.scrumalliance.org/about-scrum/overview)	Introduction to IAM Project Management
Alignment	the synchronization rate of processes and environments	Strategic Alignment and Access Governance
Applicant	A subject undergoing the processes of enrollment and identity proofing.	Defining the Problem – Identity Proofing Challenges
Architecture	Framework for the design, deployment, and operation of an information technology infrastructure. It provides a structure whereby an organization can standardize the technology it uses and align its IT infrastructure with digital transformation policy, IT development plans, and business goals.	Introduction to IAM Architecture
Architecture Overview	Describes the architecture components required for supporting IAM across the enterprise.	Introduction to IAM Architecture
Architecture Patterns	Identifies the essential patterns that categorize the IT infrastructure architecture in an organization and will guide the deployment choices for IAM solutions.	Introduction to IAM Architecture
Assertion	A formal message or token that conveys information about a principal, typically including a level of assurance about an authentication event and sometimes additional attribute	IAM Reference Architecture

	information. Sometimes this is called a Security Token.	
Assurance Level	A category describing the strength of the identity proofing process and/or the authentication process. See NIST SP.800-63-3 for further information.	IAM Reference Architecture
Asymmetric Cryptography	Any cryptographic algorithm which depends on pairs of keys for encryption and decryption. The entity that generates the keys shares one (see Public Key) and holds and protects the other (see Private Key). They are referred to as asymmetric because one key encrypts, and the other decrypts.	Practical Implications of Public Key Infrastructure for Identity Professionals
Attribute Provider	Sometimes the authority for attributes is distinguished from the authority for identities. In this case, the term Attribute Provider is sometimes used. It is a subset or type of an Identity Information Authority.	IAM Reference Architecture
Attribute-Based Access Control ("ABAC") / Claims-Based Access Control ("CBAC")	a pattern of access control system involving dynamic definitions of permissions based on information ("attributes", or "claims"), such as job code, department, or group membership.	Introduction to Policy-Based Access Controls (v2), The Business Case for IAM
Attributes	Key/value pairs relevant for the digital identity (username, first name, last name, etc.).	An Overview of the Digital Identity Lifecycle (v2)
Audit Repository	A component that stores records about all sorts of events that may be useful later to determine if operations are according to policy, support forensic investigations, and allow for pattern analysis. Typically, this is highly controlled to prevent tampering. Audit Repository is the ISO	IAM Reference Architecture

	name for this concept and is localized to the IDM. In this model, the term is generalized to indicate a service that supports event records from any part of the ecosystem.	
Authentication	Authentication is the process of proving that the user with a digital identity who is requesting access is the rightful owner of that identity. Depending on the use-case, an 'identity' may represent a human or a non-human entity; may be either individual or organizational; and may be verified in the real world to a varying degree, including not at all.	Introduction to Access Control (v3), Authentication and Authorization, Introduction to Consumer Identity and Access Management
Authentication (AuthN)	The act of determining that to a level of assurance, the principal/subject is authentic.	IAM Reference Architecture
Authenticator	The means used to confirm the identity of a user, processor, or device, such as a username and password, a one-time pin, or a smart card.	Identity and Access Management Workforce Planning, Introduction to Customer Identity and Access Management
AuthN Assertion	A security token whereby the IDP provides identity and authentication information securely to the RP.	IAM Reference Architecture
Authoritative Source	The system of record (SOR) for identity data; an organization may have more than one authoritative source of data in their environment.	User Provisioning in the Enterprise, Introduction to Customer Identity and Access Management
Authorization	Determining a user's rights to access functionality or resources within a computer application and the level at which that access should be granted. In most cases, an 'authority' defines and grants access, but in some cases, access is granted because of inherent	Introduction to Access Control, Authentication and Authorization, Introduction to Customer Identity and Access Management

	rights (like patient access to their own medical data)	
Authorization (AuthZ)	Authorization is how a decision is made at run-time to allow access to a resource. We break this down into two types: shared and local. The FICAM framework includes this as a subcomponent of the Access Management System. AuthZ is not included in the ISO or Internet2 models.	IAM Reference Architecture
Authorization Server (AS)	The OAuth2 server is able to authorize a client, issue tokens, and potentially validate tokens	An Introduction to OAuth2.0
Automatic Certificate Management Environment (ACME)	A communication protocol for automating lifecycle management of PKI certificates. Significant providers like Let's Encrypt leverage ACME to support issuing TLS certificates for web servers.	Practical Implications of Public Key Infrastructure for Identity Professionals
Bearer Token	A token whose possession is sufficient to enable access to a protected resource	An Introduction to OAuth2.0
Bilateral Federation	A bilateral federation is one that consists of only two entities: one Identity Provider (IdP) and one Service Provider (SP). This is the most common model for an enterprise identity federation.	Federation Simplified (v2)
Binding	Associating an authenticator with an identity.	Identity and Access Management Workforce Planning, Defining the Problem – Identity Proofing Challenges
Bot	Sometimes called an Internet bot, short for 'robot' but referring to a software routine that performs	Non-Human Account Management (v2)

	automated tasks over the Internet or a web robot referring to an autonomous network application, or simply a ‘bot’ referring to an automated, typically repetitive, task used for a specific purpose.	
Business to Business (B2B)	Business to Business processes in the field of IAM involve business partner access to company resources using some form of remote access (e.g., federated access).	The Business Case for IAM
Business to Consumer (B2C)	Business to Consumer processes in the field of IAM are customer or consumer access to company resources. In B2C, consumers manage their own identity in a CIAM. The company still manages access to the resources, using ABAC or PBAC methods for access control	The Business Case for IAM
Business to Employee (B2E)	Business to Employee, also called workforce IAM, includes managing identities and accounts for employees and contractors following an identity lifecycle.	The Business Case for IAM
Ceremonies	Predictable interactions that users can infrequently navigate in a well-watched place	Introduction to Identity – Part 2: Access Management
Certificate Authority Trust List (CTL)	A client maintains a list of trusted Certificate Authorities created and managed by the software provider or local administrators. The client will only trust certificates issued under one of the CAs in the CTL, so the CTL serves as a "safe list."	Practical Implications of Public Key Infrastructure for Identity Professionals
Certificate Management System (CMS)	A system that provides management and reporting layers for certificate issuance and revocation. A CMS integrates CA products with Identity	Practical Implications of Public Key Infrastructure for Identity Professionals

	Governance and Administration (IGA) systems as well as Service Desk systems.	
Certificate Policy (CP)	A document that defines the high-level policy requirement for a PKI. RFC 3647 identifies a PKI's policy framework and describes a CP's contents and outline. An enterprise operating a CA will often publish its certificate policy to external parties so they can determine whether to trust certificates issued by the CA.	Practical Implications of Public Key Infrastructure for Identity Professionals
Certificate Practices Statement (CPS)	A CP identifies the requirements for managing a CA and issuing PKI certificates. A CPS describes how a CA implements those requirements. The CPS uses the same outline as the CP, defined in RFC 3647. Unlike the CP, enterprises rarely publish their CPS in unredacted form.	Practical Implications of Public Key Infrastructure for Identity Professionals
Certificate Revocation List (CRL)	A certificate authority will publish a list of revoked certificates, called a CRL so that clients can verify that a certificate is still good.	Practical Implications of Public Key Infrastructure for Identity Professionals
Certificate Signing Request (CSR)	When requesting a certificate, the requesting entity provides a copy of the public key, their identifiers, and other information in a specially formatted binary object called a CSR.	Practical Implications of Public Key Infrastructure for Identity Professionals
Channel	The communication avenue between you and your end-user, or your agent and their customer. This could be phone, chat, social media, or others.	Managing Identity in Customer Service Operations
CIA Triad	The fundamental Information security concepts of risk classification of resources from the perspectives of Confidentiality, Integrity, and Availability.	Non-Human Account Management (v2)

Claimant	A subject whose identity is to be verified by using one or more authentication protocols.	Defining the Problem – Identity Proofing Challenges
Claimed Identity	An applicant's declaration of unvalidated and unverified personal attributes.	Defining the Problem – Identity Proofing Challenges
Claims-Based Access Control (CBAC)	See Attribute-Based Access Control (ABAC)	Introduction to Policy-Based Access Controls (v2)
Classical Computer	A computer that uses binary encoding and Boolean logic to make calculations in a deterministic way. We use the term Classical Computers in contrast with Quantum Computers.	Practical Implications of Public Key Infrastructure for Identity Professionals
Client	A client application consuming an API	An Introduction to OAuth2.0
Cloud Infrastructure Entitlement Management (CIEM)	a categorization of technologies focused on managing the granting, verification, and refinement of permissions for cloud and hybrid technologies. CIEM is often seen as a component of Identity Governance and Administration (IGA)	Techniques To Approach Least Privilege
Competency Model	A collection of tasks, knowledge, and skills (TKS) needed for effective job performance. A competency model is part of a workforce framework.	Identity and Access Management Workforce Planning
Consent	Permission for something to happen or agreement to do something.	Introduction to Privacy and Compliance for Consumers, Introduction to Customer Identity and Access Management
Consumer (or Customer) Identity and Access	CIAM is the field of IAM that focuses on the Registration, Authentication, and Authorization services for an individual or entity receiving or	The Business Case for IAM, Introduction to Customer Identity and Access Management

Management (CIAM)	purchasing services from an organization.	
Consumer Protection Law	Laws and regulations that are designed to protect the rights of individual consumers and to stop unfair, deceptive, and fraudulent business practices.	Laws Governing Identity Systems
Context	conditions under which an action on a resource is authorized for a subject, such as time of access, location of access, or a compliance state.	Introduction to Policy-Based Access Controls (v2)
Continuous Authentication	Continuous authentication is a mechanism that uses a variety of signals and measurements to determine during a user session if there is any change in the confidence that it is still the same user that authenticated at the beginning of the session, and trigger an authentication action if there is a drop in confidence.	Designing MFA for Humans
Contract Law	Laws that relate to making and enforcing agreements between or among separate parties.	Laws Governing Identity Systems
Credential	A credential allows for authentication of an entity by binding an identity to an authenticator.	IAM Reference Architecture
Credential	An object or data structure that authoritatively binds an identity—via an identifier or identifiers—and (optionally) additional attributes to at least one authenticator possessed and controlled by a subscriber.	Defining the Problem – Identity Proofing Challenges
Credential Management	How to issue, manage, and revoke authenticators bound to identities. Credential Management roughly corresponds to the IDPro term for Credential Services; we use the term	Identity and Access Management Workforce Planning

	Credential Management here to correlate to the Federal Identity, Credential, and Access Management (FICAM) initiative’s terms.	
Credential Service Provider	Following the guidance included in NIST 800-63-3, we include both the enrollment function and credential services together under the name Credential Services Provider.	IAM Reference Architecture
Credential Service Provider	A trusted entity that issues or registers subscriber authenticators and issues electronic credentials to subscribers. A CSP may be an independent third party or may issue credentials for its own use.	Defining the Problem – Identity Proofing Challenges
Credential Services	Credential Services issue or register the subscriber authenticators, deliver the credential for use, and subsequently manage the credentials. We include PKI information for IAM architectures that must include system components that need certificates and private keys. This roughly corresponds to the FICAM component called Credential Management Systems.	IAM Reference Architecture
Credential Stuffing	An attack in which an adversary tests lists of username and password pairs against a given CIAM system.	Introduction to Customer Identity and Access Management
Credentials	Any attribute or shared secret that can be used to authenticate a user.	Account Recovery (v2)
Credentials	In the context of CIAM, credentials are how individuals authenticate themselves to an organization’s CIAM system	Introduction to Customer Identity and Access Management
Cryptographic Module	A hardware or software component that securely performs cryptographic operations within a logical boundary.	Practical Implications of Public Key Infrastructure for Identity Professionals

	Cryptographic Modules store private keys within this boundary and use them for cryptographic functions at the request of an authorized user or process.	
Cryptographic Module Validation Program (CMVP)	A program allowing cryptographic module developers to test their modules against the requirements defined in FIPS-140. The computer security resource center under the United States National Institute of Standards and Technology (NIST) maintains a publicly available list of validated modules.	Practical Implications of Public Key Infrastructure for Identity Professionals
Data Controller	Defined in Article 4(7) of the GDPR: "'controller' means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data;". This article uses the term "organisation" as a synonym for "data controller", since organisations involved in IAM will normally be data controllers.	An Introduction to the GDPR
Data Mapping	"a system of cataloguing what data you collect, how it's used, where it's stored, and how it travels throughout your organization and beyond."	Impact of GDPR on Identity and Access Management
Data Processor	Defined in Article 4(8) of the GDPR for situations where an organisation processes personal data solely on the instructions of others. A Data Processor must not determine the purposes of processing, for example by processing in its own interests, or, beyond limited technical choices, the means of doing so. Data Processors are regulated by Article 28: in	An Introduction to the GDPR

	particular they must have a contract with the Data Controller that covers all the subjects listed in Article 28(3). Data Processors are excluded from some, but not all, of the liabilities and duties of Data Controllers.	
Data Protection by Design	Data protection through technology design. See GDPR Article 25 for more detail	Impact of GDPR on Identity and Access Management
Data Protection Officer	An individual who must be appointed in any organization that processes any data defined by the GDPR as sensitive. The DPO is responsible for "Working towards the compliance with all relevant data protection laws, monitoring specific processes, such as data protection impact assessments, increasing employee awareness for data protection and training them accordingly, as well as collaborating with the supervisory authorities."(See GDPR Articles 35, 37, 38, and 39 for more detail)	Impact of GDPR on Identity and Access Management
Data Subject	Defined in Article 4(1) of the GDPR (see "Personal Data" above) as the formal term for the human to whom personal data relates. This article uses the term "individual" as a synonym for "data subject".	An Introduction to the GDPR
Decentralized Identifier (DID)	An identifier that is created and anchored in a decentralized system such as a blockchain or ledger and can represent any entity in the ecosystem – an issuer, a holder, a verifier, and even an identity hub.	A Peek into the Future of Decentralized Identity

Delegated Authorization Framework	An access control framework that decouples authentication from authorization, allowing the password to stay local and protected	Introduction to Identity – Part 2: Access Management
Digital Cards	Represent verifiable credentials that users collect over time and are stored as part of the user agent or the identity hub of the user. It's somewhat simpler to refer to them as digital cards rather than verifiable credentials when speaking about them.	A Peek into the Future of Decentralized Identity
Digital Identity	the combination of a unique identifier together with relevant attributes that uniquely identifies an entity..	An Overview of the Digital Identity Lifecycle (v2)
Digital Wallet	represents a digital metaphor for a physical wallet and is generally represented by the combination of the user agent and the underlying capabilities of the computing device, such as secure storage and secure enclaves on a mobile phone. The digital wallet contains digital cards.	A Peek into the Future of Decentralized Identity
Directory	A directory is a central repository for user identities and the attributes that make up those identities. A user identity might be John Smith with firstName attribute as John, lastName attribute as Smith, title attribute as Director, and Department attribute as Marketing. The attributes in the directory can be used to make authorization decisions about what this user should have access to in applications.	Authentication and Authorization
Discretionary Access Control	a pattern of access control system involving static, manual definitions of permissions assigned directly to users.	Introduction to Policy-Based Access Controls (v2)

dPKI	A decentralized public key infrastructure and is usually implemented via an immutable blockchain or ledger – a place where DIDs can be registered and looked up alongside the associated public keys of the DID and its metadata. dPKI can be described more generally as the *verifiable data registry*, as the dPKI is just one of many possible implementations for a verifiable data registry. While this paper refers to dPKI, the reader should be aware that a verifiable data registry need not necessarily be "decentralized".	A Peek into the Future of Decentralized Identity
Electronic Identification, Authentication, and Trust Services (eIDAS)	European legislation gives legal standing to electronic signatures under eIDAS. This legislation also documents providing legally binding digital signatures with X.509 certificates to comply with Qualified Signature requirements.	Practical Implications of Public Key Infrastructure for Identity Professionals
Electronic Identification, Authentication and Trust Services (eIDAS)	European legislation that gives legal standing to electronic signatures. This legislation also documents how to provide legally binding digital signatures with X.509 certificates to comply with Qualified Signature.	Practical Implications of Public Key Infrastructure for Identity Professionals
Elliptic Curve Cryptography (ECC)	An asymmetric cryptosystem based on calculating points along elliptic curves.	Practical Implications of Public Key Infrastructure for Identity Professionals
Encryption	Processing data using a cryptographic algorithm to provide confidentiality assurance.	Practical Implications of Public Key Infrastructure for Identity Professionals
Enforcement	The mechanism that ensures an individual cannot perform an action or	IAM Reference Architecture

	access a system when prohibited by policy.	
Enrollment	Also known as Registration. Enrollment is concerned with the proofing and lifecycle aspects of the principal (or subject). The entity that performs enrollment has sometimes been known as a Registration Authority, but we (following NIST SP.800-63-3) will use the term Credential Service Provider.	IAM Reference Architecture, Defining the Problem – Identity Proofing Challenges
Enterprise Architecture	An architecture covering all components of the information technology (IT) environment	Introduction to IAM Architecture
Entitlement	The artifact that allows access to a resource by a principal. This artifact is also known as a privilege, access right, permission, or an authorization. An entitlement can be implemented in a variety of ways.	IAM Reference Architecture
Entitlement Catalog	A database of entitlements and their related metadata. The catalog includes an index of entitlement data pulled from business systems, applications, and platforms, as well as technical and business descriptions of the entitlements or their use	User Provisioning in the Enterprise
Entitlement Management	Cataloging and managing all the accesses an account may have. This is the business process to provision access.	Introduction to Identity - Part 1: Admin-time (v2)
External identifier	The means by which a person in control of a digital identity refers to that identity when interacting with a system	Identifiers and Usernames

Federal Agency Smart Credential Number (FASC-N)	A unique identifier associated with a smart card. FASC-N is used in the US Federal Government PIV standard to support Physical Access.	Practical Implications of Public Key Infrastructure for Identity Professionals
Federal Information Processing Standard **(**"FIPS") 140	A NIST standard defining "Security Requirements for Cryptographic Modules.	Practical Implications of Public Key Infrastructure for Identity Professionals
Federated Access Controls	an access control architecture that accommodates separation of user/subject authority and resource/object authority.	Introduction to Policy-Based Access Controls (v2)
Federated Identity	The means of linking a person's electronic identity and attributes, stored across multiple distinct identity management systems	Introduction to Identity – Part 2: Access Management
Fractured Identity	A case where a single end-user has multiple disparate digital identities.	Managing Identity in Customer Service Operations
Fraud Law	Laws that protect against the intentional misrepresentation of information made by one person to another, with knowledge of its falsity and for the purpose of inducing the other person to act, and upon which the other person relies with resulting injury or damage.	Laws Governing Identity Systems
Gantt Chart	A popular schedule format that displays both activity and timeframes in a single chart	Introduction to Project Management for IAM Projects
General Data Protection Act (GDPR)	Formally, Regulation 2016/679 of the European Union, in force May 25, 2018. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679	An Introduction to the GDPR

Governance	Making sure that accountable owners are demonstrably in control.	Strategic Alignment and Access Governance
Groups	A set of identities with defined permissions. In this specific context, a group contains many individuals, but the group identity is opaque, and no information is available regarding which group member took an individual action.	Practical Implications of Public Key Infrastructure for Identity Professionals
Hardware Security Module (HSM)	A hardware device that generates and protects cryptographic keys.	Practical Implications of Public Key Infrastructure for Identity Professionals
Holder	The entity that holds verifiable credentials. Holders are typically users but can also be organizations or devices.	A Peek into the Future of Decentralized Identity
Identification	Uniquely establish a user of a system or application.	Introduction to Access Control, Introduction to Customer Identity and Access Management
Identifier	The way a system refers to a digital identity. PKI Certificates support both internal and external identifiers. See Ian Glazer's article, "Identifiers and Usernames," for a generic overview of identifiers.	Practical Implications of Public Key Infrastructure for Identity Professionals
Identifier	An identifier is a means by which a system refers to a record (at the most abstract levels.) In this case, it could mean the string that a person provides that "names" their use account.	Introduction to Customer Identity and Access Management
Identity	Defining attributes for a human user that may vary across domains, e.g., a user's digital identity will have a different definition in a work	Non-Human Account Management (v2)

	environment as opposed to the user’s bank. A device identifier is sometimes referred to as its identity.	
Identity	An attribute or set of attributes that uniquely describes a subject within a given context.	Defining the Problem – Identity Proofing Challenges
Identity Analytics and Intelligence (IdA)	Identity analytics and intelligence mean looking at entitlement data, looking at the assignment of that, and trying to figure out and define what risk looks like. IdA provides a risk-based approach for managing system identities and access, with the intention of centralizing governance, visibility, and reporting for access-based risk.	Introduction to Identity - Part 1: Admin-time (v2)
Identity and Access Management (IAM)	Identity and Access Management (IAM) is the discipline used to ensure the correct access is defined for the correct users to the correct resources for the correct reasons.	Authentication and Authorization
Identity and Access Management (IAM)	The discipline that enables the right individuals to access the right resources at the right times for the right reasons.	Identity and Access Management Workforce Planning
Identity and Access Management Workforce Planning	Activities involved in ensuring an enterprise identity and access management team are staffed with the right talent to execute business and technical objectives.	Identity and Access Management Workforce Planning
Identity, Credential, and Access Management (ICAM)	Programs, processes, technologies, and personnel used to create trusted digital identity representations of individuals and non-person entities, bind those identities to credentials that may serve as a proxy in access transactions, and leverage the	Identity and Access Management Workforce Planning

	credentials to provide authorized access to an organization’s resources.	
Identity Evidence	Information or documentation the applicant provides to support the claimed identity. Identity evidence may be physical (e.g., a driver’s license) or digital (e.g., an assertion generated and issued by a CSP based on the applicant successfully authenticating to the CSP).	Defining the Problem – Identity Proofing Challenges
Identity Federation	An identity federation is a group of computing or network providers that agree to operate using standard protocols and trust agreements. In a Single Sign-On (SSO) scenario, identity federation occurs when an Identity Provider (IdP) and Service Provider (SP) agree to communicate via a specific, standard protocol. The enterprise user will log into the application using their credentials from the enterprise rather than creating new, specific credentials within the application. By using one set of credentials, users need to manage only one credential, credential issues (such as password resets) can be managed in one location, and applications can rely on the appropriate enterprise systems (such as the HR system) to be the source of truth for a user’s status and affiliation. Identity federations can take several forms. In academia, multilateral federations, where a trusted third party manages the metadata of multiple IdPs and SPs, are fairly common. [1]This article focuses, however, on the enterprise use case where **bilateral federation**	Federation Simplified (v2)

	arrangements, where the agreements are one-to-one between an IdP and an SP, are the most common form of identity federation in use today.	
Identity Governance and Administration (IGA)	a discipline that focuses on identity life cycle management and access control from an administrative perspective.	Introduction to Identity - Part 1: Admin-time (v2), The Business Case for IAM
Identity Governance and Administration (IGA)	Includes the collection and use of identity information as well as the governance processes that ensure the right person has the right access to the right systems at the right time.	Introduction to IAM Architecture
Identity Governance and Administration (IGA)	a solution for automating user management and authorizations in target systems, building on the organization's customer and human resource processes.	Strategic Alignment and Access Governance
Identity Hub or Repository	The place where users can store their encrypted identity-related information. An identity hub can be anywhere – on the edge, on the cloud, or on your own server. Its purpose is to store personal data. Some implementations may allow other entities to access the identity hub of the user if the user specifically grants such access. You can think of an identity hub as the individual's personal data store.	A Peek into the Future of Decentralized Identity
Identity Information Authority (IIA)	This represents one or more data sources used by the IDM as the basis for the master set of principal/subject identity records. Each IIA may supply a subset of records and a subset of attributes. Sometimes the IIA is distinguished from the Identity Information Provider or IIP. We use IIA	IAM Reference Architecture

	to include the service that actually provides the information as well as the root authority. This corresponds to Identity Information Source in ISO/IEC 24760-2 and Identity Sources in Internet2.	
Identity Lifecycle Management	A process that detects changes in authoritative systems of record and updates identity records based on policies.	User Provisioning in the Enterprise
Identity Management (IDM)	A set of policies, procedures, technology, and other resources for maintaining identity information. The IDM contains information about principals/subjects, including credentials. It also includes other data such as metadata to enable interoperability with other components. The IDM is shown with a dotted line to indicate that it is a conceptual grouping of components, not a full-fledged system in itself.	IAM Reference Architecture
Identity Proofing	accruing evidence to support "who this is." Identity proofing is the last, but not the least, important part of this admin-time section. This is the process of collecting and verifying information about a person for the purpose of providing an account or a corresponding credential. This is typically performed before an account is created or the credential is issued, or a special privilege is granted.	Introduction to Identity - Part 1: Admin-time (v2)
Identity Proofing	The process by which a CSP collects, validates, and verifies information about a person.	Defining the Problem – Identity Proofing Challenges

Identity Provider (IdP)	An Identity Provider (IdP) performs a service that sends information about a user to an application. This information is typically held in a user store, so an identity provider will often take that information and transform it to be able to be passed to the service providers, AKA apps. The OASIS organization, which is responsible for the SAML specifications, defines an IdP as “A kind of SP that creates, maintains, and manages identity information for principals and provides principal authentication to other SPs within a federation, such as with web browser profiles.”	Federation Simplified (v2), Authentication and Authorization
Identity Provider (IDP)	Identity Provider or IDP is a common term. We treat this as a subset of Identity Management. It consists of the service interfaces: AuthN/Assertion, Service Provisioning Agent, Session Management, Discovery Services, and Metadata Management.	IAM Reference Architecture
Identity Provider (IdP)	The party that manages the subscriber’s primary authentication credentials and issues assertions derived from those credentials. This is commonly the CSP as discussed within this article.	Defining the Problem – Identity Proofing Challenges
Identity Register	This is the datastore that contains the enrolled entities and their attributes, including credentials. See the IDM section for elaboration. The terms Directory, Identity Repository, and Attribute Store are sometimes used as synonyms.	IAM Reference Architecture
Identity Repository	The identity repository is a directory or a database that can be referenced by	User Provisioning in the Enterprise

	external systems and services (such as authentication or authorization services).	
Identity Theft Law	Laws governing crimes in which the perpetrator gains access to sensitive personal information belonging to the victim (such as birth dates, passwords, email addresses, driver's license numbers, social security numbers, financial records, etc.), and then uses this information to impersonate the victim for personal gain, such as to commit fraud, establish credit in the victim’s name, or access the victim’s accounts.	Laws Governing Identity Systems
Impersonation	A scenario where a user is able to perform actions as though they are a known user other than themself.	Managing Identity in Customer Service Operations
Infrastructure-as-code	the process of managing and provisioning computer data centers through machine-readable definition files rather than physical hardware configuration or interactive configuration tools.	Techniques To Approach Least Privilege
Internet Key Exchange (IKE)	A subordinate standard under IPsec specifying how to use X.509 certificates to establish symmetric keys for an IPsec tunnel.certificates to establish symmetric keys for an IPsec tunnel.	Practical Implications of Public Key Infrastructure for Identity Professionals
Internet Protocol Security (IPsec)	A standard for communication between two machines providing confidentiality and integrity over the Internet Protocol.	Practical Implications of Public Key Infrastructure for Identity Professionals
Intra-organizational (Single Sign-On):	A central digital identity, such as an account in a directory, is linked by	An Overview of the Digital Identity Lifecycle (v2)

	downstream systems as authoritative for authentication.	
Inter-organizational (Federation)	An organization relies on another organization’s digital identity and lifecycle management processes.	An Overview of the Digital Identity Lifecycle (v2)
Internal identifier	The way an identity management system refers to a digital identity	Identifiers and Usernames
Issuer	The entity that issues verifiable credentials about subjects to holders. Issuers are typically a government entity or corporation, but an issuer can also be a person or device.	A Peek into the Future of Decentralized Identity
Joiner/Mover/Leaver	The joiner/mover/leaver lifecycle of an employee identity considers three stages in the life cycle: joining the organization, moving within the organization, and leaving the organization.	Introduction to Identity - Part 1: Admin-time (v2), The Business Case for IAM
Journey-based Creation	The process that guides a customer through a series of interactions prior to establishing a digital identity. For example, capturing the minimum basic information needed from a customer to enable creation of an identity.	An Overview of the Digital Identity Lifecycle (v2)
Just-in-time (JIT) Access	a technique where a credential or a permission is granted to a principal for a temporary timeframe when they need the permission to perform an activity. Access is revoked once the activity is complete, limiting its usage.	Techniques To Approach Least Privilege
Key	In a cryptosystem, a Key is a piece of information used to encrypt or decrypt data in a cryptographic algorithm.	Practical Implications of Public Key Infrastructure for Identity Professionals

Knowledge-Based Authentication (KBA)	A method of authentication that uses information known by both the end-user and the authentication service but is not necessarily a secret.	Account Recovery (v2), Managing Identity in Customer Service Operations
Knowledge-Based Authentication (KBA)	Identity-verification method based on knowledge of private information associated with the claimed identity. This is often referred to as knowledge-based verification (KBV) or knowledge-based proofing (KBP).	Defining the Problem – Identity Proofing Challenges
Least Privilege	Also known as the Principle of Least Privilege; a resource, such as a user, must only be able to access the resources (e.g., applications, data) that are necessary for it to function.	Introduction to Identity – Part 2: Access Management
Least Privilege	The principle that a security architecture should be designed so that each entity is granted the minimum system resources and authorizations that the entity needs to perform its function. (NIST Information Technology Laboratory)	Techniques To Approach Least Privilege
Lifecycle	In the context of CIAM, lifecycle refers to the stages that an individual or entity might experience over the course of their relationship with an organization, beginning with the formation of a relationship (such as being hired into an organization or signing up for service) and ending with the severance of that relationship (such as termination or closing an account)	Introduction to Customer Identity and Access Management
Local Authorization	Local authorization is handled by the RP.	IAM Reference Architecture
Metadata Management	The processes and techniques that allow the collection, use, and eventual	IAM Reference Architecture

	deletion of control data used by the IDM to recognize and trust the Relying Party. This corresponds to Relying Party data in the Internet2 model.	
MFA Prompt Bombing	Also known as MFA fatigue, MFA prompt bombing is a cyber-attack technique that describes when an attacker bombards a user with mobile-based push notifications, which sometimes leads to the user to approve the request out of annoyance which might lead to an account takeover.	Multi-factor Authentication
Multi-Factor Authentication (MFA)	An approach whereby a user's identity is validated to the trust level required according to a security policy for a resource being accessed using more than one factor (something you know (e.g., password), something you have (e.g., smartphone), something you are (e.g., fingerprint).	Account Recovery (v2), Introduction to Access Control
Multilateral Federation	A federation that consists of multiple entities that have agreed to a specific trust framework. There are several forms of multilateral federations, including hub-and-spoke and mesh. Multilateral federations are the most common model for academic identity federations.	Federation Simplified (v2)
National Institute of Standards and Technology (NIST)	): A US Government agency that defines and publishes various standards. One department within NIST, the Computer Security Resource Center (CSRC), publishes the Federal Information Processing Standards (FIPS) series. While these standards are only mandatory for US Government Agencies, many countries	Practical Implications of Public Key Infrastructure for Identity Professionals

	recognize them as de-facto global standards.	
Non-Human/Person Account	Any account not used by a person, such as accounts used for devices, services, and servers.	Non-Human Account Management (v2)
Non-Person Entities	Any unique combination of hardware and software firmware (e.g., device) that utilizes the capabilities of other programs, devices, or services to perform a function. Non-person entities may act independently or on behalf of an authenticated individual or another NPE.	Practical Implications of Public Key Infrastructure for Identity Professionals
OAuth 2.0	OAuth 2.0 is an open-source protocol that allows Resource Owners such as applications to share data with clients by facilitating communication with an Authorization Server. That data takes the form of credentials given to applications to obtain information/data from other applications. The Authorization Server is usually the Identity Provider (IdP). The Authorization Server (AS) may provide authorization directly or indirectly. For example, the AS may supply attributes or profile data of the Resource Owner or provide access to data that can later be used for authorization purposes, such as entitlements from an Identity Management or Governance Solution.	Federation Simplified (v2)
Online Certificate Status Protocol (OCSP)	A protocol that allows a client to query the Certificate Authority or a Validation Authority for the status of an individual certificate rather than downloading a CRL.	Practical Implications of Public Key Infrastructure for Identity Professionals

OpenID Connect (OIDC)	OpenID Connect is a simple identity layer on top of the OAuth 2.0 protocol. It enables Clients to verify the identity of the End-User based on the authentication performed by an Authorization Server, as well as to obtain basic profile information about the End-User in an interoperable and REST-like manner.	Federation Simplified (v2)
Passwordless	Any means of authenticating a user account that does not require a static stored shared secret. Techniques include one-time passwords and passkeys.	Introduction to Customer Identity and Access Management
Path Discovery and Validation (PDVal)	The process to determine whether a certificate is valid and trusted by the validator.	Practical Implications of Public Key Infrastructure for Identity Professionals
Permission	a statement of authorization for one or more subjects to perform one or more actions on one or more objects.	Introduction to Policy-Based Access Controls (v2)
Personal Data	Defined in Article 4(1) of the GDPR: "'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;". Note: "natural person" (human) is used to distinguish from companies and other corporate entities that are "legal persons".	An Introduction to the GDPR

Personal Data	Personal data are any information which are related to an identified or identifiable natural person.	Account Recovery (v2), Impact of GDPR on Identity and Access Management
Personal Identification Number (PIN)	A numeric secret commonly used to unlock a private key container in software or hardware.	Practical Implications of Public Key Infrastructure for Identity Professionals
Personal Identity Verification (PIV)	A US Government program designed to enable strong authentication for all government employees and contractors, based on Public Key Infrastructure.	Practical Implications of Public Key Infrastructure for Identity Professionals
Policy Administration Point (PAP)	The location where the different types of owners define the access policy.	Introduction to Access Control
Policy Decision Point (PDP)	The policy engine validating Access requests and provided attributed against the Access Policy (as defined in the Policy Administration Point).	Introduction to Access Control
Policy Enforcement Point (PEP)	The authority that will only let an Access Requester connect to the Access Supplier if the Policy Decision Point allows it.	Introduction to Access Control
Policy Engine	It is a security component that validates whether an actor is allowed to access a protected resource, following the requirements in an access policy.	Introduction to Access Control
Policy Information Point	The authority that refers to the (external) trusted providers of attributes that will be used in the Access Decision. An example is the myacclaim.com service that administers Open Badges of certifications, such as CISSP and MSCP.	Introduction to Access Control

Policy Store	A repository that houses configuration information for the CIAM system and serves as an Authoritative Source for that information. For example, OAuth token Lifecycle policies or Authorization policies.	Introduction to Customer Identity and Access Management
Policy-Based Access Control (PBAC)	a pattern of access control system involving dynamic definitions of access permissions based on user attributes (as in ABAC) and context variables for permitting or denying access.	Introduction to Policy-Based Access Controls (v2), The Business Case for IAM
Preferences	Choices that individuals or entities make in administering the relationship they have with an organization. These choices may include topics of interest or approved communication methods. Often, Preferences are stored with Profile information.	Introduction to Customer Identity and Access Management
Principle of Least Privilege	an information security best practice ensuring that users in an access control system do not have more access to resources than is necessary for their intended activities.	Introduction to Policy-Based Access Controls (v2)
Privacy	An abstract concept, with no single, common definition	Introduction to Privacy and Compliance for Consumers
Privacy Law	Laws that regulate the collection, use, storage, and transfer of personal data relating to identified or identifiable individuals.	Laws Governing Identity Systems
Private Key	A key that a single entity exclusively and privately controls. It corresponds to a public key that the entity may share for data encryption or signature verification.	Practical Implications of Public Key Infrastructure for Identity Professionals
Privileged Access Management	A mechanism for managing temporary access for accounts with high-risk permissions. PAM often involves	Techniques To Approach Least Privilege, The Business Case for IAM

	check-out and check-in of a credential generated for a single use.	
Privileged Account Management (PAM)	focusing on special control for risky high-level access. Privileged Account Management (PAM) is a mechanism for getting those special accounts under control.	Introduction to Identity - Part 1: Admin-time (v2)
Processing	Defined in Article 4(2) of the GDPR: “‘processing’ means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction”. Note that even this long list of activities is not exhaustive: other activities may also fall within the definition of “processing”. Additional rules, in Article 22, apply to “automated individual decision-making, including profiling”. These generally have the effect of strengthening the rights of information and objection described later and may limit the use of automation for some high-impact decisions.	An Introduction to the GDPR
Profile	A collection of attributes about an individual. The individual may provide it directly, or the organization may gather it indirectly.	Introduction to Customer Identity and Access Management
Progressive Profiling	A technique to reduce customer friction by gathering Profile, preference, and Consent information	Introduction to Customer Identity and Access Management

	over time (when needed) rather than all at once.	
Project	A time-limited activity to achieve a defined outcome(s)	Introduction to Project Management for IAM Projects
Project Charter	Documented authority for the project manager to proceed with a project; it will usually include a succinct statement of the project's purpose	Introduction to Project Management for IAM Projects
Project Plan	A document that describes a project; it will usually include a scope statement, schedule, resource plan, communications plan, and quality plan	Introduction to Project Management for IAM Projects
Protected Resource	An API in the OAuth2 terminology	An Introduction to OAuth2.0
Public Key	A key that an entity publicly distributes. It corresponds to a private key that the entity exclusively and privately controls.	Practical Implications of Public Key Infrastructure for Identity Professionals
Public Key Certificate	A certificate containing a public key, one or more identifiers for the private key holder, an identifier for the Certificate Authority, and additional metadata to support security requirements.	Practical Implications of Public Key Infrastructure for Identity Professionals
Public Key Infrastructure	A set of tools, standards, and related policies designed to manage trust based on public/private key pairs and certificates.	Practical Implications of Public Key Infrastructure for Identity Professionals
Protected Resource	A system, a process, a service, an information object, or even a physical location that is subject to access control as defined by the owner of the resource and by other stakeholders, such as a business process owner or Risk manager.	Introduction to Access Control

Reconciliation	The process of identifying and processing changes to users and user access made directly on target systems.	User Provisioning in the Enterprise
Refresh Token	The OAuth2 token that allows a client to renew an access token when it is expired without the user’s presence	An Introduction to OAuth2.0
Registration	See Enrollment	Defining the Problem – Identity Proofing Challenges
Registration	The creation of a relationship between an individual and an online system that is initiated by the individual and results in the creation of a user account or Profile.	Introduction to Customer Identity and Access Management
Registration Authority (RA)	An individual, system, or business function which provides registration and identity proofing for entities receiving certificates and manages the certificate issuance and renewal process. The most important responsibilities of an RA include identity proofing and binding the private key to the identity.	Practical Implications of Public Key Infrastructure for Identity Professionals
Relying Party (RP)	A component, system, or application that uses the IDP to identify its users. The RP has its own resources and logic. Note that the term ‘relying service’ is used in the ISO/IEC standards to encompass all types of components that use identity services, including systems, sub-systems, and applications, independent of the domain or operator. We will use the more common Relying Party (or RP). An RP roughly corresponds to the Agency Endpoint in the FICAM model or to Identity Consumers in the Internet2 model.	IAM Reference Architecture

Remote	*In the context of remote authentication or remote transaction*, an information exchange between network-connected devices where the information cannot be reliably protected end to end by a single organization's security controls.	Defining the Problem – Identity Proofing Challenges
Resource or Object	an asset protected by access controls, such as an application, system, or door.	Introduction to Identity - Part 1: Admin-time (v2)
Return on Investment (ROI)	Return on Investment is the economic measure of value of an investment, using costs, revenues, interest rates, and lifecycle as parameters.	The Business Case for IAM
Revoke	Revocation is the announcement that clients should no longer trust an individual certificate.	Practical Implications of Public Key Infrastructure for Identity Professionals
Revised Payment Systems Directive (PSD2)	PSD2 (the Revised Payment Services Directive, Directive (EU) 2015/2366) is an EU Directive, administered by the European Commission (Directorate General Internal Market) to regulate payment services and payment service providers throughout the European Union (EU) and European Economic Area (EEA). It contains many requirements specifically related to Strong Client Authentication.	Designing MFA for Humans
Risk Context (RCTX)	Risk Context consists of additional facts that can be brought to bear to improve the overall security of the ecosystem. Internal or external events and facts can be applied to enable, limit, or terminate access. This is similar to the section Monitors and Sensors under FICAM's Governance Systems and to many of the inputs of the Policy Decision Point in the NIST	IAM Reference Architecture

	Special Publication 800-207, a paper on Zero Trust.	
Role Management	a way to group access rules to make them more manageable	Introduction to Identity - Part 1: Admin-time (v2)
Role-Based Access Control (RBAC)	the use of roles at run-time; a way to govern who gets access to what through the use of roles.	Introduction to Identity - Part 1: Admin-time (v2)
Role-Based Access Control (RBAC)	A pattern of access control system involving sets of static, manual definitions of permissions assigned to “roles”, which can be consistently and repeatably associated with users with common access needs. Role-based access control is a control scheme in which roles are granted to identities, and those roles determine what access to resources those identities should have. Basic roles might be “admin” and “read-only user” – an admin would be able to make changes to a system and a read-only user would only be able to view resources.	Introduction to Policy-Based Access Controls (v2), Authentication and Authorization
Roles	A set of permissions. A role must be associated with an individual user, and the user gains the associated authorization when they are associated with the role.	Practical Implications of Public Key Infrastructure for Identity Professionals
RSA	An asymmetric cryptosystem based on large prime numbers. The acronym RSA stands for the three principal inventors, Ron Rivest, Adi Shamir, and Len Adleman.	Practical Implications of Public Key Infrastructure for Identity Professionals
S/MIME	A standard for constructing and sending digitally signed or encrypted messages using asymmetric cryptography	Practical Implications of Public Key Infrastructure for Identity Professionals

Schedule	A document that defines the activity and resources required to achieve the planned deliverable(s) and outcome(s)	Introduction to Project Management for IAM Projects
Scope	A string designating a (part) of a protected resource that a client is authorized to access.	An Introduction to OAuth2.0
Secure Socket Layer (SSL)	A deprecated standard for encrypting data in transit; TLS has superseded it.	Practical Implications of Public Key Infrastructure for Identity Professionals
Security Assertion Markup Language (SAML)	SAML is an XML-based communication protocol between SPs and IdPs. Usually, the enterprise hosts the IdP, whereas applications (including cloud services) are the SPs.	Federation Simplified (v2)
Segment	a grouping of subjects that may be useful for authorizations, such as full-time employees, undergraduate students, IT administrators, or clinicians.	Introduction to Policy-Based Access Controls (v2)
Self-sovereign Identity	A term that describes a digital movement that is founded on the principle that an individual should own and control their identity without the intervening administrative authorities.	A Peek into the Future of Decentralized Identity
Sender Constrained Token	A token whose possession is not sufficient to enable access to a protected resource (additional proof of identity by the client application is required)	An Introduction to OAuth2.0
Server Account	An account with privileged access rights to a server's operation typically used for configuration purposes.	Non-Human Account Management (v2)
Server-based Certificate	A protocol that allows a client to query a server to determine whether a certificate is valid and trusted. The	Practical Implications of Public Key Infrastructure for Identity Professionals

Validation Protocol (SCVP)	server does not need to be associated with the issuing CA. SCVP does two things; (1) it determines the path between the end entity and the trusted root, whereby the client doesn't need to trust any intermediate CAs. (2) it also performs delegated path validation according to policy.	
Service Account	An account used by a computer application to access other applications or services for a specific purpose.	Non-Human Account Management (v2)
Service Provider (SP)	Defined by the OASIS organization, which is responsible for the SAML specification, as “A role donned by a system entity where the system entity provides services to principals or other system entities.” This usually takes the form of an application that offers services requiring authentication and authorization to a user.	Federation Simplified (v2)
Session	A period of time after an authentication event when an RP grants access to resources for the principal/subject. The duration of the session and the mechanism for enforcement vary by implementation.	IAM Reference Architecture
Session Management	A coordinating function provided by an IDP to control sessions of subscribing RPs.	IAM Reference Architecture
Shared Authorization	Shared authorization is provided by a facility outside of the RP. It is shown here as part of the access management grouping.	IAM Reference Architecture
Signature	Processing data using a cryptographic algorithm to provide integrity assurance.	Practical Implications of Public Key Infrastructure for Identity Professionals

Single Sign-On	Single Sign-On is a service that enables SPs to verify the identities of End Users by facilitating communication with IdPs. SSO acts as a bridge to decouple SPs and IdPs. This can happen via numerous protocols such as agent-based integrations, direct LDAP integration, SAML, and OpenID Connect, to name a few.	Federation Simplified (v2)
Social Engineering	Social engineering is a method of manipulating people so they give up confidential information, such as passwords or bank information, or grant access to their computer to secretly install malicious software.	Account Recovery (v2), Designing MFA for Humans
Sources of "Truth"	where authoritative data about individuals live.	Introduction to Identity - Part 1: Admin-time (v2)
Special Category Data (SCD)	Categories of data that are regarded as particularly sensitive, so subject to additional regulation. Defined in Article 9(1) of the GDPR as "personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation"; Article 10's "personal data relating to criminal convictions and offences" requires similar treatment, so is normally considered as another category of SCD.	An Introduction to the GDPR
Step-Up Authentication	A method to increase the level of assurance (or confidence) the system has regarding a user's authentication by issuing one or more additional	Designing MFA for Humans

	authentication challenges, usually using factors different from the one(s) used to establish the initial authenticated session. The need for increasing the level of assurance is typically driven by the risk associated with the sensitive resource the user is attempting to access.	
Subject Alternative Name	One or more identifiers for a certificate subject that certificate issuers can use to carry application-specific identifiers such as email address or User Principal Name (UPN).	Practical Implications of Public Key Infrastructure for Identity Professionals
Subject Distinguished Name (Subject DN)	A unique identifier for the subject within the scope of the Certificate Authority. Issuers structure the subject DN like an LDAP entry name.	Practical Implications of Public Key Infrastructure for Identity Professionals
Subscriber	A party enrolled in the CSP identity service.	Defining the Problem – Identity Proofing Challenges
Sunk cost	Expenses that have already been made in the past and that are unrecoverable.	The Business Case for IAM
System Account	A generic term for a privileged account that has extensive permissions that enable system configuration changes.	Non-Human Account Management (v2)
Task	Lowest level of defined activity; multiple tasks will typically be grouped into stages of project phases	Introduction to Project Management for IAM Projects
Threat Modeling	Threat modeling is an analysis technique used to help identify threats, attacks, vulnerabilities, and countermeasures that could impact an application or process.	Account Recovery (v2), Designing MFA for Humans
Tort Law	The body of law that covers situations where one person’s behavior causes injury, suffering, unfair loss, or harm	Laws Governing Identity Systems

	to another person, giving the injured person (or the person suffering damages) a right to bring a civil lawsuit for compensation from the person who caused the injury. Examples include battery, fraud, defamation, negligence, and strict liability.	
Transport Layer Security ("TLS")	A cryptographic protocol designed to provide confidentiality and integrity of communications between two endpoints.	Practical Implications of Public Key Infrastructure for Identity Professionals
Trust Federation	a trust framework between multiple entities with the purpose of leveraging identity and access management information in a controlled fashion	Introduction to Identity – Part 2: Access Management
Trust Framework	This component represents the legal, organizational, and technical apparatus that enables trust between the IDM and the RPs.	IAM Reference Architecture
Trust Root	A technical structure that provides the IDP and RP the ability to recognize each other with a high degree of certainty. This is similar to the concept of Trust Anchor (NIST SP.800-63-3), but we allow for a structure that relies on a mutually agreed-upon third party. A trust root derives from the operation of a Trust Framework.	IAM Reference Architecture
Two-Factor Authentication (2FA)	A specific case of Multi-Factor Authentication (see: IDPro's Consolidated Terminology) where two factors must be checked to validate a user's identity.	Designing MFA for Humans
Universal Resolver	An identifier resolver that works with any decentralized identifier system through DID drivers. The purpose of a universal resolver is to return a DID document containing DID metadata	A Peek into the Future of Decentralized Identity

	when given a specific DID value. This capability is very useful because DIDs can be anchored on any number of disparate dPKI implementations.	
User or Subject	a person or entity who may receive access within an access control system.	Introduction to Policy-Based Access Controls (v2)
User Agent	A user agent is any software that retrieves, renders, and facilitates end-user interaction with Web content.	Cloud Service Authenticates Via Delegation – SAML
User Provisioning	The means by which user accounts are created, maintained, and deactivated/deleted in a system according to defined policies.	User Provisioning in the Enterprise
User Provisioning and Lifecycle Management	how user records get where they need to be but only as long as they are needed	Introduction to Identity - Part 1: Admin-time (v2)
Username	a common term used for an external identifier	Identifiers and Usernames
Username	An identifier unique to the authentication service used in conjunction with a shared secret to authenticate a user.	Account Recovery (v2), Managing Identity in Customer Service Operations
Validator	An entity that verifies a certificate and confirms that the other party controls the private key in the transaction.	Practical Implications of Public Key Infrastructure for Identity Professionals
Verifiable Credentials	Attestations that an issuer makes about a subject. Verifiable credentials are digitally signed by the issuer.	A Peek into the Future of Decentralized Identity
Verifiable Presentations	The packaging of verifiable credentials, self-issued attestations, or other such artifacts that are then presented to verifiers for verification. Verifiable presentations are digitally signed by	A Peek into the Future of Decentralized Identity

	the holder and can encapsulate all the information that a verifier is requesting in a single package. This is also the place where holders can describe the specific terms of use under which the presentation is performed.	
Verifier	The entity that verifies verifiable credentials so that it can provide services to a holder.	A Peek into the Future of Decentralized Identity
Workforce Framework	An outline of the job categories, work roles, and competency models needed to execute workforce planning.	Identity and Access Management Workforce Planning
Workforce IAM	The application of IAM sub-disciplines such as access governance, authentication, and Authorization for employees as opposed to the applications of such disciplines for customers.	Introduction to Customer Identity and Access Management
Workforce Planning	Activities that ensure an organization has the right talent to execute business and technical objectives.	Identity and Access Management Workforce Planning
X.509	An ISO standard from the X.500 series that defines the basic rules for encoding public key certificates.	Practical Implications of Public Key Infrastructure for Identity Professionals
Zero Standing Privilege (ZSP)	a state where JIT access is used for all permissions and no long-standing permissions are assigned to principals.	Techniques To Approach Least Privilege
Zero Trust	From NIST Draft Special Publication 800-207, “Zero trust assumes there is no implicit trust granted to assets or user accounts based solely on their physical or network location (i.e., local area networks versus the internet)”	Introduction to Identity – Part 2: Access Management

Introduction

Introduction to Identity - Part 1: Admin-time (v3)

By Ian Glazer, edited by Espen Bago

To comment on this article, please visit our GitHub repository and submit an issue.

Table of Contents

Abstract

This article introduces the concepts of digital identity and identity and access management (IAM). It also discusses the constituents that identity professionals serve, compares and contrasts business-to-employee (B2E) and business-to-consumer (B2C) identity use cases, and considers IAM technologies from the perspective of administrative, or admin-time, technologies. IAM technologies and use cases that focus on active, live interactions, or run-time, are mentioned for comparison.

Introduction: How to Approach Identity and IAM

Digital identity is a big topic; it touches every aspect of an enterprise's technical systems and services. This article is not going to offer a taxonomy of identity. Instead, it supports the idea that every individual and organization will likely approach digital identity from a different perspective and level of understanding, given their specific (yet perfectly valid) needs for their local identity system or service.

Identity is an often-debated term. Long-time practitioners and new members of the industry alike struggle with what "identity" means. This article suggests there is not a one-size-fits-all, definitive definition of identity. Instead, it encourages the reader to consider their own local context and adapt the rough definitions here to fit their organization.

This article takes a contextual approach, showing some possible ways of dividing up the IAM world and offering some examples of usage in context. Keep in mind that IAM is not just about technology. It is about the profession itself and us as practitioners.

Terminology

Joiner/Mover/Leaver: The joiner/mover/leaver lifecycle of an employee identity considers three stages in the life cycle: joining the organization, moving within the organization, and leaving the organization.

Sources of "Truth" - where authoritative data about individuals live.

Identity Governance and Administration - a discipline that focuses on identity life cycle management and access control from an administrative perspective.

Privileged Account Management - focusing on special control for risky high-level access. Privileged Account Management (PAM) is a mechanism for getting those special accounts under control.

Identity Proofing - accruing evidence to support "who this is." Identity proofing is the last, but not the least, important part of this admin-time section. This is the process of collecting and verifying information about a person for the purpose of providing an account or a corresponding credential. This is typically performed before an account is created or the credential is issued, or a special privilege is granted.

User Provisioning and Lifecycle Management - how user records get where they need to be but only as long as they are needed

Entitlement Management – the business process to provision access

Role Management - a way to group access rules to make them more manageable

Role-Based Access Control (RBAC) - the use of roles at run-time; a way to govern who gets access to what through the use of roles.

Access Certification – the business process to verify that access rights are correct

Entitlement Management – Cataloging and managing all the accesses an account may have.

Identity analytics and intelligence mean looking at entitlement data, looking at the assignment of that, and trying to figure out and define what risk looks like. IdA provides a risk-based approach for managing system identities and access, with the intention of centralizing governance, visibility, and reporting for access-based risk.

Constituencies - who is it that we serve?

It is easy to lose the forest for the trees in the world of IAM, as there are so many little bits, nuances, abbreviations, and random factoids. Thinking about the ultimate stakeholder for whom the identity work is being done is one way to keep your focus on the big picture.

There are a variety of different constituencies that we serve as identity professionals, which means a variety of different technologies are needed to help them. These groups may include the traditional employee or the more complex groups such as customers, non-paid employees, contractors, and those not within the usual confines of an organization.

Whether that constituency covers employees, business partners, citizens, or students, in everything you do as an identity professional, you should keep the individual's experience in mind. Holding the individual in mind grants more context and a broader view. This approach helps you to realize that "Hey, the reason why I am doing this automated provisioning project is that we're about to hire 5000 new people, and we've got to make them productive on their first day of work."

Business-to-Employee (B2E): Making Employees Productive

For employees and contractors, the primary concern is productivity. The business wants their staff to be productive on day one and want their access removed immediately on separation. The mission here is to get the right access to the right people at the right place at the right time. That's what identity professionals are trying to do: get the appropriate access to people so they can be productive.

More often than not, the Human Resources (HR) department is in charge of employee data, and the HR system is the source of truth. Challenges with this include:

- Potential data integrity issues

- The organization may have multiple HR systems.
- Other non-employee data may (or may not!) reside in this HR system.

Regardless of the challenges involved, this is most typically our source of truth because if someone shows up in the HR system, they are going to get paid, so we need to make them productive; that's a very practical source of truth.

If there is one quote to think about with employee identity, it is "Who has access to what?" It is about making sure that the right people have access to the right stuff. The governing lifecycle, in this case, is the one known as "Joiner/Mover/Leaver":

- People **join** an organization.
- Their roles change as they **move** within the organization.
- Eventually, they **leave** an organization.

The HR system (or systems) acts as a source of truth for employee lifecycle events and related data, such as role or job codes.

Although contractors have similar identity and access-related needs, they may not share the same sources of truth. There may be instances where the HR system does not include the contractor population. Finding a singular source of truth for a contractor can be a real challenge in many enterprises. Some use their procurement system, some use bespoke systems, some use spreadsheets, and some even use their user account provisioning system. For temporary or seasonal workers, it may be most efficient to use a social media identity provider to onboard these types of short-term staff, provided that the organization can obtain the necessary level of assurance.

Business-to-Business (B2B): Connecting to Partners

The next constituency is our business partners. In every industry, we need to connect with our business partners. This connection is really about making sure that members of your supply (or value) chain can interact with you: You are giving them apps to use to work with you, but where do the identity records for these people come from?

Ideally, partners arrive with the identity bits provided by their organization. In that case, we are dealing with the business partner's system of record, likely their HR system, and you are one degree removed from it. This distance often means that you have delegated the administration of doing life cycle management. However, in high-risk applications, the owner of the application may want to control the access rather than trusting the business partner.

From an IAM perspective, B2B and B2E are very similar. The key difference is the source of truth. Often the enterprise doesn't have a system that specifically tracks individuals who are employees of their business partners. Instead, they delegate the management of those people to other systems, either in their own enterprise or in the partner's organization. More often than not, the IAM systems become a de facto source of truth for individual partner identities.

Business-to-Consumer (B2C): Digitally Engage

Last but not least is Business-to-Consumers (B2C). B2C is about bringing whatever the awesome thing is that your organization does or sells to the people. When you talk to the people in your business building the consumer-facing service, you'll often hear them describe the way a consumer interacts like this: "The person is going to do this, and then the person is going to do this." And an identity professional would ask questions like "How did that person get there?" and the answer would be, "Well, yeah, they logged in." And suddenly, you realize that the people building the service have no idea what we as identity professionals do at all. This lack of understanding is an amazing opportunity to make that awesome thing that your organization does get to the right people. That is your mission.

But in this world, the life cycles are different. It is about the individual, the citizen, the consumer. In many ways, they are in control of the life cycle, not you, and you have to be able to accommodate that.

The mission of the business is, "I want to deliver an awesome experience." No one is in the business of just giving people an account and calling it a day. In a B2C setting, you cannot say, "Great! You can log in; I am done here." No, that is just the beginning of the relationship. There is a focus on the customer experience, and we as identity professionals are helping deliver that experience. We are a critical onramp for it.

B2C use cases illustrate that we, as identity professionals, are not alone in our enterprises. We cannot get our jobs done without our peers in security and privacy. There are three legs in this stool to make it work. For privacy, identity provides operational controls, especially in the context of access to data. And for security, identity offers a valuable framework. We put the "who" in the "who the heck is on my network" kind of questions. So if you are working in a B2C (or B2B) setting, and you have not met your peers in the privacy and the security team, go seek them out. They have valuable tools that you can help enrich and that can help you as well.

Technologies Involved - Admin-time vs. Run-time

Having established the constituencies we serve, it is time to look at some of the technologies we use to do that. One approach among many valid ways of sorting out the various technologies and terms is to split the world into administrative (or admin-time) and run-time discussions.

Essentially, the technologies and disciplines used to set things up are on the admin-time side, and the things that are being used when the user is logging in or going through a forgot-password process are on the run-time side.

Admin-time Technologies

The three main areas within the admin-time sphere are:

- Sources of “Truth” - where authoritative data about individuals live.
- Identity Governance and Administration - a discipline that is really about life cycle management and access control from an administrative perspective.
- Identity Analytics and Intelligence - of particular interest to large firms to help assure access is correct.

Two additional areas are also admin-time but do not always fit in the same bucket. Some industry analysts like to add these categories:

- Privileged Account Management - focusing on special control for risky high-level access.
- Identity Proofing - accruing evidence to support “who this is.”

Sources of “Truth”

How do I know who someone is? That may be too difficult a question to answer from both a metaphysical and practical perspective. We can instead rephrase it to: “How can I find reasonably good, authoritative records about people? I need to send their paycheck somewhere.” Or, “I need the shipping address of my business partner. How do I find this data?”[i]

For employees, the answer tends to be HR. For partners, it tends to be that delegated admin one step removed from their HR system. And in consumer settings, things get more complicated. In low-risk areas, the answer is the individual. They are the authoritative source for much of the information you will use. For convenience, this may come from a social media profile, for instance. But in higher risk areas such as financial or medical, the answer may include authoritative sources such as their institution of higher learning or their local government. In an educational setting, a student information system may serve as a source of truth for students.

Data quality is an essential element here. We depend on data for doing things like ensuring people have the right access. But the data from the source of truth is not always reliable, so we may have to operate under the assumption that data quality issues may exist.

Identity Governance and Administration

These are the tools that manage who has access to what. They are the tools that rely on a source of truth (the who) to govern entitlements (the access) in target systems (the what) via connectors.

Identity Governance and Administration (IGA) tools are traditionally more focused on employees, contractors, or students. These tools can often be thought of as more traditional, enterprise-centric tools related to ERP systems.

This area is considerably larger than the other five areas of the admin-time sphere, and our coverage here will focus on the following subsections of it:

- User Provisioning and Lifecycle Management - how user records get where they need to be but only as long as they are needed
- Entitlement Management – the business process to provision access
- Role Management - a way to group access rules to make them more manageable
- Role-Based Access Control (RBAC) - the use of roles at run-time
- Access Certification – the business process to verify that access rights are correct

User Provisioning and Lifecycle Management

User provisioning is the mechanism that helps create, maintain, and eventually remove user accounts in target systems. This mechanism can listen to joiner/mover/leaver events from sources of truth (for example, a connector to the HR system listening for events such as the addition of a new hire). That event then triggers the provisioning system to evaluate the user through business rules in order to undertake required actions, such as create a new user account in Active Directory. The mechanism also has rules describing what those triggered actions are, such as to start setting up access based on some attributes from the new hire data. That typically means assigning entitlements, which can be something that requires approval. For basic entitlements like “birthright” access, we may not need approval. For example, all employees should get access to the productivity suite and email, none of which require approval. If, on the other hand, someone wants to obtain access to the mainframe as a sysadmin, that is going to take some approval. You will have both types — access requiring explicit approval as well as access that does not — in almost all organizations.

A common mistake is to try to automate everything. Avoid this! There are hundreds, if not thousands, of systems and services in your enterprise. Trying to automate provisioning to all of that is just diminishing returns. So what then should be automated? The candidates to look for are the systems with the largest user populations or the highest turnover in those systems. Automation is essential for high-volume or high-velocity systems. Other candidates are systems with too many requests for your helpdesk team to manage, or the ones so sensitive that you want to lock down the rules of who gets access to it. Those make sense to automate.

Day one access systems are excellent candidates for automation. Partly because it is to some extent non-controversial; you get email, you get productivity, you get inside the employee portal, maybe you get VPN. Creating these user accounts has to happen for all new employees and represents a large administrative burden ripe for automation.

After day one onboarding and for the vast number of remaining systems, you are going to provision additional access manually. This means either people will ask for access and/or you will manually create the account (often because you do not need to do it very often.) And in some cases, that system that you want to create an account in exists away from your sphere of direct influence; you will not have a connector to the system. For such systems, the only way you can get to it is by opening up a support ticket, and a human will have to directly access the system to create or change the user account. These typically do not need to be automated.

Lastly, provisioning systems are often involved in setting up passwords. This involvement means that provisioning systems often need to have aggregate password content rules. That exposes all sorts of challenges because different systems can have radically different internal rules and password capabilities. For example, you may have a password content rule that mandates the inclusion of a special character. Because of system proclivities, a person could provide a password with a special character that the Oracle database could not accept, but Active Directory could. User provisioning (or password management) systems have to deal with these potential problems as gracefully as possible.

Entitlement Management

Now we have a source of truth and users flow into a data repository, and that triggers our user provisioning systems and starts creating users in our target applications or services. But it is not enough just to create a user account; we also have to know what that account can do. This set of actions is what we call entitlement management. Entitlement management can get really detailed really fast because the total of all the little privileges that govern what a user can do in a system can be extremely numerous. It is not unheard of to have hundreds, if not thousands, of individual privileges in a system. Those privileges are often aggregated into user groups or roles, which can also become quite numerous. It is like grains of sand on the beach, which is why we try to aggregate them together. Imagine you have three employees, one system, and four privileges in it: Create Purchase Order (PO), Update PO, Read PO, and Delete PO. Connecting each person to the right collection of privileges is possible, but it becomes unmanageable very quickly.

That’s where layers of abstraction come in: We put this thing in between the user and the privileges called an entitlement. We say, “This allows you to manage purchase orders.” And it is these things that the provisioning system hands out, instead of the detailed privileges themselves, because there are way too many discrete privileges to keep track of. We abstract the details and instead say, “Here’s an ability,” or “Here’s something associated with your job responsibility.” Unfortunately, those discrete, detailed privileges still need to exist in order to allow the level of granularity an organization’s business processes require, and to provide the level of instruction to the system that can be coded into the environment.

Entitlement management means cataloging all the accesses a person can have, which can be a massive undertaking. For example, a medium-sized bank may have ten major systems (but often a lot more), which means you may have thousands upon thousands of privileges, which are aggregated into a thousand or so entitlements. You then need to figure out how to map that to the business needs. Entitlement management is this cataloging process.

Ideally, you are bundling privileges together into sets that make some semblance of sense for people and the organization. For example, imagine that you want to gather all the entitlements together that someone who works in purchasing would need. Or that you want to make sure you have put together the relevant entitlements that someone who is a business partner - at the gold tier but not the silver tier - has access to. This level of efficiency is what you and your identity colleagues are working towards to make access to enterprise resources manageable.

It also tends to be mandatory work if you’re ever going to do Segregation of Duty[ii] analysis, for example for Sarbanes-Oxley (SOX)[iii] compliance or General Data Protection Regulation (GDPR)[iv] compliance requirements, where we have to identify combinations of accesses that together are dangerous. For example, people who can authorize payments to partners should not be able to create a fraudulent partner and pay them.

Role Management

But even when we have worked these entitlements down to a level where they are manageable entities in themselves, using them effectively will be very challenging. The answer to that challenge for many, though not all, organizations is to try to do something called Role Management. You may have heard of it as Role-Based Access Control (RBAC). The essence of it is as follows: in some organizations, job functions are very regular. Regular job functions are most typically found in hierarchical organizations. On the other end of the scale, this works quite poorly in matrixed organizations; that is because it is hard to pinpoint, for example, the three top job responsibilities, as they are always shifting.

Role management can be useful for saying, “These types of job responsibilities need this kind of access, so let’s call that thing a Role.” Additionally, sometimes you have this thing called a technical role, which is saying, “Here are the low-level bits you’re going to need to do your job,” and it becomes a handy bundle to assign to people. Imagine roles as a grouping to which you might provide access in a common way. You should only create roles if you are going to provision or control access to a group differently. At the highest level, you could only have a handful of roles, and you should review them regularly as your organization evolves.

Role-Based Access Control

RBAC is just a way to govern who gets access to what through the use of roles. There is no need to overthink it. It works great in regular, hierarchical, homogeneous organizations. Not surprisingly, it works really well for places like the US Department of Defense. It does not work great for the 150-person start-up; do not try to do that.

When overthinking RBAC, or over time in general, you can get what we call a “role explosion,” where you have more roles than you have people. Some of the salty veterans say, “oh yeah, I survived that. They told me that was a good idea. It was a horrible idea.” Try to avoid this situation; it rarely ends well.

Access Certification

At this point, people have access. They are productive on day one. They are productive on day two. On day three, they start to become a privileged threat because they have too *much* access. And the best way to get more and more capabilities in an enterprise is to simply change jobs. You go from doing this job to another job, and if your business rules didn’t explicitly prevent the continuation of access, experience shows that you do not lose your old access, you just keep it on top of new accesses.

Another instance of accumulating accesses is when you onboard someone new, with the dreaded situation of “Whose access should we model you after.” Anyone who has gone through any kind of IT security audit knows how horrifying an audit can be. A common answer to the question of whom to model a set of new accesses to give to a new employee is their boss. The boss is likely to be the biggest source of access violation around because they have accreted access over years. They are a horrible person to emulate for this purpose, from a risk management perspective.

Certification is the ongoing review of who has which accesses, a process that became popular with the introduction of the Sarbanes Oxley law (SOX) in the United States. It is a great tool to prevent people from keeping access they no longer need. An auditor might say that this should be done quarterly, something that quickly becomes very fatiguing. Better methods may be to trigger reviews based on changes to entitlements, changes to overall user risk, or to try to detect if someone deviates from a norm. In other words, we want to certify whether, if compared to a set of peers, you are an outlier. We want to figure out why that is. It is a powerful way to make sure you don’t have issues with the access and entitlements that you’ve assigned in a non-automated fashion.

All of the mentioned elements add up to a lot of data flowing around. With all the users, times all the systems, times all the entitlements, times all the roles, times all the privileges, the total is staggering. How can we make sense of it all?

Identity Analytics

One of the ways is through identity analytics (IdA) and intelligence, which is more than just reporting on "Who has access to what."

Identity analytics and intelligence mean looking at entitlement data, looking at the assignment of that, and trying to figure out and define what risk looks like. What does a normal user look like? Compared to that, what does a heavily privileged user look like? What does the model of a system administrator look like compared to developers, someone working in finance, or someone working in the field sales organization?

The goal is to find commonalities of outliers among user populations and to understand what access-related risk looks like in the organization. Other goals of IdA include being able to group commonly assigned entitlements together as candidate roles, to identify over-privileged users, to discover undocumented high privileged access rights assigned to regular, non-privileged accounts. IdA can also accurately measure and report on user, account, entitlement, application, departmental, and organizational risk posture.

IdA provides a risk-based approach for managing system identities and access, with the intention of centralizing governance, visibility, and reporting for access-based risk.

It uses dynamic risk scores and advanced analytics to determine the associated level of risk and to derive key indicators for automating account provisioning, de-provisioning, authentication, and privileged access management. This approach reduces the identity attack surface by identifying (for remediation) unnecessary, unused, and outlier access.

Another feature of this admin-time function of risk determination is that the indicators and data it produces can be integrated with, and used by, your run-time systems. For example, during the login process: If we know that a person is not particularly risky, then they might not need to be challenged for additional authentication factors. But if, on the other hand, that person has a lot of privilege and power in the systems, and maybe they deviate from the norm in their job role, then they might need to provide additional verification. Run-time risk determination analysis such as this can be partly or fully automated, depending on the quality of the indicator data and the maturity of the organization using them.

Privileged Account Management

Some of the user accounts out there are special. Your sysadmin accounts, your root accounts, and so on. These accounts are not necessarily tied to people. But they are super privileged user accounts. We may have a whole team of people that have to act like the root administrator. Privileged Account Management (PAM) is a mechanism for getting those special accounts under control. You can essentially check them in and out as needed: “Hey, I need to go in and apply a zero-day patch, so I need to act like the root administrator for this,” and the system will grant the relevant access for that purpose, after validating who the user is. It may also record the screen, so that as the user is performing their actions, what’s going on is being logged. One use case for this could be when having a third-party service vendor who’s going to come in and do maintenance on specialized pieces of equipment, where we need to have an audit log of the actions that they took. This log would be like the record function in privileged account management. Another important function is scrambling the password for these special accounts, so that no one retains the password to the root or sysadmin account after the job is done, such as the patch job in the example above.

Identity Proofing

Identity proofing is the last, but not the least, important part of this admin-time section. This is the process of collecting and verifying information about a person for the purpose of providing an account or a corresponding credential. This is typically performed before an account is created or the credential is issued, or a special privilege is granted. It also tends to be a lengthier process the first time we encounter a particular individual, as opposed to the secondary proofing required for purposes of account recovery.

The process is often found in regulated industries, such as in banking, with requirements for doing Know Your Customer (KYC) and anti-money laundering. These require government documents to be presented in some fashion, proved to be accurate and valid, and then associated with the individual. This is the proofing process.

Depending on the account or credential to be issued, there are different ways of doing proofing, many tied to government-issued identity. In contrast, others are based on what we call self-asserted.

In an enterprise setting, B2E, relating to employees, proofing is a very common process, involving background checks and showing documentation (for example, your passport or a driver’s license) to get your job. For employees, we want to do this because this is how we will get a new job. But for a B2B setting, it is a very different situation. How do we onboard a new business partner? In addition to making sure who that person is, we may need proof that this is the organization we want to work with and that *this* person is someone we want to work *within* that organization. These different criteria make for a very different kind of proofing process.

What about identity proofing in B2C use cases? How does one know and trust a new customer who makes a claim about themselves? Here it is a question of how much we need to care about that. There is a trade-off in the B2C world between velocity versus veracity. For some organizations and apps, the priority will be velocity - to get people registered quickly and into the app as fast as possible. The user journey is optimized for this. They'll have very limited access, and the threshold for user registration is very low.

For others, the priority is veracity, either because of the brand experience, because of the business they're in, because of what they want to deliver in terms of value, or related to the chance of fraud. In this case, the enterprise wants more verifiable data about the person. The enterprise determines it is important to have a higher level of assurance that the new customer is really the person they claim to be.

Conclusion

Digital identity, as we indicated at the beginning, is a big topic. We've touched on the constituencies IAM serves, the technologies involved, governance, analytics, privileged accounts, and identity proofing. Each of those topics can (and hopefully will, in future versions of the IDPro Body of Knowledge) fill out an entire chapter by itself.

The article offered the IAM practitioner a chance to understand some of the major considerations that will impact their systems and services; readers need to consider their own local context and adapt the rough definitions offered here to fit their own unique organization. A future article will dive into the concept of run-time technologies.

Author Bios

Ian Glazer

Ian Glazer is the founder and president of Weave Identity – an advisory services firm. Prior to founding Weave, Ian was the Senior Vice President for Identity Product Management at Salesforce. His responsibilities include leading the product management team, product strategy, and identity standards work. Earlier in his career, Ian was a research vice president and agenda manager on the Identity and Privacy Strategies team at Gartner, where he oversaw the entire team's research. He is a Board Emeritus and the co-founder of IDPro and works to deliver more services and value to the IDPro membership, raise funds for the organization, and help identity management professionals learn from one another. During his career in the identity industry, he has co-authored a patent on federated user provisioning, co-authored and contributed to user provisioning specifications, and is a noted blogger, speaker, and photographer of his socks.

Espen Bago

Espen Bago realized in 2002 that as system administrator, he'd been working in identity already for a while and decided from there to fully explore what this Identity thing was all about. He's been an independent Identity Advisor and coordinator to large enterprises for the last few years, but in 2021 became Identity Manager for the Norwegian Labour and Welfare Administration. As such, his goal is to make certain that identities – and the real persons this represents – are not forgotten when governments inevitably go all-in digital. He's also a founding member of IDPro and a member of the IDPro Body of Knowledge Committee and the IDPro Certification Committee.

Change Log

Date	Change
2023-10-27	V3 published; removed extra Access Certification definition; updated Ian Glazer's bio
2021-06-30	V2 Editorial updates; addition of a Terminology section; update to B2E section
2020-03-31	V1 published

[i] An organization does not always need to know "who" a person is to any level of specificity. They just need to know things like "is this the same person each time" or "is this account authorized to perform this action."

[ii] AICPA, "Segregation of Duties," accessed on 11 January 2020, https://www.aicpa.org/interestareas/informationtechnology/resources/value-strategy-through-segregation-of-duties.html

[iii] United States. 2002. *Sarbanes-Oxley Act of 2002,* [Washington, D.C.]: [U.S. G.P.O.], https://www.govinfo.gov/content/pkg/PLAW-107publ204/pdf/PLAW-107publ204.pdf.

[iv] *EU General Data Protection Regulation (GDPR):* Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1, https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en.

Introduction to Identity - Part 2: Access Management

By Pamela Dingle

Table of Contents

Abstract

Who are you, and what are you allowed to do? In digital systems, these questions are the domain of Identity and Access Management (IAM). Access management systems provide the mechanisms for deciding who is who, and to evaluate and enforce decisions about who should get access to what. Part 2 of the introduction to the IDPro Body of Knowledge explores the big picture of access management from a historical perspective. You can expect a little advice, a lot of context, and an experience-based overview of what we do in access management and why our contributions matter.

Terminology

- Ceremonies - predictable interactions that users can infrequently navigate in a well-watched place
- Delegated authorization framework - an access control framework that decouples authentication from authorization, allowing the password to stay local and protected.[i]
- Federated Identity - the means of linking a person's electronic identity and attributes, stored across multiple distinct identity management systems.[ii]
- Least privilege - also known as the Principle of Least Privilege; a resource, such as a user, must only be able to access the resources (e.g., applications, data) that are necessary for it to function.[iii]
- Trust federation - a trust framework between multiple entities with the purpose of leveraging identity and access management information in a controlled fashion.
- Zero trust - From NIST Draft Special Publication 800-207, "Zero trust assumes there is no implicit trust granted to assets or user accounts based solely on their physical or network location (i.e., local area networks versus the internet)"[iv]

Introduction

What is access management, and why is it so exciting? There is something thrilling and urgent about the moment a decision is made, a gate is lifted, and a precious resource is made available to a stranger. Did we make the right person productive, or did we make a risky mistake? Good access management depends on good identity data; it also requires policies that represent corporate rules, an accurate understanding of current environmental and contextual factors, and tools that can enforce according to a defined risk tolerance. A lot of preparation and consideration goes into the run-time decisions that are made every day and that operate with all kinds of granularity at infrastructure, middleware, and application layers.

If you are an experienced identity professional, you have watched our tools evolve - but if you are just starting, it can be valuable to hear some perspective on why things are the way they are. Hold on to your hats: this introduction is not even remotely objective, but it will give you one perspective on how we got here and how the concepts discussed in later chapters have evolved into our current access management landscape.

To kick off the ride, here are a few critical realities to keep in mind in the world of access management:

Resources need stability

Company secrets, financial transactions, and personal communications are just a few examples of the precious resources that identity professionals are tasked with protecting. Resources may be exposed through application programming interfaces (APIs), web interfaces, or native mobile applications. Adding externalized access management capabilities to a single resource is relatively easy, but adding to a hundred or a thousand is exhausting. Owners of these applications rarely want to make frequent changes. After the first time, you as an identity professional try to schedule an application access management update within the change management windows of a hundred different applications, you will feel the same way.

Resources should not perform local identity management

If every resource you deploy performs its own login functions, it is nearly impossible to ensure that they follow the kinds of best practices detailed in places such as NIST 800-63B or adhere to unified corporate policies.[v] Hundreds of applications each separately attempting to store credentials, protect a login page, and secure an account recovery process present an immense attack surface and make it likely that users will reuse passwords across applications. This pattern means an attacker who guesses the password to one application has a credential that can be replayed to gain access to other applications, and you have no way to know which applications are at risk.

Humans need challenges, but not obstacles

While resources need stability and consistency, humans need empathy. We require users to interact with computer systems to show they are the proper operator of the digital account they claim to have a right to; this process should be easy for a good user and tough for an impostor. The best practice is to create “ceremonies” - predictable interactions that users can infrequently navigate in a well-watched place. While authentication is the best-known ceremony, there are many other ways in which humans are asked to interact, such as self-service registration or account recovery, notifications, or transactional approval. We want users to notice when an unusual ceremony takes place because it may alert them that fraud is happening. Ceremonies are guaranteed to change as new attacks force administrators to try additional techniques, including changes in user experience (UX), authentication factors, and risk detection. While it is important to keep the attackers out, the experience of the good users is critically important. Faced with a tough problem, humans often behave predictably, and that predictability is an attack vector in itself. If you as the administrator make your users’ lives too hard, you become the problem: Users will circumvent the controls you put in place to try to protect them.

Garbage In, Garbage Out

The most visible parts of access management are decisions made in the moment, but those decisions do not exist in a vacuum. Before any access management decision is made, someone has to set up digital rules and policies that closely approximate the business goals of the organization (see “Introduction to Project Management for IAM Projects” for more on managing an IAM project).[vi] User, group, and role context must exist, and some combination of device, network, and risk context as well. By the time a user attempts to access a given resource, all of the data that might go into an access choice should be available. Never forget: It doesn't matter how good your access management infrastructure is if decisions are based on bad input.

Now, on to the Fun Part

Identity professionals end up at the forefront of an age-old problem. We have resources to protect, users who want access, and attackers who want access as well and are really hard to distinguish from users. We need a system that is accurate, but no system will be 100% accurate, so the system must also follow the principles of zero trust, starting with least privilege. We must strongly authenticate users and leverage the environmental context to detect fraud.   We must apply a single consistent policy view across a disparate landscape of resources. And we have to verify all the time that our systems are working the way we think they are.

Access Management as an Evolution

This body of knowledge will give you all sorts of data about the basic concepts that are deployed in an access management regime - but why do those mechanisms exist? They evolved in response to both business requirements and security threats. Administrators found themselves lacking in control and created best practices that made administration at scale easier and attacks at scale more difficult.

Password Proliferation Gave Us Directories

When businesses first began accumulating business programs within their private network, every new program required that user accounts be created and deleted. Every program asked each user to set a password. As businesses grew to have hundreds and thousands of programs, users hit the limit of how many usernames and passwords they could remember. Some programs let users choose their own usernames, and as a result, usernames varied wildly across programs. Many programs had wildly varying password policies. It was the wild west and from that wild west came the concept of "directories". Instead of a hundred programs separately storing usernames and passwords, applications began to call out to an external directory of users, often using LDAP (Lightweight Directory Access Protocol).[vii],[viii] Suddenly, users could use one password everywhere, and administrators didn't have to maintain thousands of applications individually. All was well… for a while.

Password Fatigue Gave Us Web Access Management

The upside to user directories and LDAP was that users only had to remember one password. The downside was that even if all applications at the time were within the same network perimeter and were all LDAP-integrated, the user was still prompted for their password every time they used a new application - over the course of a day, that was a lot of typing. The resulting innovation was a new access management technique called "Web Access Management" (WAM).[ix] With web access management, users would authenticate once with their password, and then a (usually encrypted) domain-wide session cookie would be generated that could be read by multiple applications. Instead of performing an LDAP "bind," the application could check that the user had a valid cookie. Around the same time, other technologies to address password fatigue developed, including Kerberos.[x] These technologies finally give users some relief; a user could log in one time and access multiple applications. The concept of logging in once to access multiple apps has come to be known as 'single sign-on' (SSO).

Perimeter Limitations Gave Us Federation

As long as businesses were operating within their network perimeters, access management functions like Kerberos and WAM provided both convenience and security. But the Internet was opening up, and many companies wanted to begin allowing not only their employees to access resources, but also partners and customers. Businesses wanted to create trust relationships with other businesses and enable their users to access each other's applications. This desire was met through a standard called SAML (Security Assertion Markup Language).[xi] Businesses pre-establish a trust "federation" between two domains and then request a secure introduction whenever a user attempts to access a resource. SAML and other federated identity specifications allowed businesses to retain control over the activities of their own users both in their own domains and across domains. Federated identity remains a backbone of access management, and SAML is still the gold standard for cross-domain access management.

Mobile & API Innovation Gave Us OAuth & Delegated Authorization Frameworks

Federation and SSO are what we call in the industry "user-present" scenarios. We can tell that the user is present in a federation request because the activity occurs using a browser, and browsers don't have brains - they are 'passive' clients, and somebody has to be there to push the buttons and click the links. Around 2007, most business application delivery was focused on the browser - but the release of the first "smartphone" changed the game. Mobile applications could be downloaded from an app store and render data accessed from cloud APIs, just as cloud platforms were becoming popular. Suddenly an 'active' software client became a desirable way to talk to users.

Even as users got excited about the power of mobile applications, identity professionals ran into a problem: applications were calling APIs when users were not present, and even worse, many mobile applications wanted to consume and display data from cloud platforms that they were not affiliated to. If a mobile app wanted to access an unaffiliated cloud platform, the only answer was to ask the user for their password and then replay the password within every single API fetch. The result was something called the **password anti-pattern**: users got used to giving away their cloud platform passwords to any client that asked for it, and those clients had to cache user credentials on mobile devices so they could execute API calls in users' absence.

SAML was not a perfect fit in a mobile context. XML parsers were not built into mobile platforms, and cryptographic requirements were heavy. The resulting access management paradigm was OAuth 1.0, a “delegated authorization framework” that could layer with federated protocols. OAuth addresses the ‘user not present’ scenario: applications ask for and receive an “access token” that does not introduce the user; instead, access tokens represent the ability to access a tightly scoped set data and services on behalf of a user.

Maybe access tokens don’t sound like such a big deal, but when you consider that you can pass access tokens to APIs instead of primary credentials, the results are significant. You prevent API endpoints from ever collecting or validating primary user credentials, thus removing multiple attack vectors around data leakage, man-in-the-middle-attacks, and rogue administrators harvesting credentials. Because the mechanism for authorizing the API is decoupled from the mechanism for authenticating users, the door opens to a world where a user could authenticate with factors other than a password without causing work for applications. Access tokens act as a stable currency that can be centrally architected and scalably deployed.

Multi-Factor Authentication (MFA) Is and Was and Will be Again

Through all of the above antics and shenanigans, password attacks were haunting identity administrators. All sorts of conventions evolved to try to keep attackers out of accounts they didn’t own: we forced people to change their passwords regularly; we forced them to set longer and more complex passwords; we designed our LDAP directories and login forms to stop responding if too many incorrect attempts were made. Despite all these attempts to mitigate the risk, almost any password a human could set and remember without help is trivially attackable. If you doubt this statement, read “Your Pa$$word doesn’t matter“ by Alex Weinert (@alex_t_weinert).[xii] Be prepared to weep.

The revelation that passwords are fundamentally flawed is not new - dating back to at least the ’70s, there has been research on how to get around the need for a human brain in the authentication process.[xiii,xiv] We developed the simple idea that passwords are “something you know,” but also described other options for validating a human’s ownership of a digital account could also include “something you have” or “something you are”. The idea is not that validating the thing you have can replace the thing you know, but rather that a combination of things you have, are, and know would require an attacker to compromise both digital and physical information. Today, the state of the art in multi-factor authentication is very sophisticated. A growing number of users protect their phone with a biometric, navigate an SMS message to confirm a transaction, or use an OTP (one-time password) to improve security without any need to understand the underlying principles.

We all know that MFA must continue to improve in usability to become ubiquitous. Specifications like FIDO2 are industry-changing for access management, not because the problem is solved - but because the problem is ***decoupled*** - FIDO2 (W3C WebAuthn and

FIDO CTAP2) has separated the problem of negotiating cryptographic keys from the problem of requiring user gestures.[xv] The cryptographic key exchange can now stay reliable, while we focus on innovation - and possibly even revolution - in user interactions.

The Best Security is Invisible Security

In addition to the visible ceremonies we put in front of those who attempt access to resources, a lot is happening beneath the surface. We increasingly rely on context to supplement active user challenges in calculating the risk of any given transaction. Adjacent areas to identity are now critical stakeholders in our attempts to prevent identity fraud - Cloud Access Security Brokers (CASBs),[xvi] Unified Endpoint Management (for example, Mobile Device Management or MDM),[xvii] and EUBA (Entity and User Behavioral Analysis)[xviii] fortify our access management regimes. Attackers have learned to defeat static access management processes, so we have evolved our defenses beyond password complexity: if you are not checking passwords against a rapidly updated set of banned strings including lists of newly known-to-be-breached passwords and augmenting this with real-time threat intelligence you are in serious trouble.

And the Moral of the story is…

That brings us to now. Identity professionals today still struggle with all of the anecdotal issues listed here, but we have tools at our disposal and conventions on how to best deploy them. The better we can get as a profession at working together to eliminate fraud, detect abuse, and guide our users towards successful interactions, the better off everyone is. Everyone before you leveraged the work of their contemporaries to take a step forward. Now you have the opportunity to take the next step.

What Will Access Management look like in the Future?

When we look back on today's world of access management, what stories will be our contribution? There will be an assessment of our success in helping users to adopt multiple factors - did we succeed? Did we miss opportunities? As long as we are timid, a huge chunk of our immediate future will be spent mitigating attacks that we already know are mostly preventable. Dragging your feet on MFA as an access management professional today is like catching up on social media when you know you have a report due (a behavior common enough to have its own name: akrasia)[xix]. After the fact, we will ask ourselves why we got in our own way, and there will likely be no good answer.

At some point, when enough administrators adopt MFA and eliminate the easy jackpots that are single-factor passwords, our industry will win this amazing prize:

A whole new wave of inventive attacks!

That may not sound so great, but it really is. Today, attackers can spend almost no money or time and still make a living from doing nothing fancier than running free phishing scripts from the Internet. A strongly authenticated world does not eliminate jackpots, but it does make the pool of criminals able to win those prizes a much more distinguished group. Attackers will move to post-authentication attacks like token theft and consent abuse. And the whole time, identity professionals and others will be making new things! Inventing better ways! Introducing resources and content that businesses want! We will embrace wearables as security devices, perform secure transactions even in hostile places, make the measure of least privilege even tighter. We will get better at tracking the promises that products make to us and better at punishing those who mess with our data. We will find a way to share private things and have true confidence that those private things will never become public. We will weather quantum meltdowns and new social networks, and it will all be a fight worth fighting.

The identity management professional who has read this far is clearly dedicated - and that is a great thing. We need the next generation of professionals to pick up the torch, question all assumptions, and push us into a future where risk is low, productivity is high, and new challenges keep our lives interesting.

Author Bio

Pamela Dingle is a long-time member of the identity management world, and is a Director, running the identity standards team at Microsoft. The identity standards team works with standards bodies like W3C, IETF, and the OpenID Foundation on specifications like OAuth 2.0, FIDO, SCIM, and OpenID Connect, and Pamela works to ensure that customers, product groups, and the industry all understand the value of standards and other identity best practice patterns. Pamela spent eight years as an identity architect and eight years in the office of the CTO at Ping Identity and is a founder of Women in Identity.

[i] Raible, Matt, "What the Heck is OAuth?" DZone Security Zone, 28 April 2018, https://dzone.com/articles/what-the-heck-is-oauth.

[ii] Wikipedia contributors, "Federated identity," Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/w/index.php?title=Federated_identity&oldid=949399706 (accessed June 6, 2020).

[iii] Wikipedia contributors, "Principle of least privilege," Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/w/index.php?title=Principle_of_least_privilege&oldid=950981064 (accessed June 6, 2020).

[iv] Rose, Scott, and Oliver Borchert, Stu Mitchell, Sean Connelly, "Zero Trust Architecture (2nd Draft)," SP 800-207 (Draft), National Institute of Standards and Technology, February 2020, https://csrc.nist.gov/publications/detail/sp/800-207/draft.

[v] Paul A. Grassi, James L. Fenton, Elaine M. Newton, Ray A. Perlner, Andrew R. Regenscheid, William E. Burr, and Justin P. Richer. 2017. Digital identity guidelines - Authentication and Lifecycle Management. Technical Report. NIST Special Publication 800-63B.

[vi] Graham Williamson and Corey Scholefield. Introduction to IAM Project Management for IAM Projects. IDPro Body of Knowledge, volume 1, issue 1, 31 March 2020. https://bok.idpro.org/article/id/25/.

[vii] Wikipedia contributors, "Lightweight Directory Access Protocol," Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/w/index.php?title=Lightweight_Directory_Access_Protocol&oldid=960496535 (accessed June 6, 2020).

[viii] "The Most Complete History of Directory Services You Will Ever Find," blog, Easy Identity, 13 April 2020, https://idmdude.com/2012/04/13/the-most-complete-history-of-directory-services-you-will-ever-find/ (accessed June 12, 2020).

[ix] Wikipedia contributors, "Web access management," Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/w/index.php?title=Web_access_management&oldid=959341667 (accessed June 6, 2020).

[x] Wikipedia contributors, "Kerberos (protocol)," Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/w/index.php?title=Kerberos_(protocol)&oldid=960957884 (accessed June 6, 2020).

[xi] Wikipedia contributors, "Security Assertion Markup Language," Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/w/index.php?title=Security_Assertion_Markup_Language&oldid=956779073 (accessed June 6, 2020).

[xii] Weinert, Alex, “Your Pa$$word doesn’t matter,” Azure Active Directory Identity Blog, Microsoft Corporation, 9 July 2019, https://techcommunity.microsoft.com/t5/azure-active-directory-identity/your-pa-word-doesn-t-matter/ba-p/731984.

[xiii] Morris, Robert, and Ken Thompson. "Password security: A case history." Communications of the ACM 22.11 (1979): 594-597.

[xiv] Feldmeier D.C., Karn P.R. (1990) UNIX Password Security - Ten Years Later. In: Brassard G. (eds) Advances in Cryptology — CRYPTO’ 89 Proceedings. CRYPTO 1989. Lecture Notes in Computer Science, vol 435. Springer, New York, NY

[xv] “FIDO2:WebAuthn & CTAP,” FIDO Alliance, https://fidoalliance.org/fido2/.

[xvi] Wikipedia contributors, "Cloud access security broker," Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/w/index.php?title=Cloud_access_security_broker&oldid=949494699 (accessed June 6, 2020).

[xvii] Raam, Giridhara, “The What, Why, and How of Unified Endpoint Management,” Integration Zone, DZone, 8 July 2019, https://dzone.com/articles/the-what-why-and-how-of-unified-endpoint-managemen.

[xviii] Petters, Jeff, “What is UEBA? Complete Guide to User and Entity Behavior Analytics,” Inside Out Security Blog, Varonis, 29 March 2020. https://www.varonis.com/blog/user-entity-behavior-analytics-ueba/

[xix] Clear, James, “The Akrasia Effect: Why We Don’t Follow Through on What We Set Out to Do and What to Do About It,” excerpt from Atomic Habits, https://jamesclear.com/akrasia (accessed June 6, 2020).

Words of Identity

A pragmatic guide to the confusing terms and words of Identity and Access

By Espen Bago

Table of Contents

Abstract

Identity and Access is a complex topic covering a wide range of topics and sub-areas. Getting a grasp on this is difficult for anyone; understanding the subject well enough to explain it to others and collaborate is even more challenging.

A natural starting point for understanding a complex topic or area is to seek to learn the native language, such as the professional jargon of the area. And many of us start precisely that way, by searching out dictionaries containing words and terms of "IAM" and its sub-components.

There are many such dictionaries to be found. And the subtle—and not-so-subtle—differences in the definitions found in these illustrate the challenge this article aims to alleviate. We may have a de facto "*language of Identity and Access*," but this language has no formal structure to its semantics.

Our area of technology and business is characterized by an abundance of words having multiple meanings. These are confusing or frustrating for beginners and experienced practitioners alike.

This article is intended as a guide for keeping track of ambiguous words and terms often encountered in our industry, as well as to show that this deficiency in understanding may be even more significant in Identity and Access than in other places.

The article offers examples of where terms are ambiguous and definitions seem to vary across the industry. These examples serve as an aid both to lessen confusion and encourage better and clearer usage of the terms. This article also discusses reasons for the differences and offers some suggestions on countering this challenge in the line of work.

Introduction

None of us in this industry work with bricks and mortar or other tangible, real objects. Everything we do—in IT, not just in Identity and Access—is instead a digital representation, an abstraction, of something that might exist in the real world.

Identity and Access are the glue for many of those digital representations. This puts a lot of responsibility on our representations to be extra reliable, understandable, and able to be proven correct. This concept of representation may be the most important thing to understand when considering, interpreting, and choosing between the different possible meanings of words.

Practitioners new to Identity and Access quickly realize that many of the words they encounter have different meanings than they first thought. One of the first words encountered is “identity” itself. Some will think they know what it means, and others will stop and think and ask. Does “identity” mean the same as “user”? Does “user” mean a person, or does it mean some digital object within IT systems (like a “user account”)? The difference is often obvious to the author or originator, but less so for the rest of us.

But since many people—newcomers and old hands alike—are reluctant to show (perceived) weakness in front of perceived experts, questions are too often not being asked when they are unsure. As in any industry, a typical consequence of miscommunication is that the end product or project is of lower quality or takes longer to deliver.

Another aspect of the problem is the differences in dialects between separate companies and organizations. Learning the local dialect may be achievable, but realizing that other organizations and products have divergent definitions can be a surprise.

There is no quick fix available for the ensuing confusion, but it may help to be aware of the most commonly diverging meanings and their context. The following list is a sample of words and terms where Identity Professionals have experienced significant ambiguity.[i]

Words

Notes on Specific Words and Terms in Identity and Access

This terminology section highlights how common terms are defined differently within the same industry. It is not intended to suggest definitive language for any term included—the focus here is on showing existing usage variations.

Access Right, Entitlement, Permission, Privilege, Profile, Role (and More)

There are multiple words that (mostly) mean almost the same as the term *access right* or simply *access*. One challenge is that sometimes they are used interchangeably as pure synonyms. At other times, each word is ascribed a slightly different meaning, often denoting different granularity of access in a hierarchy when one word is meant as being a subtype of another. But such usage is only defined by local customs rather than universally. Often, we see such specific usage as part of a specific vendor terminology or in a particular standard. The suggestion here is to by default assume these words to be synonyms, and if there is a need for them to have distinct and significant meanings, describe these meanings locally and make certain the description follows the text wherever it is used. Each potential synonym listed above has separate entries below, noting some of their possibly distinct meanings in particular contexts.

Account

The word *account* has its origin in the act of counting something. Identity and Access often denotes "*user account,*" as in an IT system's digital user object or user record. But *accounts* in a bank, insurance company, or customer relationship system differ from the *user accounts* an IT department might speak of.

In such situations, using "account" in documentation and description will cause confusion unless it is made unambiguously clear how to understand it.

But user accounts and accounts do not exist isolated from each other. Financial systems exist where "users" (or persons) can have one or more "accounts." Similarly, Customer Relationship Management (CRM) systems exist where customers (or persons in general) can have one or more "accounts." Both need to interact with Identity and Access systems. CRM and finance are just two examples of a word taking on a different meaning when the context changes or varies.

Authentication

Authentication is often described as "the process or action of proving or showing something to be authentic, true, genuine, or valid." Note that this does not necessarily mean the entity is mapped to a known, verified, natural person. This is often a prerequisite to *allowing* access to *resources* in an information system. In that context, authentication is often confused with authorization, as in many erroneously thinking that if someone authenticates successfully to a protected resource, they should also have access to it. The *authorization* process does not follow automatically from *authentication*, and each of the process steps needs to be clearly and distinctly described.

Authentication Factor

Continuing from *authentication* above, further potential for confusion is related to the varying understanding of the individual building blocks or elements of that process. For example, *authentication* often requires multiple "factors in Identity and Access." But "factor" is often interchangeably used to mean both "categories of factors" and "specific or individual factors." This ambiguity tends to make understanding harder. Descriptions of *authentication* also often contain confusing usage of components such as *identifier* (e.g., the identifying key, text string, numeral) and *authenticator* (e.g., the password, hardware key, biometric fact) of the process.

Authorization

Authorization is, as indicated above, sometimes confused with *authentication,* although they are different processes. Even in the IDPro Body of Knowledge, the definitions diverge slightly based on the context of the article.[ii] Apart from this, the complexity and lack of one standard for authorization gives rise to confusion. For more examples and information, see the IDPro blog post, The State of the Union of Authorization.

Entitlement

Entitlement is often used as a synonym for *access rights,* as mentioned above. Since "being entitled" in general means inherently having a right to something, as opposed to having been granted a privilege, the terms are sometimes used in Identity and Access to denote different levels of access rights in a hierarchy. Such usage is discouraged because it relies on subtle differences that are hard to understand, especially for non-native speakers of English. If the context of the situation requires such a hierarchy, it is better to explicitly describe and explain it than to depend on minute implicit differences in meanings.

Identification

Identification is listed here mainly to complement the words authentication and authorization*,* as it is a process related to those two terms but is often conflated with *authentication.* These processes may be implemented in very different ways depending on the context and requirements, so identification, authentication, and authorization are sometimes merged and implemented as one. But in another context, keeping identification separate and distinctly defined might be essential. In some contexts, there might not even be a need for identification at all (meaning there is no need for an identifier to be used in the *authentication* and *authorization* processes). That might be the case if the only requirement for granting access to a resource is that payment has been made. For the sake of completeness: Other sub-processes are also related to and often required by those discussed above, such as (identity) validation, proofing, vetting, etc.

Identity

First: *Identity is almost never a synonym for* just *identifier.* But the word is often used as if it were.

In almost every case*,* *identity* in our industry is shorthand for *digital identity.* It is often a representation of a real, natural person or something that acts like a person, such as a robot, or something that acts on behalf of a person, such as many Internet of Things (IoT) entities. Anything that requires *authorization* or *authentication* must have an *identity*, even though it does not always *have to* be reliably linked to an actual person. But what it means "to have" an *identity* in a specific context or situation is often not explained. And *identity* is often used interchangeably to mean different things that are not immediately apparent to the reader or audience. The difference often lies in the level of complexity intended for the given *identity* object. For example, sometimes *identity* needs to mean a very specific set of required data attributes that together—completely, for that given context—make up the *identity*. At other times, *identity* refers to a user object in a digital system, possibly including corresponding data attributes as well. And sometimes, *identity* is used for just referring to the *identifier* or *username* itself, without any notion of further complexity.

It may seem useful to have the word *identity* be so flexible, but when it switches back and forth between meanings, for example, "the person using the service" and "the user object representing them," readability suffers. The mix-up of *identity* and *user*—neither of which are clearly defined terms—is very common in the industry. At the time of writing, several examples can be found, for instance, in the Microsoft Azure AD documentation, if searching for both the words *identity* and *user* and seeing how they are used.[iii] Microsoft is no worse than any other party in this regard, unfortunately.

The NIST definition of identity also demonstrates this uncertainty about what it means to "have an identity."[iv] It states that a (digital) identity is "The set of physical and behavioral characteristics by which an individual is uniquely recognizable." The word "individual" here leads us to think about a person since there is nothing else following. Still, the definition is unclear about what it means for identities not directly intended to uniquely represent an actual, physical individual person.

It bears mentioning that there is no such thing as a "human digital identity" versus a "non-human digital identity." In digital systems, any identity is a digital object, which with varying degrees of certainty and through several layers of abstraction, might represent a real person, or it might not.

This problem of not distinguishing *identity* from *identifier* becomes even harder when using the widespread abbreviations ID and ident. These started out as shorthand for *identification* but now often mean either *identity* or *identifier* or both at the same time, making it easier to write a text about them but much harder to understand what it means. It is strongly advised to only use these abbreviations with clear guidance on their intended meaning in the given context. And whenever encountering them, it is advisable to investigate what exactly they stand for instead of guessing.

Owner

Most of the words in Identity and Access are used to *represent* something physical in the digital realm. As such, there is always the concept of relation and linking, which is often accompanied by the concept of ownership. A person may "own" an Identity, which may in turn "own" various user accounts/objects, which may, in turn, be assigned (ownership of) individual Access rights directly or grouped via Roles. In these cases, "ownership" and what it means will not be self-explanatory and needs to be clarified.

Permission

Permission is one of the common synonyms of access or access rights. In Identity and Access, permission has the same general meaning as entitlement and privilege (see below). However, it may also denote the lowest level in a hierarchy of access rights.

Person

For some vendors, *user* denotes the actual human accessing the service, while others use *person* for this. Others again do both at the same time. See User and Identity.

Privilege

As a synonym for entitlement, access rights, and so on, *privilege* is discussed above. In general usage, *privilege* is not a synonym for *right*, which is worth noting. Think of the sentence: "Education is a right, not a privilege." In Identity and Access, where *entitlement*, *access right,* and *privilege* represent further digital abstractions of something, such distinctions are seldom practical nor constructive.

Privilege in Identity and Access is associated with an even more common challenge. It is used both, as in the above, to denote any access right because any access right is a privilege granted. What causes confusion is when *privilege* is additionally used to mean *special access rights that imply an extra high level of privilege.* A whole specialty area of Identity and Access deals with such special access rights, including administrative access, access to sensitive information, accesses that can cause extra harm if misused, etc. This area has taken the name "Privileged Access Management," abbreviated PAM.

Where *privilege* sometimes refers to special access and sometimes to *any* access, it is advisable to make this distinction very clear by other means than just the word itself.

Along the same line, it may sometimes be better to use a different term, such as Higher Privilege Management or Higher Privilege Governance, for situations covering only a defined set of *special* access rights to emphasize the focus on special or "higher in importance than the others."

A related concept is the principle of "least privilege," used both in general in information security and risk management as well within the Zero Trust security model. Determining what constitutes the "least privilege" necessary for doing a particular job or task will also require being able to group and distinguish between different access rights (privileges) according to the corresponding risk.

Profile

Profile (and the similar *group* as used in Active Directory Security Group) is typically used for describing a collection of something, often a set of access rights or attributes about an entity. In Identity and Access, there is often no significant difference between using a group, profile, label, type, category, or a similar word to mean "a grouping" of, for example, access rights. But often the developer, the designer, or the author of the text had a distinction or special meaning in mind, so it is important to determine and describe what special characteristics or dependencies that specific grouping is intended to have.

Role

Role is often used to represent a grouping of something. This has become the general meaning of *role*. But *role* can generally group *anything* from individual access rights to people, tasks, and responsibilities. All these different meanings are relevant in Identity and Access, but exactly what things a given *role* groups, and under which rules, is rarely spelled out. When using the word *role,* it is almost always necessary to specify how that role is different from all other places the word is used. If, for example, using a term such as "Business role," "Technical role," or "Application role," always supply a precise definition for the term.

Token

There are many tokens in use in Identity and Access since most of the work relates to creating digital representations or symbols of something else—something which may also be abstract. Whether it is *explicitly* a token or just *implicitly* a representation—like a token can represent a valid and authenticated user, or an identity can represent an actual, physical person as well as a non-physical robot—the description of what is being represented and how cannot be implicit. Care must be taken to ensure that the reader or audience understands the relationship and what is represented by what.
An example from standards is the difference between a SAML token and a hardware token such as a FIDO security key. In NIST 800-63-3, the latter was changed to be an authenticator, and the former is still a token to help avoid confusion.

User

The meaning of the word *user* often overlaps with Identity and Person. It is often used to represent a person, such as the physical person who is meant to use a specific digital service, as well as simply representing the *identifier* or *username* of a digital object in the system. Without keeping these two meanings clearly apart, it will be hard for an audience to understand when it means one or the other. Another distinction that needs to be made clear for *user* is when it represents both internal *users* of software systems as well as external *users*. The former are often, but not always, administrative users or employees, and the latter are often, but not always, customers. Context determines the correct usage, but since the context is often not known, it needs to be specified.

The list above primarily aims to showcase the most common ways typical words in Identity and Access are used confusingly. Other lists aim to provide commonly used definitions - one of these is the *Terminology in the IDPro Body of Knowledge*, the list of words and terms used in articles of the Body of Knowledge, describing how they are used and maintained by the IDPro.

Causes and Consequences

An understanding of sources of ambiguity may be useful here, as this can make it easier to detect potential misunderstandings as well as manage their impact.

As noted above, Identity and Access have a language of their own. It is a language consisting of technical terms and abbreviations, but it also includes many *common* words that have taken on *special* meanings. These commonly known words comprise one such source of ambiguity. This organic growth of potential meaning stems from the fact that adding extra meanings to a word is much easier than taking it away. Consequently, the original meaning of the word is, for most people, still present in their minds. Unfortunately, they must also guess what exact interpretations have been added. The lack of a single, authoritative vocabulary for Identity and Access means that such extra meanings may and will diverge over time.

Whether one has learned these meanings from a list—found by searching on the Internet—or learned them from a mentor, colleagues, or presentations at conferences, they are valid in one or more specific contexts. If there are different possible contexts, there will also be multiple possible meanings.

One reason this is plaguing the area of Identity and Access is that this is an industry, not a discipline of science, and a young industry. It's an industry where practice is developing faster than standards and theory.

There is also the fact of multiple stakeholders. Identity and Access are relevant across various sectors (e.g., finance, healthcare, education, government), and each sector brings its own needs and interpretations to terms used in their environments.

But the stakeholder type probably most useful to be aware of is Marketing. For every term, technical or not, there is a risk that, in the end, "Marketing owns everything."[v]
No one has enough bandwidth to fight a battle for every term, so regarding which terms and concepts we find essential to retain ownership and definition power over, we must prioritize; we "have to choose our battles."[vi]

The relevance and viability of Identity and Access across sectors drive financial investment in vendors and products, resulting in companies' desire to put their stamp on Identity and Access terms and have their specific words correspond to their specific product or expertise. This desire leads to a multitude of competing words and/or meanings for terms like "privileged access rights" or "zero trust" or creates new terms overlapping with old, such as "IdM" vs. "IGA" vs. "CIEM," or "UEBA" vs. "ITDR."

The proliferation of such terms, created primarily to distinguish products from others, or attempt to take the name of a method or framework and connect it to a product, is something Identity Professionals get used to seeing over time. That does not mean it is necessarily a sustainable situation for the industry, and investigations into potential long-term solutions might be constructive to pursue. A discussion of potential longer-term solutions and the change the industry might go through is outside of the scope of this article.

On the other hand, it is within the scope to highlight the issue for the sake of better understanding and suggest how to approach the issue in the short term.

Short-term Solutions

To begin with, the best thing that Identity and Access practitioners can do is be *aware* that the terms used in the industry are confusing and ambiguous. When *hearing or reading* words of Identity and Access, this means:

- Continuously being aware of the problem.
- Setting aside time and patience for questioning.
- Questioning everything that:
    - Seems to have a different meaning than expected.
    - May have a meaning not immediately understood.
    - Seems ambiguous.

When *using*—writing or speaking—words of Identity and Access, awareness means consciously practicing clear and precise language.

- One important guideline is to always think about which of the chosen words may be understood differently if read or heard by persons from different backgrounds. If so, further explanation may be necessary. The list of words in this article is a good reference point for potential confusion.

- Consider whether a word is chosen because it can convey a fact or concept clearly, or whether it just looks good on paper.
- Imagine a theoretical difference between an identity *engineer* and an identity *evangelist*, the former needing to be unambiguous, the latter needing to be convincing.[vii]
- See also the note on Marketing above.

Create and maintain *local sources of truth (definitions)* where needed and when the universal terminologies do not precisely fit your local purpose.

- Use such lists to maintain a local authority to clarify in which context the meanings are valid.
- Try to keep the use of these words to only the intended local context.
- When it is necessary to collaborate with someone *outside* of the local area**:** Describe and explain the local context and purpose of the list.

In addition to the general awareness noted in the first bullet point, maintain an additional awareness of *specific words* within Identity and Access.

- These are regular words whose specific meanings get confused more often than others.
- See the Terminology section for examples of such words.

Conclusion

The specific context of a word is often unclear or unknown. And very few of these words and terms have exact, universally agreed-upon meanings. Consequently, unresolved debates about correctness or truth are common in Identity and Access. In many more cases, no one wants to admit that they are unsure about the meaning, and there isn't even room for a debate that might lead to resolution.

With so much opportunity for misunderstandings and miscommunication, the language of the industry is unnecessarily complex. This complexity hurts the recruitment and diversity efforts of the industry, as the impression individuals come away with is that one must be an expert in the field to participate. At the same time, there are no authoritative places to become an expert since the meanings are not universally agreed upon. And as practice develops faster than standards, individual actors in the industry tend to further develop standards in different ways, leading to competing versions. One example is the ISO 18013-5:2021 for Mobile driving license (mDL) application, where different vendors have been building solutions based on different draft versions of the standard.

Even being experienced and an 'insider' does not ensure correct understanding. Despite years of experience, individuals will find that words such as *user* or *identity* have multiple and contradicting meanings in a sentence.

There are potential solutions for this chaotic ambiguity of terms, some of which are immediately available and might be applied in the short term**.** Possible solutions for the long term**,** however**,** require more planning and coordination by the industry and affected parties.

In summary: The vocabulary of Identity and Access is vague and contradictory, and as such is not the best possible tool to build reliable Identity and Access solutions. It is a problem that only the smallest startup companies can ignore if they will never have any customers.

Awareness and carefulness around ambiguous words and terms—and knowledge about them—can help in the short term.

Author Bio

Espen Bago realized in 2002 that as system administrator, he'd been working in identity already for a while and decided from there to fully explore what this Identity thing was all about. He's been an independent Identity Advisor and coordinator to large enterprises for the last few years, but in 2021 became Identity Manager for the Norwegian Labour and Welfare Administration. As such, his goal is to make certain that identities – and the real persons this represents – are not forgotten when governments inevitably go all-in digital. He's also a founding member of IDPro and a member of the IDPro Body of Knowledge Committee and the IDPro Certification Committee.

[i] Based on conversations and questions about the issue in the IDPro Slack channels and in the industry in general.

[ii] Flanagan (Editor), H., (2022) "Terminology in the IDPro Body of Knowledge", *IDPro Body of Knowledge* 1(9). doi: https://doi.org/10.55621/idpro.41

[iii] https://docs.microsoft.com/en-us/azure/active-directory/fundamentals/active-directory-compare-azure-ad-to-ad

[iv] https://csrc.nist.gov/glossary/term/identity

[v] Vittorio Bertocci at the Identity at the Center podcast episode #167 - https://www.identityatthecenter.com/listen/episode/24656bde/167-2022-gartner-iam-summit-vittorio-bertocci-with-auth0

[vi] Ibid.

[vii] It’s sometimes necessary to do both at the same time—the point being made is to prioritize clarity over seeming convincing.

Authentication and Authorization (v2)

By Mark Morowczynski (Microsoft) and Michael Epping (Microsoft)

Updated by the IDPro Body of Knowledge Committee for v2

To comment on this article, please visit our GitHub repository and submit an issue.

Table of Contents

Abstract

This article describes the fundamentals of authentication and authorization, two core components of Identity and Access Management. It also delves into federation and Identity Providers, common tools for performing authentication and authorization in an organization.

Introduction

This article describes authentication and authorization, two core components to a sound Identity and Access Management strategy. Organizations typically have multiple tools that leverage authentication and authorization, both on-premises and in the cloud. The core concepts of each are described, and common ways authentication and authorization are used are explored.

Terminology

Many of these terms have been sourced from the "Terminology in the IDPro Body of Knowledge".[i]

Term	Definition
Access Control Lists	Access Control Lists are definitions around who or what are allowed or denied access to a resource. For example, a file share may have an Access Control List that allows Marketing Department users to read and write, IT Department users to read-only, and denies all other users' access.
Attribute-Based Access Control (ABAC)	A pattern of access control system involving dynamic definitions of permissions based on information ("attributes", or "claims"), such as job code, department, or group membership.
Authentication	Authentication is the process of proving that the user with a digital identity who is requesting access is the rightful owner of that identity. Depending on the use-case, an 'identity' may represent a human or a non-human entity; may be either individual or organizational; and may be verified in the real world to a varying degree, including not at all.
Authorization	Determining a user's rights to access functionality with a computer application and the level at which that access should be granted. In most cases, an 'authority' defines and grants access, but in some cases, access is granted because of inherent rights (like patient access to his/her own medical data). Authorization is evaluating what access or rights an identity should have in an environment.
Directory	A directory is a central repository for user identities and the attributes that make up those identities. A user identity might be John Smith

	with firstName attribute as John, lastName attribute as Smith, title attribute as Director, and Department attribute as Marketing. The attributes in the directory can be used to make authorization decisions about what this user should have access to in applications.
Identification	Uniquely establish a user of a system or application.
Identity Federation	An identity federation is a group of computing or network providers that agree to operate using standard protocols and trust agreements. In a Single Sign-On (SSO) scenario, identity federation occurs when an Identity Provider (IdP) and Service Provider (SP) agree to communicate via a specific, standard protocol. The enterprise user will log into the application using their credentials from the enterprise rather than creating new, specific credentials within the application. By using one set of credentials, users need to manage only one credential, credential issues (such as password resets) can be managed in one location, and applications can rely on the appropriate enterprise systems (such as the HR system) to be the source of truth for a user's status and affiliation. Identity federations can take several forms. In academia, multilateral federations, where a trusted third party manages the metadata of multiple IdPs and SPs, are fairly common.
Identity and Access Management	Identity and Access Management (IAM) is the discipline used to ensure the correct access is defined for the correct users to the correct resources for the correct reasons.
Identity Information Authority, aka Sources of "Truth"	This represents one or more data sources used by the IDM as the basis for the master set of principal/subject identity records. Each IIA may supply a subset of records and a subset of attributes. Sometimes the IIA is distinguished from the Identity Information Provider or IIP. We use IIA to include the service that actually provides the information as well as the root authority. This corresponds to Identity Information Source in ISO/IEC 24760-2 and Identity Sources in Internet2.
Identity Provider	An Identity Provider (IdP) performs a service that sends information about a user to an application. This information is typically held in a user store, so an identity provider will often take that information and transform it to be able to be passed to the service providers, AKA apps. The OASIS organization, which is responsible for the SAML specifications, defines an IdP as "A kind of SP that creates, maintains, and manages identity information for principals and provides principal authentication to other SPs within a federation, such as with web browser profiles."
Multi-Factor Authentication	An approach whereby a user's identity is validated to the trust level required according to a security policy for a resource being accessed using more than one factor (something you know (e.g., password),

	something you have (e.g., smartphone), something you are (e.g., fingerprint).
Relying Party	A component, system, or application that uses the IDP to identify its users. The RP has its own resources and logic. Note that the term 'relying service' is used in the ISO/IEC standards to encompass all types of components that use identity services, including systems, sub-systems, and applications, independent of the domain or operator. We will use the more common Relying Party (or RP). An RP roughly corresponds to the Agency Endpoint in the FICAM model or to Identity Consumers in the Internet2 model.
Role-based access control	A pattern of access control system involving sets of static, manual definitions of permissions assigned to "roles", which can be consistently and repeatably associated with users with common access needs. Role-based access control is a control scheme in which roles are granted to identities, and those roles determine what access to resources those identities should have. Basic roles might be "admin" and "read-only user" – an admin would be able to make changes to a system and a read-only user would only be able to view resources.

What is Identification

Identification is the act of determining which identity is in use or being interacted with by uniquely establishing a user of a system or application. Before an identity can be authenticated, it must be determined which identity is being used. A common way this occurs is through a user providing their username. The username is used to identify the user, while the password is used to authenticate the user.

Some types of credentials can provide both identification and authentication simultaneously, such as FIDO credentials or some biometrics.

What is Authentication?

Conceptually, authentication, sometimes abbreviated as AuthN, is the process of ensuring ownership of an account at the time the account is used to access a resource or establish a session. You complete authentication dozens of times a day and don't even realize it. When you log in to your computer with your username and password, you just did authentication. Then when you log in to check your email through a browser or an application like Outlook, you again authenticate to prove you own or are otherwise responsible for that email account. When you pick up your mobile device and use a biometric like a fingerprint or your face to unlock the device, you again complete authentication. If you go to the ATM to withdraw money, you first need to provide a card and then a PIN. If you successfully authenticate, then the ATM can trust that you are the owner of the account. Authentication can take different forms for different resources:

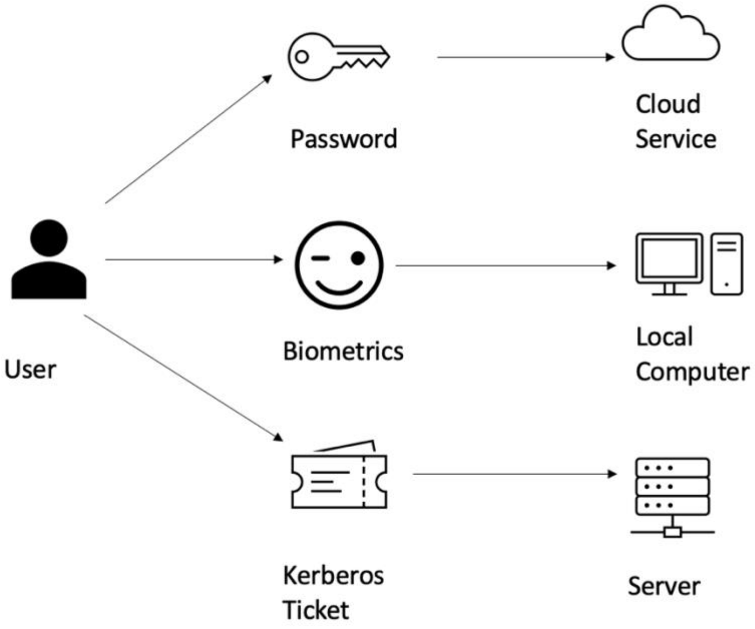


Figure 1: The user and their different authentication factors

There are many different possible authentication factors, such as memorized secrets, hardware tokens, and biometrics. These are often referred to as "something you know," "something you have," and "something you are." The most common factor is username and password. Also growing in popularity is the use of multifactor authentication methods. The most common is a text message or phone call, although these are no longer the strongest options available. There are also methods like One Time Passcode (OTP) software apps or hardware keys, where the password can be used only once and is usually valid for a limited duration. There are also different authenticator apps where a push notification is sent to the device and approved by the end-user. Physical FIDO2 security keys and biometrics like fingerprints and facial recognition are becoming more common and passkeys are rolling out to replace the standard password authentication ceremony entirely.[ii] Non-human identities also need to authenticate. Computers and services authenticate to each other using things like certificates, shared secrets (really just a password for an application), or other protocols developed for this purpose. Authentication, it's not just for people!

Authentication is often the first step when an entity wants to access a resource. We must first determine which identity is trying to access the resource and determine if it is the legitimate identity or an imposter. Then we can move on to the next step, determining what access, if any, should be granted or denied to the identity.

What is Authorization?

The next critical part of Identity and Access Management is authorization, sometimes abbreviated as AuthZ. Conceptually you can think of this as what an entity is allowed to do. Once the system or services knows who you are through authentication, you will be granted rights or permissions to do things through authorization. Authentication helps verify you are the same subject every time; authorization determines if you as the subject are allowed to access or do whatever action you are trying to do. These rights can be as simple as viewing a file (a grant permission) or denying the ability to view a file (a deny permission). You've probably experienced this when someone sent you a file or a link to a site and you received an "Access Denied" error message. You don't have the authorization to access that resource. You've also experienced this when you were able to view a file or access a site. There was just no message saying you were allowed to do it! You've probably come across hundreds if not thousands of authorization decisions a day and not even realized it (unless you get stopped, of course).

Authorization decisions can be made based on many factors. To start with a common one, if you have a specific role assigned to your account, you might have permissions in the system to add, modify, delete, or view things. This authorization architecture is commonly referred to as Role-Based Access Control (RBAC). For example, if you hold the role of administrator in a system, you might be able to manage all aspects of that system. Alternatively, if you hold the role of a reader in the system, then you may be able to view all the same things as the administrator, but you don't have the ability to make any changes.

Similarly, there are Attribute-Based Access Control (ABAC) systems where users may be granted specific rights depending on attributes on their account. For example, if you are a member of the sales organization, you would probably be a member of a sales group in your corporate directory or have a Department attribute set to "Sales". This group membership or attribute would grant you access to the sales shared network folder or a file-sharing site. But you wouldn't be able to access the engineering shared network folder or engineering site. Only those that were a member of the engineering department would be able to. These decisions are made typically by Access Control Lists (ACLs) determined by the system administrator. RBAC and ABAC are large topics unto themselves, deserving of their own articles.[iii]

Another example of authorization based on information about the user could be their job title. When a regular user logs into their HR application, they see information about themselves. How many hours they've worked, their manager, their pay stub, and information about their benefits. They are only authorized to view their own information. Their manager has a similar view about their own information but may also have additional information they can see about their employees. They can see all the hours worked for their direct employees but can't see that about other employees in the organization. Based on their title, they are only authorized to see that additional information about their direct reports. Finally, the head of HR might expect to see a wide range of information about the company. They might expect to see total hours worked for everyone in the company, total payroll, and benefits spent. Because they hold the title of Head of HR, they are authorized to see all this information.

Authorization applies to non-human accounts as well. A service account can hold roles in most directories. It would have the same permissions as any human account with that role. Service accounts can also be members of groups. A common example of this is the service that runs the backups on Windows servers. Depending on the design, it might require membership to a high privilege group, like Backup Operators, in order to backup and restore files on the system.[iv]

At this point, the concept of authorization should be clear and may seem straightforward. Authorization grants or denies permissions to various resources for both human and non-human accounts. However, the implementation details of this can be extremely complex. In our example above, the sales team and engineering team have access to separate corporate resources. But what do we do when they need to collaborate on something? Engineering has a new product coming out, and the sales team needs to be able to sell it. Do we add the sales team to the engineering group? Should we add the engineering team to the sales group? Or do we create a NEW group called Sales-Engineering and add the sales group and the engineering group to that new group? This addition of a new group might seem like the correct solution, but what do we do when the operations group also needs to work with engineering to ensure the production of the product meets engineering standards. Operations also need to work with sales to ensure the supply chain is aligned with their sales projections. Do we create more groups for all three teams to work together? As you can see, this starts to grow and get out of hand. Having an authorization design for these types of scenarios is important before you start implementing an Identity and Access Management (IAM) solution as well as how you will handle exception cases that will arise.

Lastly, we also need to make sure we are following the concept of least privilege when it comes to authorization. Least privilege is part of a robust strategy to ensure that users and service accounts only have the minimum permissions necessary to perform their. It is easy to grant more permissions such that things will work in an effort to make the authorization process simpler, but we'll pay the price later for those decisions, often in catastrophic ways. It's also often much more difficult to remove permissions from users and non-human accounts after they have been implemented. Take the time at the start to ensure least privilege is being followed for authorization decisions. Your future self will thank you.

The Role of Identity Providers and Federation

Both authentication and authorization may occur within a single system or application or may be externalized via an identity federation. If you have an application that doesn't reside on your corporate intranet (i.e., is a cloud-hosted service), your users will still need to authenticate.[v]

The identity provider, frequently abbreviated as IdP or IDP, handles the authentication of the user. The authentication can be via a web browser using forms-based authentication, integrated windows authentication (IWA), or an application using a web API. It's really user authentication as a service. There are common on-premises IdPs as well as cloud services that can be used as IdPs. These IdPs are commonly also doing some degree of authorization. Suppose a user is not able to authenticate to the IdP because they do not have an account or they do not have access assigned to a particular application. In that case, the IdP will not issue the user any assertion that can be used to access the application. If the user successfully authenticates, then the IdP issues assertions to the application/relying party.

Assertions, sometimes also referred to as claims, are pieces of information that are sent to the application/resource provider that, in this case, identifies the user and any additional information about the user that the application needs to function. These pieces of information are also referred to as attributes. The firstName attribute may be provided as an assertion and have values such as "John" or "Jane". The information requested and sent varies from application to application, but information such as title, manager, employee ID, etc., can be included in the assertion.

Before a user can authenticate and have information sent as an assertion to the application and access it, a federation trust needs to be set up.[vi] The setup details vary between federation protocols, but the IdP and the application will essentially exchange some information, such as the IdP public key and the application's endpoints for authentication. This information is typically in the metadata of the trust. Standards, such as FastFed, define how this metadata should be formatted to establish application and IdP trust.[vii]

Federation and IdPs allow us to control authentication and authorization for applications even outside the corporate network. These are important tools, especially in modern environments where cloud applications and services continue to proliferate. Organizations must be able to authenticate users, validate they are who they say they are, authorize them, and grant them the appropriate access based on who they are, everywhere – including on-premises and the cloud.

Conclusion

This document is a review of two core IAM concepts: authentication and authorization. These concepts are used in every organization to validate identities and grant those identities the appropriate access once they've been determined to be legitimate. Validating the legitimacy of an identity is crucial to keeping attackers out of organizations' systems. Granting the least permissions necessary to the identity is also recommended; it mitigates the damage if and when the wrong user or a compromised account accesses or has higher-than-necessary level of privilege in a system, thus reducing the blast radius of any nefarious actions as much as possible. Federation via Identity Providers (IdPs) is a common way to perform this authentication and authorization today, as applications and services are increasingly found outside corporate networks. Authentication and authorization techniques can protect these resources and identities regardless of location.

Author Bios

Michael Epping is a Program Manager in the Azure AD Engineering team at Microsoft. He is part of the customer experience team; his role is to accelerate the adoption of cloud services across enterprise customers. Michael helps customers deploy Azure AD features and capabilities via long-term engagements that can last years, as well as working within the engineering organization as an advocate on behalf of those customers. Michael has more than nine years of experience working with customers to deploy Microsoft products like Azure AD, Intune, and Office 365.

Mark Morowczynski (@markmorow) is a Principal Program Manager on the customer success team in the Microsoft Identity division. He spends most of his time working with customers on their deployments of Azure Active Directory. Previously he was Premier Field Engineer supporting Active Directory, Active Directory Federation Services, and Windows Client performance. He was also one of the founders of the AskPFEPlat blog. He has spoken at various industry events such as Black Hat 2019, Defcon Blue Team Village, GrayHat, several BSides, Microsoft Ignite, Microsoft Inspire, Microsoft MVP Summits, The Experts Conference (TEC), The Cloud Identity Summit, SANs Security Summits, and TechMentor. He can be frequently found on Twitter as @markmorow arguing about baseball and sometimes making funny gifs.

Change Log

Date	Change
2021-09-30	V1 published
2022-12-15	V2 published; terminology section expanded (ABAC, Identification, Identity Information Authority, Relying Party); included reference to passkeys; removed information on PKI

[i] "Terminology in the IDPro Body of Knowledge," IDPro Body of Knowledge, updated 30 September 2021, https://bok.idpro.org/article/id/41/.

[ii] For more information on FIDO2, see Fido Alliance, "FIDO2: WebAuthn & CTAP – Moving the World Beyond Passwords," website, https://fidoalliance.org/fido2/ (accessed 28 September 2021); ; for more information on passkeys, see FIDO Alliance, "Passkeys," website, https://fidoalliance.org/passkeys/ (accessed 10 November 2022).

[iii] For more information, see Koot, André, “Introduction to Access Control,” IDPro Body of Knowledge, 17 June 2020, https://bok.idpro.org/article/id/42/, and McKee, Mary, “Policy-Based Access Control,” IDPro Body of Knowledge, 19 April 2021, https://bok.idpro.org/article/id/61/.

[iv] For more information on managing non-human accounts, see Williamson, Graham and André Koot, “Non-human Account Management,” IDPro Body of Knowledge,30 October 2020, https://bok.idpro.org/article/id/52/.

[v] For more information on identity federations and sources of truth, see Lunney, Patrick, “Federation in the Enterprise,” IDPro Body of Knowledge, 19 April 2021, https://bok.idpro.org/article/id/62/, and Dingle, Pam, “Introduction to Identity - Part 2: Access Management,” IDPro Body of Knowledge, 17 June 2020, https://bok.idpro.org/article/id/45/.

[vi] For more information on IAM architectures, see Dobbs, G. B., (2021) “IAM Reference Architecture”, *IDPro Body of Knowledge* 1(6). doi: https://doi.org/10.55621/idpro.76

[vii] OpenID Foundation, Fast Federation (FastFed) Working Group, website, https://openid.net/wg/fastfed/ (accessed 31 August 2021).

The Business Case for IAM

By André Koot

To comment on this article, please visit our GitHub repository and submit an issue.

Table of Contents

Abstract

Businesses are under enormous pressure to deliver their products and services in ways that profit the company. Areas that do not directly bring in funding are often moved lower in priority, resulting in a competition for resources that can see internal projects in areas such as IAM struggle to succeed. Projects that move to the top of the priority pile in this competition are ones that provide a compelling business case. This article focuses on how to develop a positive business case for your IAM programs.

Introduction

Identity and Access Management (IAM) is often seen as one of many expenses that must be controlled within an organization. Businesses need to see the benefits of an IAM program before they are willing to invest in IAM programs. This circular demand can leave IAM improvements stuck in a never-ending game of catch-up. Businesses fail to see the strategic value in a solid IAM program until they see tactical improvements directly attributed to IAM services.

A solid business case helps break this deadlock by providing different perspectives on the overall Return On Investment (ROI) that IAM can bring to an organization. The best business cases include:

- the concept of the quantitative versus the qualitative components of the business case for IAM;
- the perspective from different IAM domains (e.g., internally facing IAM requirements from the enterprise, externally facing IAM requirements from the customers, cybersecurity requirements); and
- the recognition of the different strategic and operational requirements for both IT and the business.

Of course, different companies will respond better to different types of business cases. Some will be driven purely by the finances, while others will respond better by putting IAM in context with other services in an organization. Some may instead be primarily driven by the regulatory requirements governing their specific business operations (e.g., finance industry regulations).

Terminology

Term	Definition
Attribute-Based Access Control (ABAC)	Attribute-Based Access Control is a pattern of access control involving dynamic definitions of permissions based on information (“attributes” or “claims”), such as job code, department, or group membership.
Business to Business (B2B)	Business to Business processes in the field of IAM involve business partner access to company resources using some form of remote access (e.g., federated access).
Business to Consumer (B2C)	Business to Consumer processes in the field of IAM are customer or consumer access to company resources. In B2C, consumers manage their own identity in a CIAM. The company still manages access to the resources, using ABAC or PBAC methods for access control
Business to Employee (B2E)	Business to Employee, also called workforce IAM, includes managing identities and accounts for employees and contractors following an identity lifecycle.
Consumer Identity and	Consumer Identity and Access Management, or Customer Identity and Access Management, involves providing access to

Access Management (CIAM)	company resources through a digital identity managed by the customer.
Identity Governance and Administration (IGA)	Identity Governance and Administration is a discipline focusing on identity life cycle management and access control from an administrative perspective.
Joiner, Mover, and Leaver (JML)	The joiner/mover/leaver lifecycle of an employee identity considers three stages in the life cycle: joining the organization, moving within the organization, and leaving the organization.
Policy-Based Access Control (PBAC)	Policy-Based Access Control is a pattern of access control involving dynamic definitions of access permissions based on attributes (as in ABAC) and context for authorized access.
Privileged Access Management (PAM)	Privileged Access Management is a mechanism for managing temporary access for accounts with high-risk permissions. PAM often involves check-out and check-in of a credential generated for a single use.
Role-Based Access Control (RBAC)	Role-Based Access Control involves using roles at run-time to govern control access. It is a pattern of access control involving sets of static, manual definitions of permissions assigned to "roles," which can be consistently and repeatedly associated with users with common access needs.
Return on Investment (ROI)	Return on Investment is the economic measure of value of an investment, using costs, revenues, interest rates, and lifecycle as parameters.
Sunk cost	Expenses that have already been made in the past and that are unrecoverable.

Acronyms

C-level	Chief Executive Level, including Chief Executive Officer, Chief Financial Officer, Chief Information Officer, etc.
BC/DR	Business Continuity/Disaster Recovery
CI/CD	Continuous Integration/Continuous Deployment
GDPR	General Data Protection Regulation
HIPAA	Health Insurance Portability and Accountability Act
HR	Human Resources
IAM	Identity and Access Management
IT	Information Technology
ROI	Return on Investment
SSO	Single Sign-On
Y2K	Year 2000
ZTA	Zero Trust Authorization

Starting an IAM program

When working in IAM, the question often arises as to whether the costs and investments of an IAM program are worthwhile. Organizations generally ask for a financial business case since that is a traditional way to argue for an investment decision. It takes significant effort to convince decision-makers to look beyond the financial viewpoint.

Most IAM programs are started to solve one of three enterprise problems:

- Operations management (HR, IT)
  - for increasing employee efficiency, enhancing data quality, and cost-effectiveness[i]
- Enterprise, IT, or security architecture
  - for aligning with current best practices such as new controls for API access, support for Zero Trust, and supporting multi-factor authentication along with resolving issues of technology debt
  - for realizing newly defined strategic business initiatives, such as implementing a Consumer IAM (CIAM) strategy for revenue generation, improving customer services, and easing digital transformation[ii]
- Chief Executive Level (C-level)
  - for responding to audit findings in a management letter or directives from a supervisory agency or as the result of a security incident or data breach[iii]

Regardless of where the IAM program starts, a lot of money will be required from multiple cost centers before the program is complete. It often takes several budget cycles and significant organizational commitment to realize an effective IAM initiative. The program's sponsors must be prepared to make a business case to justify the organizational effort and the financial costs. Even if the C-level initiates the IAM efforts, a business case must often remind all stakeholders why this initiative is critical to the organization.

So, what elements of IAM investments can be identified that make an investment worthwhile?

This article looks at the business case for IAM from different perspectives.

- The first viewpoint is based on the difference between a business case's quantitative and qualitative components.
  - Quantitative means an objective calculation of the financial costs and benefits of an investment
  - Qualitative means that the costs and benefits of an investment cannot be calculated objectively, but the components have value for the business or bring additional trouble.
- A second viewpoint looks at IAM from different domains: B2E (i.e., Workforce IAM, Identity Governance and Administration (IGA)), B2C and B2B (i.e., CIAM), and Privileged Access Management (PAM).

- The third viewpoint is the organizational viewpoint: strategic, tactical, and operational reasons for implementing IAM.

There is no easy, complete formula for calculating the ROI of an investment in IAM. But at least these views can help to convince the stakeholders to look beyond the purely financial impact of IAM.

The Added Value of IAM

Preventing Negative Impacts

IAM strengthens businesses in many ways, from supporting business continuity to protecting business resources and reputation. For example, there are many reasons to consider IAM as a method for Business Continuity and Disaster Recovery (BC/DR). As organizations grow, access to resources becomes a liability: access to resources becomes more challenging overall, and delegating tasks and responsibilities becomes a bigger problem. If an organization is not in control of its data, including information on who may access that data, its ability to function in the case of significant business interruption is at risk. Even in a disaster, maintaining a record of who has in the past and can in the future access systems and data is critical.

Possibly the most famous example of a disaster directly related to a poorly managed and enforced IAM program is that of the Enron scandal in the late 1990s.[iv] In IAM terms, the scandal was partly a result of executives circumventing management controls, possibly because of the lack of fitting access controls. The best practice of Segregation of Duties was circumvented by greed, organizational culture, and practices.

At the same time, investments in IAM suffer from the prevention paradox. Investing in IAM rarely brings immediate, visible improvements. Finding the benefits (in terms of concrete cost savings) may be hard to achieve. Would the effects be the same with fewer costs and efforts of the IAM investment? It may seem like the Y2K crisis all over again.[v]

Supporting Positive Impacts

Of course, not every organization suffers from the same malicious drivers as Enron did. Still, that case highlights the need for access control from the perspectives of business continuity, governance, and compliance. To be in control, the need for managing identities and especially the management of authorizations is demonstrated by this case and many others.

But there are more reasons for investing in IAM. In many organizations, the need for IAM comes from the need for efficiency and high data quality. Manually creating identities for personnel and adding and revoking authorizations is inefficient, while the manual execution of these tasks can result in a lack of data quality. Automating the process can provide higher quality with less expensive results.

Organizations will also find benefits in improving their user experience. Developing single sign-on (SSO) services and self-service access requests improves not just the efficiency of the process but also the user's satisfaction. The continuing development of external access and the move toward API access led to the need for IAM-related programs.

These lines of reasoning, while valid, may be too far away from daily operations and immediate, visible improvements in efficiency. Businesses and boards experience the need for short-term insight as well as long-term improvements before they make further investments. In other words, companies need a specific and complete business case for IAM.

Different Dimensions of IAM Business Case

Quantitative versus Qualitative Business Case

A simple formula for calculating the ROI looks thus:

$$ROI = NetReturnonInvestment\ /Cost\ of Investment * 100\%$$

(see Guide to calculating ROI)

The simple formula can be used to calculate if an investment is anything good, financially speaking. Suppose you want to invest 1 million and later sell the investment for 1.1 million, resulting in a profit of 100,000; the ROI would then be
(1,100000 – 1,000,000) / 1,000,000 * 100% = 10%

Of course, the calculation would be a little more complex for projects. It is unlikely that you would invest 1 million in IAM and later sell the investment, making a profit. This simple formula also hides the fact that indirect returns are both critical for the overall measure of ROI and extremely hard to quantify.

First, let's look at the distinction between the quantitative business case and the qualitative business case.

- The *Quantitative Business Case* is all about money. It is about calculating the costs and benefits objectively, at least as much as possible. This enumeration is relevant for managers to calculate all investments in an organization to prioritize investments. To a lesser degree, the business case can be input for a cash flow analysis. In this article, we classify topics as objectively quantifiable, but just like in risk management, some entries cannot be calculated objectively. For example, the risk of penalties does not result in an absolute value. It is an approximation of the cost of the risk. The cost of the risk could be calculated as the chance of discovery of the non-compliance times the potential maximal amount of a fine.

That means that, just like any approximation, it has to be taken with a grain of salt.

- For most governance, risk, and compliance managers, the *Qualitative Business Case* will be the preferred justification for investments in IAM solutions. The entries in the overview below may not be objectively quantifiable, but that does not mean that they should not be considered when prioritizing investments.
- An interesting example of a financial business case is the situation of banks and insurance companies who have to undergo a stress test to find if they can survive a financial crisis. The capital requirements are higher or lower depending on the risk level. High capital requirements impact the money-making capabilities; a high reserve is a lot of unused capital. These rules and regulations have been defined in the EU Basel IV and Solvency 2 regulations, which have also been adopted by the Federal Reserve in the US.[vi]
  If a bank has sufficient assurance about authorizations because of adequate access control, then data quality will be better, and risk (uncertainty about access) and capital requirements will be lower, resulting in a significant impact on revenue creation capabilities.

The Business Case for Different IAM domains: IGA, PAM, and CIAM

Another view is that the business case for different types of IAM-related programs may have different focal points because they focus on different things. For example:

- Identity Governance & Administration (IGA), which focuses on the internal account and authorization management for employees and contractors with enterprise access, has a root cause in automation, efficiency of performing JML processes, and assigning and revoking roles. While IGA investment decisions will benefit from a quantitative approach to the business case, a purely quantitative approach will not be enough to make the case. Costs and benefits will probably lie in different cost centers that measure success in different ways (e.g., in improved efficiency, in lower risk to security, in regulatory compliance). So, unless the business case is calculated companywide, the business case will be negative.
- Privileged Access Management (PAM) is all about managing risks of critical authorizations and remote access for internal accounts with broad access to sensitive resources. Its focus lies in governance and compliance. In this case, the business case is more likely to start off with a qualitative focus and miss out on some of the critical quantitative aspects that will strengthen the argument. The business case will be qualitative at first sight, but a secondary point of view may be limiting the risk of penalties and fines from laws and regulations.
- CIAM (used for B2C and B2B connections and also applicable for IoT and OT access) focuses on self-service identity management of consumers or customers. That moves convenience and consumer appreciation into a competitive advantage. The quantitative approach may not be sufficient; business continuity may be at risk for lack of investment.

This means that the business drivers for these domains are different and that the business case will contain other components.

Strategic, Tactical, and Operational Viewpoints

The third way of looking at the concept of the business case is the organization's viewpoint. In traditional organizational theory models (e.g., the Anthony triangle[vii]), we can identify the strategic, tactical, and operational layers. And if we follow up on these separate layers, there are also strategic, tactical, and operational considerations for implementing IAM:

Strategic

This topic is all about implementing business governance of Access, putting the business in control of IAM, and taking IAM out of the realm of IT. The underlying principles are:

- Governance Risk and Compliance: to be able to show that the organization is in control, to be compliant with laws and regulations, and to prevent 'Enron' issues.
- Competitor initiatives, competitive advantage: either to follow industry best practices (for example, a competitor implemented IGA) or to lead the market (for example, by implementing a leading CIAM platform).

These issues can also be seen as qualitative components in the business case.

Tactical

The tactical drivers may include enhancing business processes and information flows, structuring the organization to be more agile, and supporting merger and acquisition processes. But another driver could be to reduce technical debt that prevents innovation and agility. Older identity management solutions that are end-of-support or do not scale well to the cloud should be replaced.

The tactical components can be both quantitative and qualitative.

Operational

Operational considerations are related to the effectiveness and efficiency of people, processes, and technology. The automation of manual processes, increasing efficiency through self-service activities, and improving user experience are relevant topics for the business case.

These manual processes can be automated:

- User account management - In the JML processes, the workflow and the lifecycle can be automated based on transactions in the source system for identities (HR, student management, customer relationship management (CRM), etc.). For example, when Role-Based Access Control (RBAC) is implemented, granting and revoking of roles can also be automated. So, user and account management, as well as role management, can be automated, resulting in less manual work, faster processing, better data quality, and cost savings.

- Password reset – establishing a self-service mechanism for password resets increases user satisfaction and customer service efficiency.
- Reporting, certification, and attestation processes - these can be automated, resulting in more transparency.
- Data processing disclosure - Informing customers about the processing of their data can be automated in CIAM portals.
- Single Sign-on (SSO) – SSO enhances user convenience and reduces all kinds of service desk-related calls.
- Automated logging and auditing – Automated logging will facilitate security operations and forensic readiness.

Many of the operational issues can be regarded and calculated as quantitative components in the business case.

One Invalid View

One argument for not investing in IAM is the notion that an organization may have already invested heavily in IAM solutions, resulting in capital expenses that have not yet been written off.

This is not how an organization should react to an identified need for change. Costs based on decisions in the past should not be used in future decision processes; past decisions would lead to lock-in or in-agility for keeping up with the old choices. This kind of reasoning is referred to as the 'sunk cost fallacy' where people as well as organizations often continue with an action even as the costs outweigh the benefits.[viii] A useful counterargument to combat this fallacy is that, in hindsight, individuals would make different decisions for their organization.

Overview of Business Case Topics

This section offers an overview of the different components of the business case for IAM. It is by no means a complete overview, but it gives an indication of arguments for convincing anyone of the positive effects of investing in an IAM program. The tables suggest both the quantitative and the qualitative components of the business case for each of the three example domains: IGA, CIAM, and PAM. These examples can act as templates for other domains; practitioners will need to adapt the specifics to suit their own organizations and use cases. The strategic, tactical, and operational components can be recognized as components in the qualitative and quantitative columns of the tables.

The first table shows the components of the business case for Identity Governance and Administration (automating JML and implementing RBAC). In this table, both positive (green background) and negative (red background) components of both the quantitative aspects (left column) and qualitative aspects (right column) of the business case are explained.

The consecutive tables show comparable topics for both CIAM programs and PAM programs.

The negative financial components (investments, licenses, costs) are comparable for all three domains.

A basic cost savings formula is shown for some of the financial and quantitative components as guidance. It will, however, be meaningless without a good explanation of the benefits.

Business case considerations for Identity Governance and Administration programs

Quantitative Business Case: $, €, etc.	**Qualitative Business Case**
● Benefits: Cost reduction ○ Reducing manual tasks within the JML processes ■ Self-Service password reset ● Typically, a password ticket amounts to > $25 each. Implementing self-service password reset would save that workload. The net result will probably be less since service desk agents hardly ever are dedicated password reset employees. If, however, the service desk is outsourced, savings on out-of-pocket costs will be big. ***Formula***: saving = #password resets * (ticket price + (#minutes waiting for reset * hourly rate)) ■ Access Request management ● Automating access requests by removing them from service would save ticket costs but also the costs of manual handling of the process, both at the service desk and for application administrators and line managers. ● ***Formula***: savings = #requests * #cost per transaction (manual time * hourly rate)	● Benefits: Better Governance Risk and Compliance ○ Legal and regulatory obligations ■ Laws ● Laws and regulations result in controls that can be implemented and enforced by IAM solutions. ■ NIST / ISO standard compliance ● NIST and ISO standards and architecture patterns can be integrated with IAM solutions ■ Better compliance with Export Control regulations ○ Managerial Insight ■ Attestation, (re)certification ○ Supporting Organizational Agility ■ Mergers & Acquisitions, Due Diligence ■ Restructuring ○ Access Governance ■ Roles and rules ● By implementing roles and rules, the authorization models can be formalized and automated. This will reduce the level of ad-hoc access management and enhance the level of control of access ■ Reports

        - Provisioning
            - Provisioning of accounts, roles, and authorizations will save a large amount of manual labor by system and application administrators. Using birthright roles (roles that can be granted automatically based on department or manager), the performance can be impacted even more positively. The same is true for de-provisioning.
            - ***Formula*** (per connector): saving = #accounts * $cost per transaction (manual time * hourly rate)
        - Reduced costs of remediation of lack of data quality caused by manual data entry and lack of correlation between different identity repositories.
            - ***Formula***: savings = data entry *error rate* (circa 5-10%) * #accounts * $cost per transaction (manual time * hourly rate)
    - Reducing Cost of Compliance
        - Attestation
            - Automating the certification process saves all manual verification of accounts and authorization. Lowering administrator efforts to create the reports and views and lowering manual verification by managers.
            - ***Formula***: savings = #reports * $cost per analysis (manual time * hourly rate)
        - Audit reports
            - IGA solutions typically have dozens of specific IAM-related reports that can be ordered from the self-service portals without assistance from the IT department or Business Intelligence experts.
        - Ownership
            - In Access Governance, multiple stakeholders are responsible for defining access decisions. By implementing roles and rules, as well as workflows, the ownership will be implemented by default. Otherwise, no access rules can be defined. Accountability will result.
- Implementing an Access Control scheme (e.g., RBAC, ABAC, PBAC, etc)
    - Popular access control schemes offer methods for defining access policies. In order to do implement these properly, IGA needs to be in place.
- Adding quality of service by moving responsibility for access control to the business from IT
    - Traditionally, IAM is a responsibility of the IT department. And that means that the ‘business’ is a victim of the SLA with the IT dept. By moving the responsibility and execution to the business, the burden of IT processes for the business is lowered. It does, however, imply that the burden now rests at the business level.
- User Convenience
    - Self-service
    - Faster processing, less idle time

● Auditors require reports. In some cases, they run their own reports (requiring specific authorizations, requiring additional governance) and analyze all results. Data drive audits are expensive. IGA solution can provide an auditor portal to use the available data. A process-oriented audit is more cost-efficient than a data-based audit. ● ***Formula***: savings = #reports * $cost per analysis (manual time of external auditor * hourly rate of external auditor + manual time of administrator * hourly rate of administrator) ■ Portals ● Using the workflow, engines-based self-service portals of IGA solutions are more cost-efficient than having data scientists or IT personnel generate reports for different stakeholders. ● ***Formula***: savings = #reports * $cost per analysis (manual time * hourly rate) ○ Lower License costs ■ Software licenses are typically user-based. In manual deprovisioning processes, removing licenses is not always performed, resulting in unused licenses. When using automated workflows for Moving and Off-boarding, deprovisioning can be used to remove licenses from user accounts. ○ Lower idle costs: Automation leads to faster processing	○ SSO ● Reducing Technical debt ○ Replacing old technology (lack of development, end-of-support type of software) with modern solutions ○ Preparing for cloud enablement

    - In manual (de)provisioning, the workflow will generally take much longer for transport time, waiting time, and idle time. Depending on the request type, this may be blocking personnel from performing actual work.

- Positive: Reducing the risk of fines and penalties
  - Fines and Penalties can occur when an organization is not in control and not compliant. By lowering the risk of non-compliance, the risk of fines and penalties will also be reduced. This may not be a financial business case, but lowering the risk will also be beneficial in accounting terms. Lower risks will also mean lower capital requirements, lowering the capital reserves and unused capital requirements.
  - Reduction of risk of data breaches
    - Privacy, GDPR, HIPAA, etc.
      - If there is more assurance about the granted access, and if the (re)certification/attestation is implemented in an effective way, the risk of incorrect authorizations is lower, and so the risk of fines will be lower.
    - Risk of negative impact on Brand value
      - Data breaches and security incidents can (in the short term) have a negative impact on brand value or stock value for listed companies. If the risk of data breaches is reduced, the risk of lower value is also reduced.

○ Reducing compliance penalty risks ■ The risk of security incidents can be reduced by implementing security controls at the user level, like Segregation of Duties as required in various laws and regulations for high-risk business processes. ***Formula***: savings = (percentage of chance of discovery) * (max fine for non-compliancy) ○ Reducing Basel4 / Solvency2 cost risks ■ If financial institutions can lower their capital requirements, their costs will be reduced, and income will rise accordingly. ● Positive: Better business reputation ○ Increasing Consumer Confidence (e.g., data is kept secure, not shared with others without consent; organizations have the ability to let the consumer know who has accessed their data; consumers have the ability to opt-out, etc.)	
● Costs: Investment ○ Cost of the program, architecture, design, procurement ■ Before starting IAM programs, lots of analysis will be made, architectures and designs, and other overhead costs, like procurement and tendering costs. These costs may not be assigned to one specific project, but the costs cannot be neglected. ○ Licenses, maintenance, and support costs (the latter for open source) ■ Most IGA software solutions come from commercial	● Costs: Ways of working ○ (Sentiment of) reduced autonomy/sovereignty for impacted business units ■ If businesses are organized in some federated way, and each dept has a degree of autonomy, the implementation of a central IGA solution may feel like impacting the autonomy of a dept. This sentiment should, of course, be reduced by pointing to the configurable access policies, workflows, and reports of modern IGA solutions.

vendors. There is a limited number of open-source products.

License fees are usually based on the number of users. On-premises solutions require an investment fee with an annual maintenance or support fee.
SAAS Cloud products are usually subscription-based.

Additional costs may occur because of

- adding/developing/configuring connectors to source and target systems
- training and certification courses.

- ○ Cost of Implementation
  - ■ IGA solutions will be implemented by an integration partner (who also usually sells the licenses for IGA). Implementation costs can be high, depending on the level of customization. Even simple configuration changes can be hard, but custom code should be avoided as much as possible. Custom code results in lock-ins, making upgrades hard and even more expensive.
  - ■ The cost of implementation can be high if the proposal leaves too many loose ends: ask the following default pricing for an IGA implementation, with one source system (HR), two target systems (AD + one DBMS connected system),

implementation of the JML workflow, and attestation report. Do not implement RBAC (incl. mover workflow) from the start of an IGA project; authorization management is too complex for full-fledged RBAC. Begin with just a few birthright roles and only start using RBAC when governance is in place. ○ Operational costs ■ Additional costs of managing the IGA solution, modeling roles and workflows, performing authorization management tasks ■ Moving decentralized (almost unidentifiable costs) JML processes to a central solution, so additional central costs, paid for by decentralized saving (this should be at least budget neutral, or could potentially lead to big cost savings, but dept versus corp makes a difference)	

The business case for Privileged Access Management programs	
Quantitative Business Case	**Qualitative Business Case**
● Benefits: Cost reduction ○ Consolidation of password management solutions ■ ***Formula***: savings = #accounts * license fee (for every password manager)	● Benefits: Governance, risk, and compliance ○ Better Governance Risk and Compliance ■ Reducing anonymous access to critical accounts ● MFA for critical access

- Consolidation of remote access solutions
    - PAM solutions, by default, have good remote access capabilities. This may include admin login and authentication, incl. MFA, secure routing, (SSL) VPN, logging and monitoring, and session recording. Most PAM solutions can replace different remote access facilities in both IT and OT and can even replace vendor/supplier remote access. Thereby reducing the costs of multiple point solutions, including maintenance and support.
        - ***Formula***: savings = (license fee + maintenance costs) (for every password manager)
    - By using the monitored PAM solution, vendors and suppliers can manage their own access without requesting (remote) access from a service desk officer.
        - ***Formula***: savings = # remote access request * administrator rate
- Password management
    - Lower operational costs by admins to secure, rotate, and manage passwords and tokens for privileged accounts.
- Reducing compliance penalty risks
    - The risk of security incidents can be reduced by implementing a PAM solution.
      Formula: savings (percentage of chance of

        - Password rotation and vaulting
        - Session Recording
    - More insight into the usage of critical accounts
    - Connection between administration and service tickets
- User convenience
    - SSO for admins
    - Remote Access for admins
        - offering MFA for non-personal accounts
    - Remote access for vendors and suppliers
        - including risk-based session recording, MFA and monitoring and logging of events

discovery) * (max fine for non-compliance)	
● Costs: Financial ● See B2E for similar costs	● Costs: Ways of working ○ (sentiment of) reduced autonomy, loss of divine powers of administrators

Business Case Considerations for Consumer Identity and Access Management Programs (B2C, B2B)	
Quantitative Business Case	**Qualitative Business Case**
● Benefits: Cost reduction ○ Manual tasks for JML ■ Self-Service Identity Management for external identities, reducing the manual tasks connected to identity management, including password reset ***Formula***: savings = #accounts * (manual cost per task)	● Benefits: Business agility ○ A competitive advantage when building portals ○ Supporting Organizational Agility ■ B2B and Remote Access ○ Support innovation ■ DevOps, Continuous Integration/Continuous Deployment (CI/CD), Zero Trust Authorization (ZTA), API access ○ Access Control and Access Governance ■ Policy-Based Access Control (PBAC), Attribute-Based Access Control (ABAC)

	○ User Convenience ■ Self-service ■ SSO ■ MFA ○ Scalability ■ Federative Access ■ Scalable to access APIs and microservices
● Costs: Finance ● See B2E for similar costs	● Costs: Way of working ○ (sentiment of) loss of autonomy of customers, victimization due to privacy risks

Closing Thoughts About the Business Case

As explained before, a short-term positive real quantifiable business case can hardly ever be achieved. For instance, the real benefits of automating the JML flow with RBAC will only be apparent after several years, after adding multiple target systems across multiple lines of business, thus generating more business value. When looking through one-year project glasses, the outcome will not be financially interesting enough. IAM cannot just be seen from a financial perspective; there are many more considerations to be taken into account.

Pay attention to the following:
The issue of just focusing on the financial business case is too restrictive, more so when the investing stakeholder Is not the stakeholder who benefits from the investment—as is often the case. In many cases, the IT department is the cost center funding the investment. But, as can be seen in the business case examples, other departments profit from the investment in IAM. It is therefore essential to identify all stakeholders and the advantages they gain from the investment in IAM solutions, even if these benefits are not financial.

A second topic that should not be ignored in the financial savings area. Many manual activities are 'hidden' costs, including when users request access, and managers review existing authorizations, approve new requests, create accounts, and grant permissions or roles. These activities disappear in the 'normal', daily tasks of employees and so often go unaccounted for. By automating these tasks, employees can focus on more valuable activities. In the financial business case, quantifying this element may be an unwanted eye-opener.

Considering a multi-faceted business case for IAM is essential for every IAM program. A business case that goes beyond financial considerations will build awareness and commitment for starting a multi-year program that adds value to long-term business continuity. Approval is nice, but do not make it depend on a financial business case only.

Acknowledgments

The author wishes to thank Robert Sherwood and IDPro Principal Editor Heather Flanagan for their reviews and assistance in writing this article, turning thoughts into words.

Additional Reading

Beattie, Andrew. 2022. "How to Calculate Return on Investment (ROI)." *Investopedia*, August. https://www.investopedia.com/articles/basics/10/guide-to-calculating-roi.asp.

Azmi, A. M. (2007). Business cases for information technology projects. Paper presented at PMI® Global Congress 2007—EMEA, Budapest, Hungary. Newtown Square, PA: Project Management Institute.

James Cook University (14th February 2020). How to Write a Business Case: Tips, Resources and Examples.

Wikipedia contributors, "Business case," *Wikipedia, The Free Encyclopedia,* https://en.wikipedia.org/w/index.php?title=Business_case&oldid=1164376316 (accessed October 17, 2023).

Author Bio

André Koot is principal consultant at and co-founder of SonicBee, a Dutch IAM consultancy company (IDPro partner), focused on business consultancy and giving IAM training courses. He is also a member of the IDPro BoK committee and (co-)authored several articles in the BoK.

[i] Schueler, Chris. 2022. "Neglecting The IAM Process Is Fighting A Losing Battle To Achieve Operational Excellence." Forbes, April 8, 2022. https://www.forbes.com/sites/forbestechcouncil/2022/04/08/neglecting-the-iam-process-is-fighting-a-losing-battle-to-achieve-operational-excellence/?sh=2a6b16147977.

[ii] "Manage Technology Debt to Create Technology Wealth." Gartner. August 17, 2020. https://www.gartner.com/en/documents/3989188.

[iii] Shea, Sharon. "How IAM Systems Support Compliance." *Security*, July 2020. https://www.techtarget.com/searchsecurity/tip/Identity-management-compliance-How-IAM-systems-support-compliance.

[iv] Hayes, Adam. "What Was Enron? What Happened and Who Was Responsible." Investopedia, March 2023. https://www.investopedia.com/terms/e/enron.asp.

[v] Allen, Frederick E. 2019. "Apocalypse Then: When Y2K Didn't Lead To The End Of Civilization." *Forbes*, December 29, 2019. https://www.forbes.com/sites/frederickallen/2020/12/29/apocalypse-then-when-y2k-didnt-lead-to-the-end-of-civilization/?sh=6c4625dc475c.

[vi] "Basel IV Implementation in the EU: What Does the New Banking Package Mean for Banks?" 2022. Oxford Law Blogs. February 3, 2022. https://blogs.law.ox.ac.uk/business-law-blog/blog/2022/02/basel-iv-implementation-eu-what-does-new-banking-package-mean-banks and "Solvency II." n.d. European Insurance and Occupational Pensions Authority. https://www.eiopa.europa.eu/browse/regulation-and-policy/solvency-ii_en.

[vii] Larson, Theodore, and Daniel Friesen. n.d. "The Anthony Triangle and an Analytics Framework: Developing a Business Analytics Curriculum Conceptual Model." *CERN European Organization for Nuclear Research*. December 2020. https://doi.org/10.5281/zenodo.3996830.

[viii] "The Sunk Cost Fallacy - The Decision Lab." n.d. The Decision Lab. https://thedecisionlab.com/biases/the-sunk-cost-fallacy.

Consumer / Citizen IAM

Introduction to Customer Identity and Access Management

By Ian Glazer

Table of Contents

Abstract

Customer Identity and Access Management (CIAM) refers to the processes and technologies that facilitate secure interactions between individuals and organizations. In particular, this article focuses on those that secure digital interactions. Whether the organization is in the public or private sector, the need to interact digitally is essential in this day and age – regardless of whether those interactions are to transact commercially, access social services, attend an online class, etc. While CIAM shares some concepts and technologies with workforce IAM, the two are sufficiently distinct to warrant further investigation. This article compares and contrasts the two while highlighting the unique challenges and opportunities inherent to CIAM.

Introduction

Customer Identity and Access Management (CIAM) represents one of the most notable opportunities for identity professionals to shine. Through CIAM, identity professionals can help organizations reduce costs and reach new customers. For commercial entities, this means growing both the top and bottom lines. With these wide-ranging opportunities, CIAM is different from workforce IAM. CIAM presents IAM professionals with new challenges, vocabularies, processes, and requirements – all of which serve to ensure that individuals can interact with organizations easily and securely.

Terminology

Many of these terms have been sourced from "Terminology in the IDPro Body of Knowledge."[1]

Term	Definition
Authentication	Authentication is the process of proving that the user with a digital identity who is requesting access is the rightful owner of that identity. Depending on the use-case, an 'identity' may represent a human or a non-human entity; may be either individual or organizational; and may be verified in the real world to a varying degree, including not at all.[2]
Authenticator	The means used to confirm the identity of a user, processor, or device, such as a password, a one-time pin, or a smart card.[3]
Authoritative Source	The system of record (SOR) for identity data; an organization may have more than one Authoritative Source of data in their environment.[4]
Authorization	Determining a user's rights to access functionality or resources within a computer application and the level at which that access should be granted. In most cases, an 'authority' defines and grants access, but in some cases, access is granted because of inherent rights (like patient access to their own medical data).[5]
Consent	Permission for something to happen or agreement to do something.[6]
Customer Identity and Access Management (CIAM)	CIAM is the field of IAM that focuses on the Registration, Authentication, and Authorization services for an individual or entity receiving or purchasing services from an organization.
Credentials	In the context of CIAM, credentials are how individuals authenticate themselves to an organization's CIAM system
Credential Stuffing	An attack in which an adversary tests lists of username and password pairs against a given CIAM system.
Identification	Uniquely establish a user of a system or application.

Identifier	An identifier is a means by which a system refers to a record (at the most abstract levels.) In this case, it could mean the string that a person provides that “names” their use account.
Lifecycle	In the context of CIAM, lifecycle refers to the stages that an individual or entity might experience over the course of their relationship with an organization, beginning with the formation of a relationship (such as being hired into an organization or signing up for service) and ending with the severance of that relationship (such as termination or closing an account)
Passwordless	Any means of authenticating a user account that does not require a static stored shared secret. Techniques include one-time passwords and passkeys.
Policy Store	A repository that houses configuration information for the CIAM system and serves as an Authoritative Source for that information. For example, OAuth token Lifecycle policies or Authorization policies.
Preferences	Choices that individuals or entities make in administering the relationship they have with an organization. These choices may include topics of interest or approved communication methods. Often, Preferences are stored with Profile information.
Profile	A collection of attributes about an individual. The individual may provide it directly, or the organization may gather it indirectly.
Progressive Profiling	A technique to reduce customer friction by gathering Profile, preference, and Consent information over time (when needed) rather than all at once.
Registration	The creation of a relationship between an individual and an online system that is initiated by the individual and results in the creation of a user account or Profile.
Workforce IAM	The application of IAM sub-disciplines such as access governance, authentication, and Authorization for employees as opposed to the applications of such disciplines for customers.

Acronyms

ATO	Account Takeover
B2B	Business-to-Business
B2C	Business-to-Consumer
B2B2C	Business-to-Business-to-Consumer
CIAM	Customer Identity and Access Management
CRM	Customer Relationship Management
DAU	Daily Active Users

IAM	Identity and Access Management
IDP	Identity Provider
JML	Joiner, Mover, Leaver (used in Workforce IAM)
MAU	Monthly Active Users
OTP	One-Time Password (or Passcode)

What is CIAM?

Over the last decade, organizations in every industry, sector, and geography have sought to provide services online. Trading under the name "digital transformation" and "digital engagement," organizations have pushed to interact with people through websites, mobile apps, and connected devices to reach new customers, offer more valuable services, and lower service delivery costs. The COVID-19 pandemic further amplified the need for all organizations to have a robust online presence.

But for people to interact with these online services, they need a means to safely and efficiently identify themselves to those services. How organizations offer sign-up and sign-in services is the core of CIAM.

What Does the "C" Stand for?

Digital identity practitioners love abbreviations, but this can cause confusion. The C in CIAM is just such an example. Despite the assertions of Sesame Street's Cookie Monster,[7] C stands for more than just cookie: in this context, it also stands for customer, consumer, or citizen. The typical usage is customer, but that may have inaccurate implications. For example, it may imply that CIAM systems only apply to contexts in which the individual pays for a service from an organization. This is not the case.

All organizations need CIAM to interact with people who could or do use their services. Such organizations include public sector agencies that deliver on behalf of citizens and residents, universities that empower students and researchers, and non-profits that serve communities and engage with supporters. And yes, this also includes for-profit businesses that sell goods and services.

Why CIAM is Important

CIAM enables organizations to reach more people and offer more valuable services. In widening reach, CIAM provides a way for organizations to expand their total addressable market while reducing service delivery costs. As a result, an effective CIAM program improves both the top line and the bottom line of organizations. These benefits are equally relevant for public sector entities that aim to reach more citizens, deliver more social services, and reduce taxpayer costs. While traditional workforce IAM is an essential cost-center focused on efficiency, security, and compliance requirements, CIAM can be seen as a profit-center.

How CIAM Differs from Workforce IAM

Some readers may be more familiar with the primary goal of workforce IAM to deliver the right access to the right people at the right place and time. To meet this goal, IAM practitioners deploy, for example, automated user provisioning, birthright policies triggered by a small number of central authorities, access request systems, and authorization policies governed by a central Identity Provider (IDP).

CIAM has a different goal. It supports organizational digital engagement efforts to deliver the right experience (in addition to access) to the right people at the right place and time. In collaboration with Chief Information Security Officers, Chief Digital Officers seek to ensure engaging, personalized experiences at every touchpoint during an individual's relationship with a given organization – and doing so securely.[8] With this goal in mind, CIAM professionals deploy different tools, including just-in-time user provisioning, social sign-on, and user registration. This article will continue to draw out further differences and similarities between workforce IAM and CIAM.

B2C vs B2B vs B2B2C

Readers may have seen references to business-to-consumer (B2C)[9] and business-to-business (B2B). In some cases, CIAM focuses primarily on B2C use cases with a secondary focus on B2B. CIAM technology offerings tend to help an organization offer sign-up and sign-in services optimized for an individual to interact with an organization. Secondarily, a CIAM technology offering might also provide B2B service to facilitate trust between two different organizations, enabling employees from one to access services from another. Knowing whether a problem or project relates to a B2C or B2B context significantly impacts the requirements.

There is a third B2* permutation: business-to-business-to-consumer (B2B2C). In this case, a technology service provider offers CIAM capabilities to multiple organizations that use those services to engage with their customers. In B2B2C scenarios, delivering CIAM services with the correct brand experience is critical. This experience consists of everything, from the logos and colors on the screens to the URLs that an end-user would see. Instead of the upstream service provider brand, the customer should always see the brand of the business with whom they have a direct relationship.

(The primary focus of this article is on B2C use cases, though it will highlight some notable differences in B2B use cases. Unless otherwise specified, the reader should assume examples and guidance are oriented towards B2C use cases.)

The Stakeholders and Measurements

Successful digital engagement requires a successful CIAM strategy. Successful digital engagement also requires a very different collection of stakeholders than workforce IAM professionals might be used to. This expanded set of stakeholders has a new vernacular and a different set of goals from which the CIAM practitioner needs to derive requirements. Furthermore, the varied perspectives of these audience members require practitioners to translate the benefits and value of CIAM to different contexts. The stakeholders in digital engagement include marketing, digital, sales and distribution, product, privacy, legal, and customer service. In addition to these players, CIAM teams will also see a more familiar face: security.

A shared digital engagement mission often includes the following goals:

- **Increase Engagement**: Increase the number of people actively using whatever the organization produces, be they physical, informational, or digital
- **Reduce Friction**: Reduce the number of steps and tasks that stand in the way of an individual getting to use whatever the organization produces
- **Build Loyalty**: Ensure repeat use/engagement through products and customer service

CIAM practitioners partner to conduct this mission against a backdrop of security, appropriate data usage, and operating costs.

The stakeholders sharing this mission use different metrics than workforce IAM practitioners. In digital channels, engagement is often measured by the number of:

- Unique visitors to an organization’s site or app
- Page views
- People actively using the products and services within a given time frame, often referred to as “Monthly Active Users” (MAU) or “Daily Active Users” (DAU)
- Unknown visitors converting to either sales or registered accounts, known as “Conversion Rate.”

Further to these goals, building loyalty comes with its own set of measures, including customer satisfaction, net promoter score, and customer lifetime value. While CIAM teams might not be directly involved in gathering these metrics, they will certainly hear about it if customer satisfaction dips because of (or is inherently limited by) an onerous login process.

Although people recognize friction when they see it, defining and quantifying it is more difficult. Often, CIAM teams hear statements such as “It’s too hard to register for an account” or “It’s too many clicks to get to the content.” Abandoned account sign-ups, the number of screens or fields to register, failed logins, support calls, and even password reset rates are all indicators of friction. The organization’s need to reduce friction in its sign-up and sign-in flows demands a careful, iterative design process that finds (and seeks to eliminate) the places where people get stuck or give up in frustration. To add to the challenge, security and privacy stakeholders often seek to introduce *more* friction to thwart automated attacks, ensure regulatory compliance, and avoid harmful user choices. The balancing act for stakeholders and the implementation team is not simple.

Authorities, Lifecycles, and Administration

CIAM underpins digital engagement and enables organizations to offer products and services via digital channels as a sole channel or in addition to existing brick-and-mortar channels (e.g., phone or a physical location). This difference in context means that the sources of authoritative information about end-users, the lifecycle of those users, and the methods by which those users are administered differ from workforce sources, cycles, and techniques.

Authoritative Sources

In the workforce context, an IAM system can usually rely on human resource systems or databases to be authoritative about who is an employee, their demographics, and their roles and job responsibilities. In CIAM, no such system is consistently present and reliable. While a customer relationship management (CRM) system might exist and possess customer profile data, it is not definitive. Similarly, an eCommerce system, if present, might maintain shopper profile data that is, again, not definitive. While either might have information about an individual, neither is authoritative: the individual is the authoritative source of information. After an individual creates a user account via the CIAM system, their resulting profile is (ideally) linked to CRM, eCommerce, Customer Support, etc., using one or more unique, verified identifiers such as email, phone number, and account number. [10] One notable exception is the B2B use case in which a CRM system might be considered authoritative (about which individuals work for which organizations).

Lifecycles

The user lifecycle in CIAM may seem different from what the reader is familiar with if they come from a workforce background. In workforce scenarios, the reader might be familiar with the concept of “joiner”, “mover,” “leaver” (JML), which reflects how a new employee joins the organization, changes roles (aka moves) throughout their career, and eventually leaves the organization. Such events are recorded in authoritative sources, like a human resource system.

However, the lifecycle for a customer looks quite different: they register for an account and, ideally (from the organization's perspective), never stop using that account. There is often no event from an HR system equivalent to trigger user account creation, change, or deletion. CIAM and associated authorization systems will often query CRM systems to pull information such as “Is the person a Gold Level member?” to determine access to downstream resources at the precise time the resource is accessed. In this regard, CIAM tends to be a world of just-in-time authentication and authorization instead of admin-time, in which user accounts and associated resource access are set up in advance.[11]

Some readers might ask, “If my existing customers’ profiles exist in the CRM, can we use that to automate user account creation and distribute the credentials?” Do not do this. In a post-GDPR world (referring to the European Union’s General Data Protection Regulation[12]), such an action will be interpreted as a violation, i.e., signing the individual up for an account without their consent. The individual is in control in B2C CIAM use cases; thus, actions need to be taken just-in-time, not *a priori*.

Administration

The theme of individual control continues into the topic of administration. The individual can and must be able to control and update the information they have provided to the organization, including name and contact information.[13] The data they must be able to control includes their password, if they have one. The organization might also grant workers similar abilities in their customer service organization to help individuals who reach out to contact centers. These capabilities come with significant security risks, and the reader is encouraged to read IDPro’s BoK article entitled ”Managing Identity in Customer Service Operations.”[14]

In B2B use cases, the organization not only needs to provide user accounts and associated access to their business partners but also enable specific people within the partner’s organization to manage their own users’ access. Known as “Delegated Administration,” this capability looks similar to granting different people within the organization the ability to administer users in other parts of the organization.

Profile, Preferences, and Consent

CIAM systems are often used to enable information gathering, including demographic data (such as age and address), contact preferences (if at all), and their approved uses for any data collected. Organizations use this information to personalize experiences, deliver goods and services, as well as use data the individual shares for business purposes.

Profile

A profile is a collection of attributes about the individual. The individual may provide it directly or indirectly, such as in social sign-up and sign-in experiences. This information enables personalized user experiences, such as using an individual’s first name on the welcome screen of a mobile app. This personalization can also include providing specialized offers based on, for example, where they live.

The profile can also include information that businesses require for essential processes. For example, the individual might provide their street address so the organization can send physical goods to their home. In some cases, organizations’ business processes include evidence that an individual is old enough to use the service itself. For example, an online gambling site may have specific regulatory requirements to verify that an individual is over 18. Alternatively, the organization may be required to gather and verify legal identity information from the individual. For example, a bank must verify an individual's legal identity to adhere to “Know Your Customer” (KYC) regulations that prevent money laundering and other financial crimes.

Preferences

Commonly, individuals are not interested in every possible product and service an organization offers; similarly, the individual may prefer one contact method over another (e.g., text message vs. email). This kind of choice is captured in the form of preferences. Preferences may include topics of interest related to an organization’s offerings (e.g., sporting goods, elder care, etc.), approved communication channels (e.g., “none” or “email but not text messaging”), and frequency of communication (e.g., monthly emails not daily.) While this information is not strictly required for business processes, it vastly improves the individual’s experience with the organization.

Consent

In response to questionable practices, an increasing number of regulators require that an individual actively and positively chooses to interact or share data with an organization. This proof is referred to as consent. Said differently, an organization often needs to record evidence that the individual asserted that they want to interact with the organization. Organizations often bundle this consent with an acknowledgment that individuals agree to terms of service and conditions of use. The presentation of this choice must be clear: this means visible and accessible as well as understandable. The granularity of consent requirements varies from region to region and industry to industry. The cadence with which an organization must gather consent information may also differ. Organizations often bundle the gathering of consent with an acknowledgment that individuals agree to terms of service and conditions of use. Understanding the consent requirements for any use case or jurisdiction is critical.

Progressive Profiling

The aggregate of profile, preferences, and consent data can be considerable. Organizations are not advised to gather all this information at any one moment, like when the individual signs up for a new account or attempts to checkout during an e-commerce transaction. Doing that would ask the individual to fill out too many fields and screens. It puts too many hoops between them and the goal they set out to achieve. In e-commerce scenarios, too much friction can lead to dropouts when individuals give up and move on to a competitive offering. In the CIAM world, friction is akin to inefficiency in workforce user provisioning: it is the enemy.

To combat this enemy, organizations can employ a technique by which they ask for profile, preference, and consent information over time and not all at once. They can ask for information, such as shipping address, at the time they need it instead of when the individual first arrives at the website or service. Known as “Progressive Profiling,” this technique reduces friction by spreading it out across interactions and over a longer period.

Profile Versus Credential

It is essential to keep clear in one's head the relationship and separation between a profile and a credential. Where a profile is a collection of attributes related to an individual, the credential is the means by which the individual identifies themselves to the website or app with a certain degree of certainty. In its simplest and most basic form, a credential is a username and password combination. On the other hand, a profile can have a very rich data structure composed of many attributes of different types. Because the purpose of these resources differs, the techniques needed to manage and protect them are different. At the highest levels, profile data is within the domain of data management and privacy professionals and their tools, while credentials are squarely in the domain of identity and security practitioners and their associated tools.

This distinction leads to a critical question about ownership within the organization. In this context, do not think of ownership with a legal mindset: we are not discussing ownership like one discusses owning a candy bar. In this context, ownership is a conversation about who, within an organization, is responsible for gathering, managing, protecting, and making use of this data. Although a CIAM technology stack may be able to obtain and store profile data, it does not mean that a) the identity team owns the profile or b) that the profile is the only form or representation of a customer within the organization. Consider that organizations will have many "pictures" of a given customer in systems such as the CIAM, customer support, marketing, and operational systems. Profile data is legion within organizations.

Why dwell here? Previously, this article discussed the various teams involved in a digital engagement program. These teams will claim, with good reason, that they own the customer profile and are thus responsible for gathering and managing it. They are not wrong in this regard, and their requirements, as foreign feeling to identity teams as they may feel, are just as valid as security or regulatory requirements with which an identity team may be more familiar. Partnership here is a must: calories spent debating ownership are better applied to building better experiences for the individual.

In the case of credentials, however, these fit in the CIAM domain and are the subject of the next section.

Credentials

Where profiles are information shared by individuals to help organizations personalize their experience, credentials are how those individuals make themselves known to an organization. Said differently, credentials are how individuals authenticate themselves to an organization's CIAM system and, thus, the entire digital landscape. Generally speaking, there are two parts to a credential:

- An identifier
- An authentication mechanism

Identifier

As the name implies, identifiers are the "name" an individual uses to tell an organization's CIAM, "I am HappyCustomer01@my.mail." More often than not, email addresses and phone numbers are used as identifiers. Using them has a side benefit: it cuts down on the information an individual has to provide as a part of their profile. Because organizations want to communicate with the individual, they often ask for their preferred email address or phone number. Using email and phone as identifiers allows them to serve double duty as both an identifier and a communication channel. Importantly, the identifier is the username in the classic username and password combination.

While seemingly straightforward, identifiers and the handling thereof can be far more complicated than expected. It is strongly recommended that the reader reviews the IDPro Body of Knowledge article "Identifiers and Usernames."[15]

Authentication Mechanisms

Having provided a valid identifier, the individual is prompted to authenticate. The most well-known and entrenched are passwords, but others exist. Increasingly, these alternatives to passwords are becoming popular.

Passwords and One-Time Passwords

The most familiar authentication mechanism is the password. Passwords are shared secrets, meaning that both the individual and the CIAM system maintain the secret to verify that the individual is who they claim to be.

Passwords are the somewhat unfortunate bedrock upon which authentication has built its castle. Refer to the IDPro Body of Knowledge "Authentication and Authorization" for more on authentication.[16] Read the National Institute of Standards Special Publication 800-63B, section 5.1, to receive guidance on good practices for password composition and treatment.[17]

Because individuals' memories are fallible, organizations need to provide means for individuals to prove they are who they claim to be and then set a new password. Known as either account recovery or password reset, these processes are often overlooked and become attack vectors for adversaries. Neglecting these processes leads to difficult user experiences, constrains account protection, and increases customer support interactions. Failing to protect password reset processes can lead to account take-overs in which an adversary exploits a weak password reset process, sets a new password known to them rather than the account owner, and takes control of the account and its associated resources (e.g., emails, photos, files, social media accounts, bank accounts, etc.). Readers should review the IDPro Body of Knowledge article "Account Recovery" for more information.[18] Additionally, it is strongly recommended that identity professionals spend time at the *beginning* of a CIAM project considering their account recovery processes across all channels (web, mobile, phone, etc) through which their organization will interact with individuals.

Shared secrets are not the only game in town. Increasingly, organizations are opting for one-time passwords (OTP). These are shared secrets with a limited lifespan and, as the name implies, can only be used once. Common examples of one-time passwords include sending a code to a mobile phone or email address. OTPs can have a better user experience; they do not require the individual to remember or store a password, and operating systems and browsers perform better at automatically filling in OTPs when they detect them. However, these benefits can be outweighed by the risks of phishing and interception. One thing to note, at this time, OTPs are often considered part of the larger "passwordless" authentication world: this is both confusing and inaccurate.[19]

Passwordless

Passwords and OTPs are not the only methods that a CIAM team can choose to deploy. Increasingly, technology providers are offering truly passwordless offerings. These offerings generally rely on a combination of public key infrastructure (shielded from the user), trusted computing mechanisms for storing those keys, and a dedicated app or browser or operating system-provided user experience to strongly assert that the individual is who they claim to be. From the individual's perspective, they either provide a biometric (such as TouchID or FaceID Apple-centric environments) or interact with a mobile app protected by biometrics or PIN. These interactions "unlock" access to the site or service.

The interest and popularity of passwordless approaches are partially fueled by the acknowledgment that passwords are poor solutions for individuals and organizations. More recently, the industry is adopting the WebAuthn standard (and associated standards). WebAuthn is a standard overseen by the W3C,[20] and its implementations can be found in most modern mainstream browsers and operating systems. Most recently, agreements on how the cryptographic material needed to power WebAuthn-based passwordless authentication can be synchronized between devices and browsers to facilitate an "enroll once, use anywhere" end-user experience, known as passkeys, have driven even more excitement and interest.

Challenges still exist with passwordless approaches, including how an organization should trust an individual they have never seen before and how an individual can get back to their user account in case of a lost device. Additionally, passwordless approaches often require modern smartphones or computers, which are unavailable to many. But that said, passwordless approaches, especially those that are standards-based, represent a path from passwords to something materially stronger with an improved user experience.

Social Login

IAM professionals may choose to augment their password and passwordless sign-on offerings with social sign-up and sign-on. In this case, an individual identifies themselves to the organization by first authenticating to another service, such as a social network or email provider. In this case, the organization doesn't hold any secrets (e.g., passwords) from the user but instead records that the associate user account needs to be authenticated by the external identity provider (the social network, email provider, etc.) Organizations can not only use a social credential to authenticate an individual but also use the information that the external identity provider provides to create or pre-populate a profile for the individual; this is social sign-up.

While social sign-up can be very appealing, it does come with some downsides. It is inherently exclusive in providing a different kind of login experience to people who are members of a specific social network. While claiming hundreds of millions of members, offering a specific social network may not feel that exclusive; individuals will likely have strong preferences. This means that organizations often offer login via multiple social networks and email providers. In turn, this leads to the NASCAR problem in which a site's login page starts to resemble a NASCAR car festooned with different logos. Leaving the NASCAR problem aside, even having one social credential option means the organization is putting another organization's brand on theirs. These choices, to some, can be polarizing at the worst and off-putting. There is an inherent assumption that the external identity provider can protect secrets and offer recovery options that are superior to what the organization can do itself. That is often a reasonable assumption, but it is worth considering before deploying such offerings.

Some organizations may require higher assurance about individuals for regulatory or business process reasons. Such organizations can deploy a user experience similar to social login – one that relies on an external identity provider and is presented as a set of choices on sign-up and sign-in screens. The key difference is that the external identity provider is a government or financial sector service. This topic area is robust and requires a more advanced examination in a future Body of Knowledge article.

Functions and Components

Functions

At their core, CIAM systems perform at least user registration and authentication. User registration allows an individual to create an account and establish a credential. It may also include collecting profile, consent, and preference data. User authentication validates the credential the individual provides when they access the organization's apps and services. It is important to note that CIAM systems usually do not trigger a user provisioning process after establishing a new user credential. However, this may be more common in utilities or B2B and B2B2C scenarios. This stands in stark contrast to more traditional workforce IAM scenarios in which the detection of a new employee in an HR system often triggers user provisioning workflows. CIAM more often relies on the just-in-time (JIT) creation of user accounts brokers by single sign-on during run-time instead of user provisioning at admin-time.

Additionally, CIAM systems often provide two more capabilities: single sign-on and OAuth token management. The single sign-on capabilities offer individuals a seamless experience as they navigate across the different websites and services an organization provides. For example, this ensures that when the individual logs into the eCommerce site to purchase something, they can access the customer support site without logging in again. The OAuth token management capabilities are used to issue OAuth tokens to the individual and their apps. These tokens are used to access APIs that the organization provides. The individual may not be aware that they have been issued tokens and are using them in many interactions with the organization’s goods and services, but identity professionals and their security peers need to be aware of this – if only to take steps if an individual’s app or device is compromised. In this case, revoking the issued token(s) will prevent further access to the compromised app.

Finally, the CIAM system may provide some form of orchestration service. This service can build the user experience the individual sees as they register and integrate third-party services into user experience flows. Such integrations can further enhance the individual's experience, perform progressive profiling, or even add higher assurance that the individual is who they claim to be.

Components

While different technology suppliers’ specific architectures and components' names will vary, they generally share the same notional architecture.

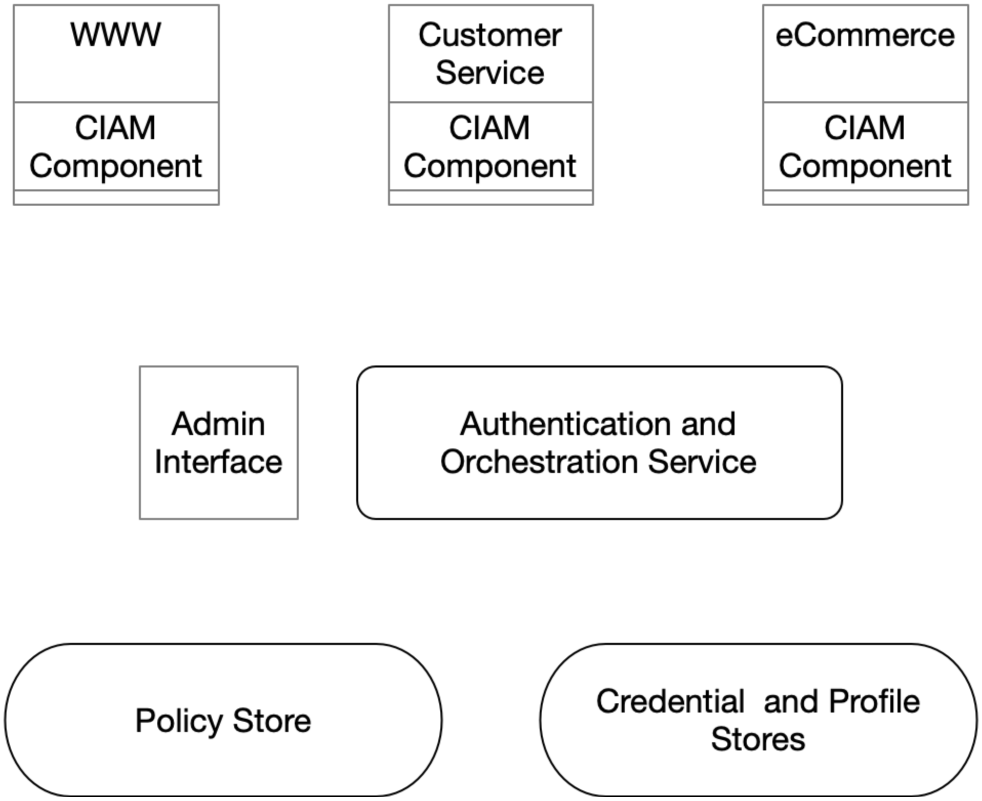


Figure 1: Components of CIAM Architecture

Credential and Profile Stores

At a minimum, a CIAM system provides a credential store or integrates with an existing one. This store is where user accounts are maintained and shared secrets, if any, are housed. The implementation of the credential store can range from a relational database to an LDAP directory to a NoSQL database and beyond.

The CIAM system may have a profile store that could contain profile, preference, and consent data. Profile storage, however, is not a strict requirement. Consider that organizations often centralize this kind of information in a customer data platform or marketing automation system. In such cases, the CIAM system will have the minimum amount of information needed for personalization and communication (e.g., a verified email address used to facilitate password resets) while the rest of the profile information is stored elsewhere.

Policy Store and Admin Interface

This repository houses configuration information for the CIAM system and serves as an authoritative source of that information. Such information could include single sign-on configurations, OAuth token lifecycle policies, and user registration workflow definitions. It might also house authorization policies that govern which resources an individual can access. The artifacts in this repository are managed by the Admin Interface or via changes to configuration files. The Admin Interface, if provided, presents a user experience where identity professionals can configure and manage the CIAM system.

Authentication and Orchestration Service

This service can serve multiple roles. At a minimum, it authenticates credentials. It can also manage the user's session and broker single sign-on. It may act as an OAuth authorization service. Lastly, it can act as an orchestration service. In this final context, it relies on policies in the policy store to determine, for example, what information must be gathered from an individual as they register, when to present an additional authentication challenge (often mediated by a third-party risk modeling service), or the steps required to reset a password.

CIAM Component

In each of the example websites in the above figure, the reader will notice the CIAM Component. This is an optional component that can help broker user registration and authentication. Often a piece of JavaScript, web component, or library, this piece of the puzzle can securely interact with the authentication and orchestration service. Often, in cases where this component is not present, the individual is redirected to a user experience that the authentication service provides to register and authenticate; they are then redirected to the site or app where they started.

Constraints and Challenges

Like other domains within the digital identity industry, CIAM comes with its own unique set of hills to climb. What follows is not an exhaustive list, and readers have likely discovered others.

Risks of Being on the Internet

CIAM systems, by their very nature, are on the public internet. After all, that's where an organization's customers are. It may go without saying that the internet is a space fraught with adversaries and risks, but it is especially important to say it about the identity systems of the internet. Every major touchpoint in a customer journey is susceptible to attack, especially sign-up, password reset, and login. Three kinds of attacks to be aware of are:

- Fraudulent Registration
- Credential Stuffing (aka cred stuffing)
- Account Takeover (ATO)

Fraudulent Account Registration

In this attack, the adversary (including bots) registers a new user account in the CIAM system using either bogus or stolen personal information. Their motivations vary from wanting to fill forums and chat groups with spam and malware links to harvesting new customer discount codes. Mitigations to these attacks can include anti-fraud systems for detection and reCAPTCHA-type puzzles, although the latter have been shown to be less effective than in years past.

Credential Stuffing

In this attack, an adversary tests whether lists of username and password pairs work in a given CIAM system. Often, adversaries acquire credentials (e.g., they may purchase them on the dark web) and test whether those credentials work at different online services. Their value on the black market is determined by the types of services those usernames and passwords can access. In many cases, the adversary is not interested in abusing an organization's service itself; instead, they are testing to see if the credentials work with your service so that they can sell it at a higher price. The reason why credential stuffing is even a thing is because people have a habit of reusing passwords. Mitigations to these attacks include specialized credential stuffing detection technology (often closely aligned with bot management and protection) and enforced multi-factor authentication (MFA).

It is important to note that credential stuffing differs from brute force attacks. In the brute force attack, the adversary is interested in testing whether an array of passwords works with a specific username. Brute force attacks can be mitigated in a variety of ways, including failed login throttling, in which multiple failed logins for the same user trigger either a slowdown in the number of times the user is allowed to log in or even a cooldown period during which all logins for the user are blocked. Credential stuffing cannot be mitigated with these measures because a CIAM system will only see one failed login per username/password pair.

Account Takeover

In this attack, an adversary possesses the means to act like the genuine authenticated user. The adversary may have the user's password (e.g., via a phishing campaign). The adversary might have found a weakness in the password reset process and forced a password change on a genuine user's account. Regardless of the means, the outcomes are the same: the adversary is in control of the user account – and may very quickly take steps to block the user from regaining access (such as changing the phone number). From that point forward, all means of nefarious actions can happen. Early detection is important but not sufficient to mitigate account recovery. Please refer to the IDPro Body of Knowledge article "Account Recovery" for more information.

Migrating CIAM Systems

Today, most organizations have an existing CIAM system. It might be tightly bound to an eCommerce platform or collaboration platform. If the organization has decided to modernize or replace its CIAM, then it is likely that IAM team members will be confronted with a migration. While migrating usernames is reasonably straightforward, migrating passwords is not. Two significant challenges are exporting passwords from the old system and getting them into the new one.

Exporting passwords presents significant challenges. It is important to note, for the avoidance of doubt, this article assumes that the word “password,” in the context of secure storage, means a password hash: systems should never store passwords in their recoverable plain text form. If exports are allowed, further data must be exported, including those comprising the security features known as “salt” and “pepper.”[21] With all three, taking extensive protective measures during migration is essential since they represent “loaded weapons.”[22]

Importing passwords requires not only the appropriate “salt” and “pepper” data but also the hashing scheme used by the previous system. Some CIAM solutions have specific features that support this process, but not all do.

Not all systems allow password exports at all. When organizations cannot migrate passwords, then at least two choices exist. Choice one involves telling the users to reset their passwords. This is not a great choice – it will certainly invite the attention of a grumpy Chief Digital Officer or other stakeholder(s). Choice two involves keeping the old CIAM alive and using it as a “dumb” credential store. When the user arrives and attempts to log in, the new CIAM tests the provided username and password against the old CIAM repository. If the credentials are good, the new CIAM records the password and marks the user as migrated. This approach is more complicated to deploy and requires that the old CIAM stays operational for a much longer period of time than the team might hope for (or want to pay for).[23]

Budget and Ownership

As discussed in the “The Team and Measurements” section, there are multiple stakeholders at the CIAM table. Besides bringing a diverse set of requirements and language, they bring their own teams and stakeholders, their motivators, their priorities, and their opinions. Who funds, operates, enhances, and is responsible for a CIAM stack can become a difficult set of questions to answer. It is not unusual to have the Chief Digital Officer take responsibility for the CIAM experience, a large percentage of the requirements, and funding. Partnered with them is the Security team, who have other requirements and are responsible for monitoring and incident response. The Identity team might be part of either organization or a separate Information Technology team. Regardless, expect that upper management will need to establish clear lines of demarcation between the various interested parties and, furthermore, to ensure there is a clear set of priorities that aligns the collective.

Topics for Future Investigation

This Body of Knowledge article is meant to be an introduction to Customer Identity and Access Management. The topic is both broad and deep: exploring the entire landscape is beyond the scope of an introductory article. The following is an incomplete list of what could and should be explored in the future:

- Incident response playbooks and documenting who to call when customers cannot register or log in
- Identity verification and proofing's role in CIAM
- High availability architectures for CIAM
- The use of fraud prevention tools to protect sign-up and sign-in
- Use of government- or financial services-issued credentials
- Emergent trends in credentials, including verified credentials
- Cross-channel or "omnichannel" CIAM

Conclusion

CIAM represents one of the biggest opportunities for identity professionals to demonstrate the value of their work. Through CIAM, identity professionals can help organizations reach new customers and grow the top and bottom lines. In this way, it is different from workforce IAM. These differences invite stakeholders from new parts of the organization – new partners, like Brand, Marketing, and Digital. Each new stakeholder brings their own set of requirements, languages, and business objectives. CIAM is, fundamentally, an internet-facing set of identity services that brings unique risks to model and mitigate. For more experienced identity professionals, CIAM may represent a fresh opportunity to reinvigorate their passion for digital identity. For newer members of the identity profession, it represents an exciting opportunity to have a meaningful positive impact on their organizations.

Author

Ian Glazer

Ian Glazer is the founder and president of Weave Identity – an advisory services firm. Prior to founding Weave, Ian was the Senior Vice President for Identity Product Management at Salesforce. His responsibilities include leading the product management team, product strategy, and identity standards work. Earlier in his career, Ian was a research vice president and agenda manager on the Identity and Privacy Strategies team at Gartner, where he oversaw the entire team's research. He is a Board Emeritus and the co-founder of IDPro and works to deliver more services and value to the IDPro membership, raise funds for the organization, and help identity management professionals learn from one another. During his career in the identity industry, he has co-authored a patent on federated user provisioning, co-authored and contributed to user provisioning specifications, and is a noted blogger, speaker, and photographer of his socks.

[1] Flanagan (Editor), H., (2022) "Terminology in the IDPro Body of Knowledge", *IDPro Body of Knowledge* 1(11). doi: https://doi.org/10.55621/idpro.41

[2] Epping, M. & Morowczynski, M., (2021) "Authentication and Authorization (v2)", *IDPro Body of Knowledge* 1(10). doi: https://doi.org/10.55621/idpro.78

[3] Ibid

[4] Glazer, I. & Robinson, L. & Hamlin, M., (2022) "User Provisioning in the Enterprise", *IDPro Body of Knowledge* 1(8). doi: https://doi.org/10.55621/idpro.84

[5] Koot, A., (2020) "Introduction to Access Control (v4)", *IDPro Body of Knowledge* 1(10). doi: https://doi.org/10.55621/idpro.42

[6] Nelson, C., (2020) "Introduction to Privacy and Compliance for Consumers (v3)", *IDPro Body of Knowledge* 1(10). doi: https://doi.org/10.55621/idpro.44

[7] Sesame Street (2009) "Sesame Street: Cookie Monster Sings C is For Cookie" https://www.youtube.com/watch?v=Ye8mB6VsUHw

[8] McKinsey and Company "Enhancing customer experience in the digital age" https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/enhancing-customer-experience-in-the-digital-age

[9] Yes, C in B2C stands for consumer and that only adds to the confusion on what the C stands for in CIAM. The writer of this article doesn't create the terms of art, just relays them.

[10] Cameron, A. & Grewe, O., (2022) "An Overview of the Digital Identity Lifecycle (v2)", *IDPro Body of Knowledge* 1(7). doi: https://doi.org/10.55621/idpro.31

[11] For more on admin-time versus runtime-time patterns, see: Bago (Editor), E. & Glazer, I., (2021) "Introduction to Identity - Part 1: Admin-time (v2)", *IDPro Body of Knowledge* 1(5). doi: https://doi.org/10.55621/idpro.27

[12] See, for example: Hindle, A., (2020) "Impact of GDPR on Identity and Access Management", *IDPro Body of Knowledge* 1(1). doi: https://doi.org/10.55621/idpro.24

[13] Ibid.

[14] Crow, A. & Rowan, J. P., (2021) "Managing Identity in Customer Service Operations", *IDPro Body of Knowledge* 1(4). doi: https://doi.org/10.55621/idpro.65

[15] Glazer, I., (2020) "Identifiers and Usernames", *IDPro Body of Knowledge* 1(1). doi: https://doi.org/10.55621/idpro.16

[16] Epping, M. & Morowczynski, M., (2021) "Authentication and Authorization (v2)", *IDPro Body of Knowledge* 1(10). doi: https://doi.org/10.55621/idpro.78

[17] Grassi, Paul, James Fenton, Elaine Newton, Ray Perlner, Andrew Regenscheid, William Burr, and Justin Richer. "Digital Identity Guidelines Federation and Assertions: Authentication and Lifecycle Management." Sectino 5.1, National Institute of Standards and Technology, U.S. Department of Commerce, June 2017. https://doi.org/10.6028/NIST.SP.800-63b.

[18] Saxe, D. H., (2021) "Account Recovery (v2)", *IDPro Body of Knowledge* 1(8). doi: https://doi.org/10.55621/idpro.64

[19] Again, the writer of this article doesn't create the terms of art but relays them with no small amount of cynicism.

[20] Hodges, J., Jones, J.C., Jones, M.B., Kumar, .A., and Lundberg, E. (2021) "Web Authentication: An API for accessing Public Key Credentials Level 2" W3C https://www.w3.org/TR/webauthn-2/

[21] Salt and Pepper reference

[22] Spacey, J. (2023) "Cryptography: Salt vs Pepper" Simplicable (Accessed on October 19, 2023) https://simplicable.com/IT/salt-vs-pepper

[23] If you didn't like passwords before, going through a CIAM migration will absolutely make you loathe them for certain.

Introduction to Privacy and Compliance for Consumers (v3)

By Clare Nelson

Updated by the IDPro Body of Knowledge Committee

To comment on this article, please visit our GitHub repository and submit an issue.

Table of Contents

Disclaimer

This article should not be considered legal advice. Identity programs that support data protection and privacy as a fundamental human right are complex and highly technical. Many of the data protection and privacy laws in countries around the world are evolving and open to interpretation. Please consult your organization's legal counsel for specific advice appropriate to your jurisdiction

Abstract

This introductory article on privacy and compliance in the consumer IAM domain sets the foundation for subsequent sections on privacy within the IDPro Body of Knowledge, providing an overview of a variety of topics, including definitions of privacy, different approaches to privacy in the consumer sector versus the workforce environment, and more.

Related Sections in the IDPro Body of Knowledge

Please refer to other forthcoming sections of the *IDPro Body of Knowledge*[i] for supporting and complementary information, notably:

- Andrew Cormack's "An Introduction to the GDPR"[ii]
- Andrew Hindle's "Impact of GDPR on Identity and Access Management"[iii]
- "Terminology in the IDPro Body of Knowledge"[iv]

Introduction to Privacy and Compliance for Consumers

Identity professionals, including enterprise solution architects, data scientists working in marketing, privacy professionals, product managers, and strategists at data brokers, have an opportunity to improve privacy plus compliance with data protection regulations and laws for consumer-facing applications. Several critical issues drive the need for improvement:

1. Unique identification of a *natural person,* such as a consumer, is easier than ever before. The smallest shred of digital exhaust, physical actions, or attributes can be collected, correlated via machine learning, and analyzed to identify a unique consumer or household with sufficient probability.
2. Consent is broken. The complexity of privacy notices, and the length of privacy policies that are rarely read, lead to a pattern of ineffectiveness, the illusion of choice, burden on the consumer, and the eventual agreement to broad terms which may not have limits or may be modified at any future date with only limited or obscure notice. As privacy and law expert Daniel Solove has stated, "Giving

individuals more tasks for managing their privacy will not provide effective privacy protection."[v] There is a growing realization that privacy laws should make stewardship of data the responsibility of the data controller and/or data processor, not the consumer.

3. Data privacy laws are years behind technology innovations, and the gap is expanding. Privacy laws cannot keep pace with technological advancements and are years behind in catching up to the 4,000 data brokers and analytics companies that collect personal data across all touchpoints, many of which are adding muscle with machine learning. Many marketing companies, plus some of the world's largest organizations, are seeking alternatives to third-party cookies in order to be able to continue to identify consumers, all within the porous guidelines of current privacy law.
4. Anonymity does not exist, and pseudonymization provides weak protection
    a. For example, researchers have proved that based on geolocation alone, anonymity does not exist.[vi]
    b. Similarly, researchers posit that it is becoming easier to re-identify a person:
        i. Once released to the public, data cannot be taken back. As time passes, data analytic techniques improve, and additional datasets become public that can reveal information about the original data. It follows that released data will get increasingly vulnerable to re-identification—unless methods with provable privacy properties are used for the data release.[vii]
    c. *Communications of the ACM* recently published an article on anonymity that confirms what many mathematicians have always known: there is still a pattern in the anonymous data and a way to de-anonymize it.[viii]
        i. "Anonymized data can never be totally anonymous: anonymization is not sufficient for private companies to avoid conflicts with laws such as Europe's General Data Protection Regulation, and the California Consumer Privacy Act."[ix]

The scope of privacy for what the GDPR calls *natural persons* keeps expanding because the ability to uniquely identify a human being with the tiniest bit of digital exhaust or trace of online or offline behavior keeps expanding. Where a person ate lunch, the way they moved their mouse to find the cursor, what they said to a service representative over the phone in light of the warning, "this call may be recorded for quality purposes," what they bought online, their device and all its related attributes, or where they bought gas may be sufficient to uniquely identify a natural person. Long gone are the days when name, address, social security number, and date of birth were the only identifiers. Now, we need to protect massive collections of personal and related data in order to provide privacy for consumers.

Terminology and Acronyms

- Consent - permission for something to happen or agreement to do something
- GDPR – General Data Protection Regulation[x]
- CCPA – California Consumer Privacy Act[xi]
- Natural Person – an individual human being
- NY SHIELD Act – New York "Stop Hacks and Improve Electronic Data Security" Act[xii]
- Privacy - an abstract concept with no single, common definition

Scope

The lofty goal of this section on *Privacy and Compliance for Consumers* is to present a global perspective. However, the initial release of this section is more focused on the GDPR, CCPA, and NY SHIELD Act with a light coverage of China and the rest of the world. As noted below in the Resources section, the International Association of Privacy Professionals (IAPP) is an excellent source of information for immediate, current, comprehensive global coverage.

Requirements for data protection and associated privacy regulations are increasing around the world. On March 21, 2020, the NY SHIELD Act went into effect, and China is working on updated privacy law. Most are familiar with the GDPR and CCPA.[xiii] The figure below highlights global regulation and enforcement as heavy, robust, moderate, or limited.

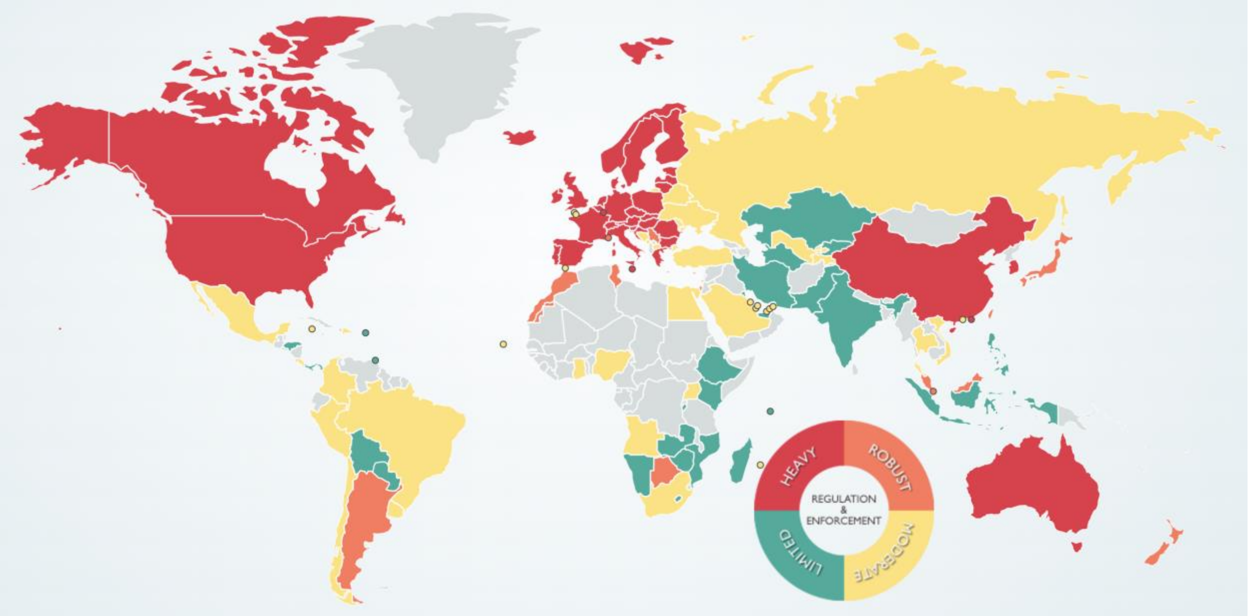


Figure 1 - Global Privacy Regulation Varies from Heavy to Limited[xiv]

In addition to a global scope, this section covers digital identities – including online services and apps – as well as physical interactions (e.g., customers entering a store or service establishment).

In this section, we consider personal data obtained, stored, or tracked through a variety of mechanisms, including cookies, electronic communications (Internet, email, messaging, apps), Wi-Fi, telephone, and Internet-of-Things (IoT). All of this data can be used to identify a unique individual and may be considered private, requiring protection as a *fundamental human right*. The first recital of the GDPR states that data protection is a *fundamental right* according to the Charter of Fundamental Rights of the European Union and the Treaty on the Functioning of the European Union (TEFU).[xv,xvi] The United Nations adopted the Universal Declaration of Human Rights in 1948, a global mandate for privacy, as articulated in Article 17:[xvii]

> *No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks.*

	EU GDPR	EU ePrivacy Directive	US CCPA, NY **SHIELD**, other	California, Oregon IoT Security Law; Singapore, others	Rest of the World
What is covered?	Personal data, cookie consent	Cookies, Internet, email, messaging, phone; tracking mechanisms, right of confidentiality	Personal data, household data under CCPA	Any connected device; has an IP address or Bluetooth address	China - none; APEC - Personal Information*

Table 1. Beyond GDPR: Personal Data Includes Cookies, Phone, and IoT

*According to the Asia-Pacific Economic Cooperation (APEC) Privacy Framework, Personal Information is any information about an identified or identifiable individual. [xviii]

The GDPR is closely related to the ePrivacy Directive. The ePrivacy Directive is soon to be replaced with the ePrivacy Regulation. The GDPR and ePrivacy Regulation are both parts of the data protection reform in the EU; where there is overlap, the ePrivacy Regulation overrides the GDPR, notably for cookies and electronic communication.[xix]

IoT law is covered below. Note that emerging laws seek first to get rid of embedded or hardcoded passwords. When an IoT device ships with the password already installed from the manufacturer, this makes it easy to breach privacy as well as to create botnet malware.

Setting the Stage

Scholars and privacy experts, including Hartzog, Zuboff, Schneier, and Maler, set the stage for examining these topics and beyond.

Hartzog

Woodrow Hartzog is a Professor of Law and Computer Science at Northeastern University, and among other roles is an Affiliate Scholar at Stanford Law School for Internet and Society. In April 2019, Hartzog co-authored *The Pathologies of Digital Consent* with Neil Richards, where they discuss defects that consent models can suffer, including: [xx]

- Unwitting consent
- Coerced consent
- Incapacitated consent

These consent defects are a far cry from the *gold standard of knowing and voluntary consent*. Hartzog and Richards conclude:

The over-use of consent in the digital context, combined with limited legal policing of the sufficiency of consent, has allowed great fortunes to be created on the basis of personal data, but it has also exposed consumers to data breaches, identity theft, and a surveillance economy unprecedented in human history, one which stretches the very notion of "consent" to say that it was ever actually agreed to.

More fundamentally, the manufacturing of consent by exploiting consent's pathologies has diminished the trust in our digital environment that is the key ingredient toward a better future. We can do better, but in order to do so, we need to recognize the pathologies of consent and limit consent to the contexts in which it is most justified. Going forward, we must rely on strategies other than fictive, manufactured, or coerced consent to minimize the risks and harms of our information economy if we seek to take advantage of its benefits in a sustainable, ethical, and progressive way.

These consent issues are discussed further in subsequent sections, including a proposed solution or *theory of consumer trust* as an alternative to an over-reliance on increasingly pathological models of consent.

Zuboff

In her recent, award-winning book, *The Age of Surveillance Capitalism: The Fight for a Human Future and the New Frontier of Power*, Harvard's Shoshana Zuboff clearly articulates her well-researched assertion that we live in a state of *surveillance capitalism.* Zuboff's message is clear:

"Surveillance capitalism unilaterally claims human experience as free raw material for translation into behavioural data.

- Although some of these data are applied to service improvement, the rest are declared as a proprietary behavioural surplus, fed into advanced manufacturing processes known as 'machine intelligence', and fabricated into prediction products that anticipate what you will do now, soon, and later.
- Finally, these prediction products are traded in a new kind of marketplace that I call behavioural futures markets.
- Surveillance capitalists have grown immensely wealthy from these trading operations, for many companies are willing to lay bets on our future behaviour."[xxi]

Schneier

Zuboff's 2019 book on surveillance capitalism makes Bruce Schneier's 2015 book, *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World,* pale in comparison. Consumers are left wondering, "Why didn't you tell me it was so bad?" because Schneier does not provide the chilling detail about how far surveillance and collection of behavioral surplus have advanced. Zuboff's words are even reflected in marketing messages of leading "data protection" vendors, "Privacy is the right of an individual to be free from uninvited surveillance."[xxii]

Maler

In her 2018 Identiverse talk, *Don't Pave Privacy Cow Paths: Retool Consent* for the New Mobility, ForgeRock CTO, then-VP Innovation and Emerging Technology, Eve Maler describes why "Consent doesn't scale for the requirements of email, laptops, and browsers, never mind mobile devices and applications.

- How much worse is the situation going to get as connected vehicles become an ever-bigger part of consumers' lives and an ever more significant integration point for every industry?" Maler establishes the "New Mobility as a critical scenario for examining consumer requirements for trust, regulatory requirements for privacy, how consent experiences and consent management must adapt, and how we can begin to meet these challenges."[xxiii]

Maler's words were prescient because, in the rush to implement consent for GDPR compliance, many companies have simply paved cow paths. Her talk describes how to refactor consent to accommodate today's architectural requirements for asynchronicity, automation, and abstraction.

What is Privacy?

Privacy is an abstract concept with many definitions and even more potential threats when that concept is attacked. There are areas such as handling of open-source intelligence (OSINT) that can have an enormous impact on the individual, but where the legal parameters are poorly specified (if they are specified at all).[xxiv] Concerns around the politicization and potential weaponization of personal data also highlight the challenges introduced by having so many perceptions of privacy.[xxv]

Privacy as a Fundamental Human Right

The protection of personal data often refers to autonomy and control over one's data. This level of autonomy and control varies depending on the context. In general, the definition of privacy differs from country to country or state by state. Even though the US is one of 48 United Nations countries that voted to adopt the Universal Declaration of Human Rights, the US does not share the EU's embrace of privacy as a *fundamental human right.* For example, instead of a comprehensive, federal law, it is building a patchwork of state-specific laws.

- In the EU, human dignity is recognized as an absolute fundamental right.
- In this notion of dignity, privacy, or the right to a private life, to be autonomous, in control of information about yourself, to be let alone, plays a pivotal role. Privacy is not only an individual right but also a social value.
- The right to privacy or private life is enshrined in the Universal Declaration of Human Rights (Article 12), the European Convention of Human Rights (Article 8), and the European Charter of Fundamental Rights (Article 7).[xxvi]

Westin. In the 1960s, Privacy pioneer Alan Westin defined privacy as "The claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others."[xxvii]

US Supreme Court. In 1989, the US Supreme Court wrote, "Both the common law and the literal understandings of privacy encompass the individual's control of information concerning his or her person."[xxviii]

China. In the People's Republic of China (PRC), a complex array of laws govern personal data and privacy, due to be eclipsed by comprehensive regulation in the future. The trend indicates "individuals are gaining significant data protection rights in the private sectors but cannot claim any remedies for the infringements of their privacy carried out by the state government."[xxix] Today, various laws apply, depending on the sector: financial services, e-commerce, telecommunications, Internet services, content providers, or healthcare.

APEC. The Asia Pacific Economic Cooperation (APEC) Privacy Framework can be downloaded here. The APEC Privacy Framework protects privacy within and beyond economies and enables regional transfers of personal information that benefits consumers, businesses, and governments. This framework is used as a basis for the APEC Cross-Border Privacy Rules (CBPR) System. The framework countries and participants include the countries in the map below.

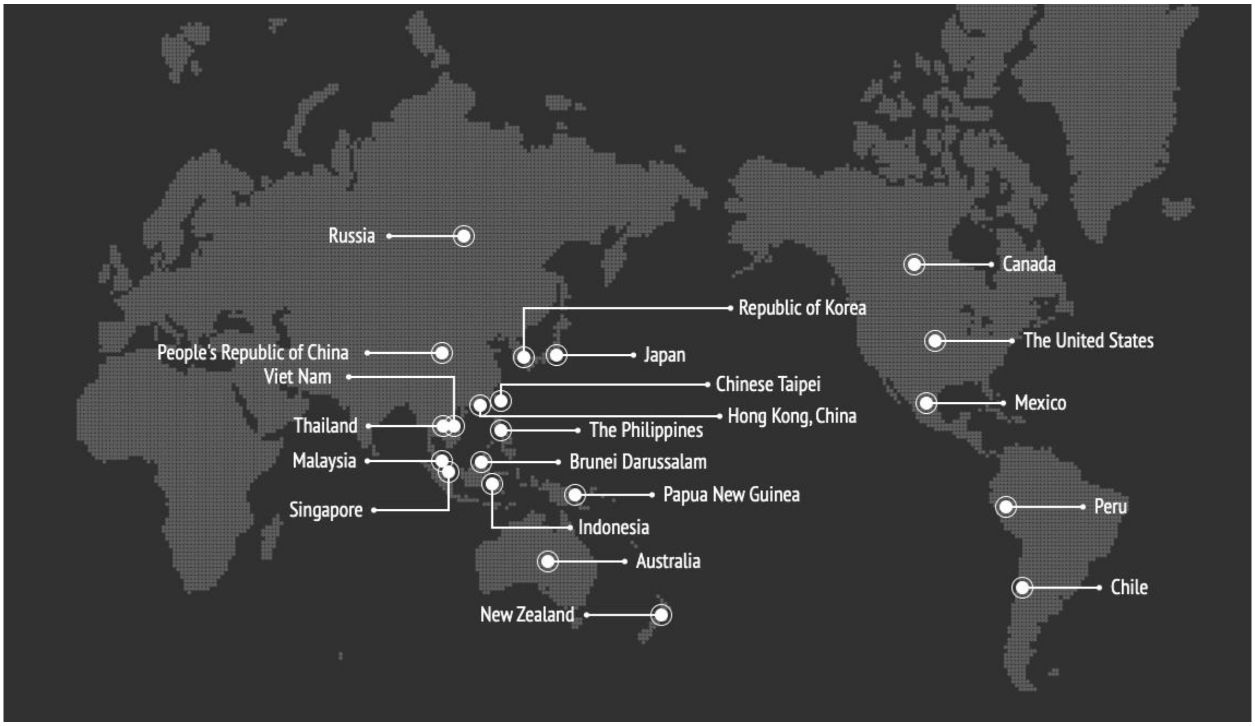


Figure 2 - Map of the APEC Cross-Border Privacy Rules (CBPR) System framework countries and participants[xxx]

Privacy Models

According to Samm Sacks, Senior Fellow Yale Law School, Paul Tsai China Center, there are two basic privacy models: 1) China, and 2) GDPR. She indicates that Viet Nam, Kenya, and India are more closely aligned with China's model.[xxxi] Even though China's privacy laws are influenced by the GDPR, they are markedly different both in detail (for example, China supports implied consent, whereas the GDPR requires explicit consent) and in spirit (in general, the rights of the state supersede individual rights). The privacy dichotomy in China is evidenced by the increased protection of consumers from technology companies such as Renren and other Chinese Facebook counterparts, even as government surveillance intensifies.

Beyond privacy law, organizations approach privacy for consumers in a variety of ways. The role of the identity professional is first to understand the organization's posture with regards to privacy, security, and risk.

Privacy Taxonomy

One of the world's leading experts on privacy law is Daniel Solove, the John Marshall Harlan Research Professor of Law at the George Washington University Law School. As depicted in Solove's privacy taxonomy below, privacy has two main parties:

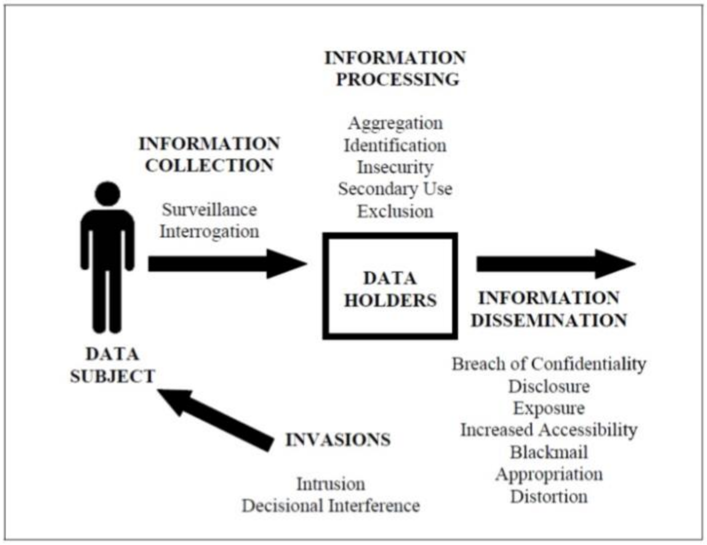
- The Data Subject, or consumer
- The Data Holders, or data processor and controller

The four main processes in the privacy taxonomy comprise:

- Information Collection
- Information Processing
- Information Dissemination
- Invasions

The four processes may be viewed in order, starting with Information Collection and ending with Invasions. The Information Collection of data about a Data Subject is done by the Data Holder, or data processor and controller to put it in GDPR terms. The collection of data about a Data Subject or individual may be done by another individual, business, government, or external organization via surveillance or interrogation. The Information Processing includes the storage of data and any additional steps taken to apply something like a software algorithm to further derive value. Insecurity refers to a lack of security. Information Dissemination includes many harmful things that might result if the information is part of a list of undesirable actions, including a Breach of Confidentiality, Disclosure, Exposure, Blackmail, or Distortion. Invasions include intrusion and decisional interference, which Solove describes as "the government's incursion into the data subject's decisions regarding her private affairs."[xxxii]

Privacy Taxonomy by Solove



Source: https://teachprivacy.com/what-is-privacy/

Figure 3 - Privacy Taxonomy by Solove

[Permission received from author to use this graphic]

Privacy by Design

Privacy by Design is the brainchild of Ann Cavoukian, one of the world's leading privacy experts; former Information and Privacy Commissioner of Ontario, Canada; former distinguished visiting professor at Ryerson University, where she was also Executive Director of the Ryerson's Privacy and Big Data Institute; and founder of Global Privacy and Security by Design Centre. Originally published in 2009, the Privacy by Design Principles, depicted in the figure below, are an integral part of the GDPR and subsequent GDPR-influenced privacy laws. Privacy by Design takes a holistic, systems engineering approach and makes it clear that compliance with regulations is not enough. Privacy by Design advances the view that the future of privacy cannot be assured solely by compliance with regulatory frameworks; rather, privacy assurance must ideally become an organization's default mode of operation.[xxxiii]

In the Privacy by Design figure below, note that *privacy by default* is one of the seven principles. The sharp contrast of cultural expectations may come as a surprise to some. As a gross generalization, the EU sensibility is to have privacy by default as the norm, whereas in the US, privacy by default is the rare exception.

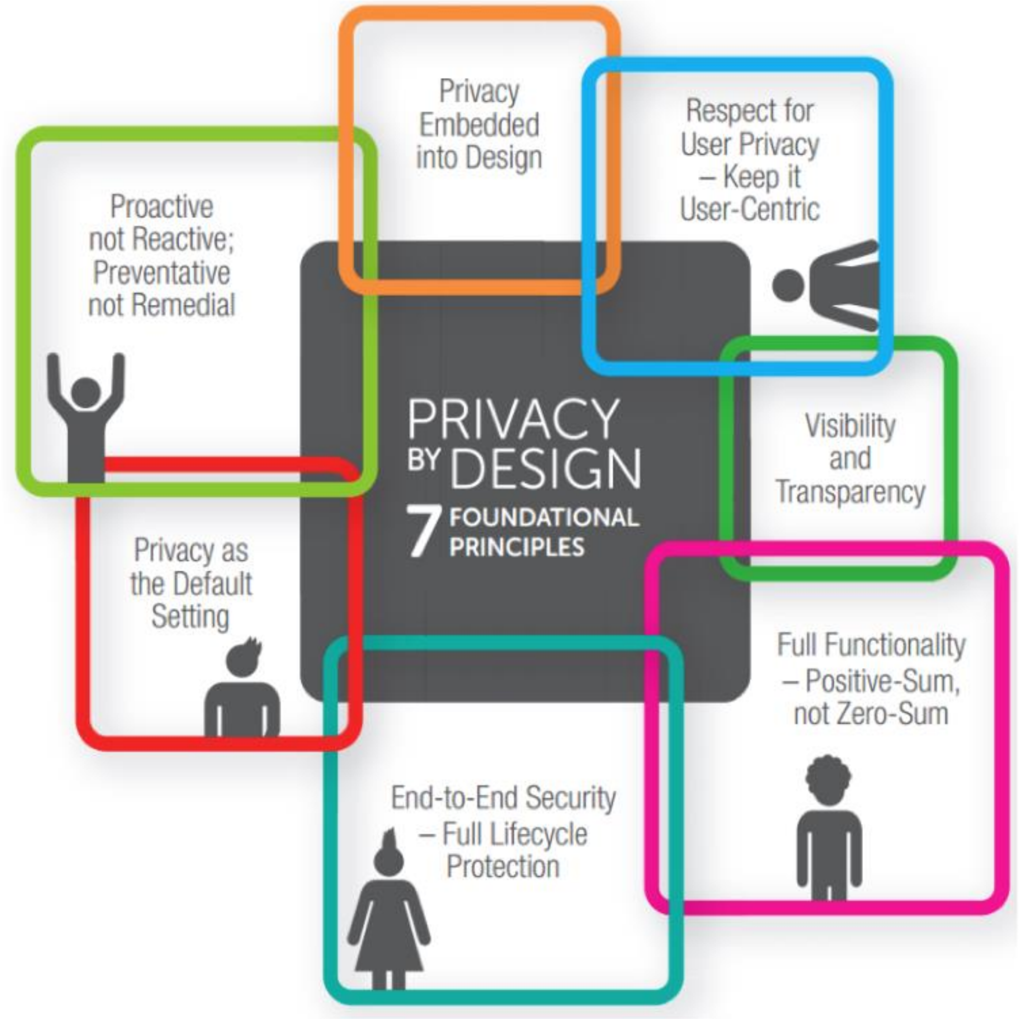

Privacy Embedded into Design
Respect for User Privacy – Keep it User-Centric
Proactive not Reactive; Preventative not Remedial
PRIVACY BY DESIGN
7 FOUNDATIONAL PRINCIPLES
Visibility and Transparency
Privacy as the Default Setting
Full Functionality – Positive-Sum, not Zero-Sum
End-to-End Security – Full Lifecycle Protection

Figure 4 - Privacy by Design, Seven Foundational Principles[xxxiv]

Cavoukian builds upon her initial Privacy by Design work in a subsequent document, *7 Laws of Identity, The Case for Privacy-Embedded Laws of Identity in the Digital Age*, where she maps privacy fair information to the Privacy-by-Design principles, resulting in "privacy-embedded Laws of Identity."[xxxv] She warns:

> *A universal identity system will have profound impacts on privacy since the digital identities of people - and the devices associated with them - constitute personal information. Great care must be taken that an interoperable identity system does not become an infrastructure of universal surveillance.*

Compliance is Necessary but not Sufficient

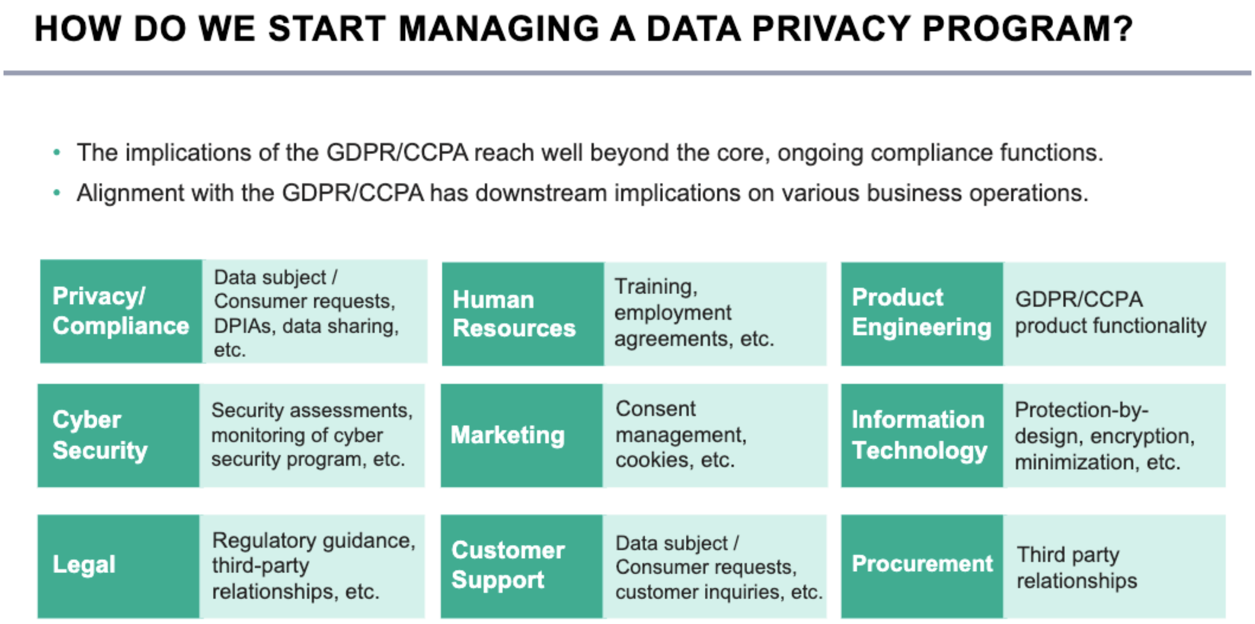
To a limited extent, privacy law enforces data protection. This section applauds the advances of privacy law, plus it explores some of the failings, flaws, and shortcomings of privacy law, including consent issues, time lag, and reactive posture because the law cannot keep pace with current innovations, and what Harvard's Shoshana Zuboff calls the asymmetric power stranglehold of Google, Facebook and others that are immune from the effective impact of privacy law because the law does not cover much of what they do with the collection of behavioral surplus.

- **Compliance ≠ Privacy or Security.** As an identity professional, remember that just because you are compliant does not mean you have achieved the appropriate level of privacy and security required by your organization (or expected by your customers), hopefully documented in its risk and privacy policies.
- **Privacy Law' Gap Growth' is Exponential.** Distinguished Fellow at Harvard Law, Vivek Wadha explains, "The gaps in privacy laws have grown exponentially. These regulatory gaps exist because laws have not kept up with advances in technology. The gaps are getting wider as technology advances ever more rapidly."[xxxvi]

Privacy and compliance capabilities are foundational for any Consumer Identity & Access Management (CIAM) program because they protect the personal data of consumers as well as safeguard organizations by defining guidelines for compliance in alignment with the organization's privacy, security, and risk management policies. If organizations do not comply, there are many negative consequences, including:

- Fines (for GDPR, up to €20 million or 4% of annual turnover, whichever is highest)
- Reputational or brand damage
- Loss of customers, loyalty erosion
- Lawsuits
- CEO may be held personally responsible

As an identity professional, you may be part of a team responsible for some or all aspects of privacy and compliance for consumers. This section will enable you to contribute and have a basic understanding of jurisdiction, consent, and data protection across the entire organization. For GDPR or CCPA compliance, you may interact with human resources, product engineering, security, marketing, IT, legal, customer support, procurement, and beyond, as shown in the figure below.



HOW DO WE START MANAGING A DATA PRIVACY PROGRAM?

- The implications of the GDPR/CCPA reach well beyond the core, ongoing compliance functions.
- Alignment with the GDPR/CCPA has downstream implications on various business operations.

Privacy/ Compliance	Data subject / Consumer requests, DPIAs, data sharing, etc.	**Human Resources**	Training, employment agreements, etc.	**Product Engineering**	GDPR/CCPA product functionality
Cyber Security	Security assessments, monitoring of cyber security program, etc.	**Marketing**	Consent management, cookies, etc.	**Information Technology**	Protection-by-design, encryption, minimization, etc.
Legal	Regulatory guidance, third-party relationships, etc.	**Customer Support**	Data subject / Consumer requests, customer inquiries, etc.	**Procurement**	Third party relationships

Figure 5 - How to Start Managing a Data Privacy Program, an Example[xxxvii]

Why Consumer Services Need Different Privacy and Compliance Strategies

CIAM and Workforce IAM

Privacy and compliance strategies for workforce IAM have some overlap with CIAM, but CIAM differs in some key regards. For this reason, simply applying workforce privacy and compliance to CIAM projects may not be optimal. Below are some of the key differences between privacy and compliance strategies for workforce versus CIAM projects:

- SCALE: CIAM scale is often orders of magnitude greater to reflect a large consumer population versus a smaller, more predictable number of employees and workforce
- CUSTOMER EXPERIENCE (CX): CX requirements for consumers are more demanding. For its members, IAPP provides a GDPR-centric document, *The UX Guide for Getting Consent:*
    - *"Consent is at the very heart of data protection and privacy,"* and while it is important, it is not the be-all and end-all of a privacy program. For example, a layered or intelligent privacy notice strategy can help make privacy interactions less cumbersome.
    - The data subject must have a say in how personal data is collected, used, shared, and destroyed.
    - Even if a choice doesn't appear to be promoted, wording, widget, and sequence matter.[xxxviii]

- LAW: Depending on the jurisdiction, the privacy law may differ in some cases for IAM versus CIAM.
- AUTOMATION: Appropriate levels of automation differ to meet spikey or unpredictable consumer demand.
- ADVERTISING: Online behavioral advertising in particular, and any advertising in general, is typically aimed at consumers, not the workforce.
- MACHINE LEARNING AND PROFILING: What the GDPR refers to as "automated processing", including profiling (automated processing of personal data to evaluate certain things about an individual); plus, machine learning is often applied to consumer data for different purposes for the workforce versus consumers.

CIAM and Social Identity

CIAM often relies on integration with social media identity providers. There are several benefits to this direction, including reducing end-user friction during sign-up and self-service registration, generating fewer usernames and passwords for the end-user to memorize, and simplified business processes that allow for outsourcing user account recovery processes. This integration is not without drawbacks, however, as integration with social media identity providers may enable cross-site tracking of users without their permission.

Security is Critical

Identity professionals need to understand their organization's risk management policies for security and privacy and work in concert with their colleagues who create those policies, as well as those responsible for the implementation of the policies. The security policy is a necessary dependency for any successful privacy policy. There is a saying, "You can have security without privacy, but you can't have privacy without security."[xxxix] Security or cybersecurity may be used interchangeably. Some also use the term information security. In the figure below from the NIST Privacy Framework, the relationship between cybersecurity risks and privacy risks makes it clear that managing cybersecurity risk may help mitigate privacy risk, but it is not sufficient because privacy risk can result from incidents outside the realm of cybersecurity incidents. For example, smart meters or smart thermostats may collect and record personal data and possibly represent a privacy risk even though they are operating as intended.

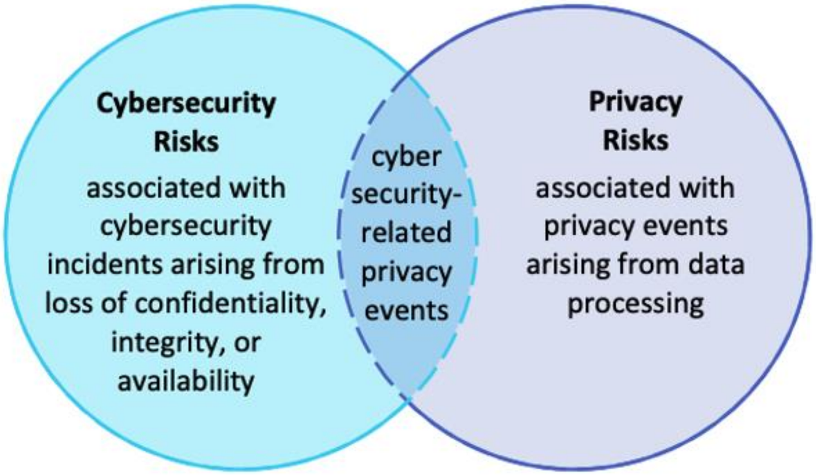

Cybersecurity Risks
associated with cybersecurity incidents arising from loss of confidentiality, integrity, or availability
cyber security-related privacy events
Privacy Risks
associated with privacy events arising from data processing

Figure 6 - Relationship Between Cybersecurity Risks and Privacy Risks[xl]

Privacy Policy is a Business Decision

An in-depth understanding of an organization's policies will provide clarity for the identity professional's role in privacy and compliance for consumers. For example, in some cases the marketing department may need to collect extensive personal data, and the organization's privacy policy may allow this. In other cases, the organization's business may depend on trust and confidentiality of personal data; and there may be ample budget to ensure data protection for consumers in a visible, transparent, and robust manner.

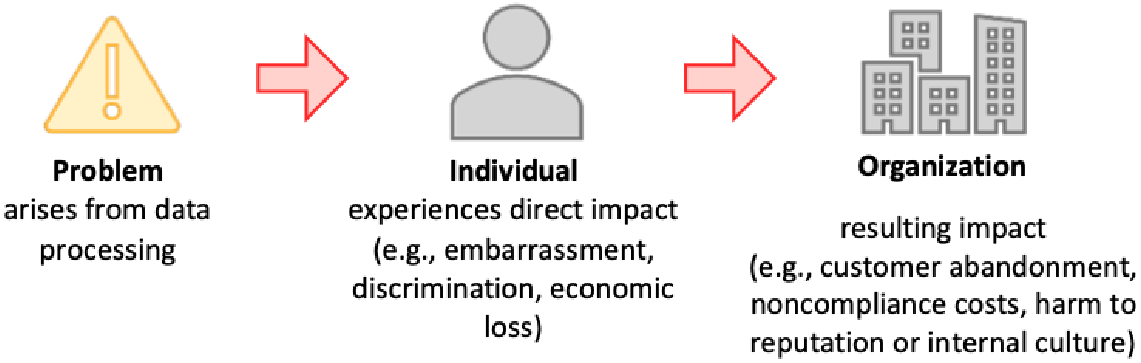

Problem
arises from data processing
Individual
experiences direct impact (e.g., embarrassment, discrimination, economic loss)
Organization
resulting impact (e.g., customer abandonment, noncompliance costs, harm to reputation or internal culture)

Figure 7 - Relationship Between Privacy Risk and Organizational Risk[xli]

How an organization deals with consumer privacy and any associated risk is a business decision; the option to mitigate, transfer, avoid, or accept risk may be made in concert with privacy policy formulation or at a later time.

- **Mitigate**. Mitigating the risk (e.g., organizations may be able to apply technical and/or policy measures to the systems, products, or services that minimize the risk to an acceptable degree);

- **Transfer**. Transferring or sharing the risk (e.g., contracts are a means of sharing or transferring risk to other organizations, privacy notices, and consent mechanisms are a means of sharing risk with individuals);
- **Avoid**. Avoiding the risk (e.g., organizations may determine that the risks outweigh the benefits, and forego or terminate the data processing); or
- **Accept**. Accepting the risk (e.g., organizations may determine that problems for consumers are minimal or unlikely to occur; therefore, the benefits outweigh the risks, and it is not necessary to invest resources in mitigation).[xlii]

Is Privacy a Competitive Advantage?

As noted above, laws and regulations typically lag innovative product and service offerings. Compliance with current and upcoming privacy laws is only the start. Privacy may be a competitive advantage or not. It depends on your organization and its consumers. In 2010, data protection pioneer and expert Alan Westin was paraphrased, "The idea that privacy can be used as a business advantage is dead, privacy controls are too complex for consumers to understand and a certification culture would be more effective."[xliii] Others take the counterargument. Organizations realize that many consumers would enjoy greater control over their data. Privacy for consumers is an opportunity to build trust. Among others, a GDPR and CCPA paper from Akamai provides "tips to build customer trust through regulatory compliance and identity governance."[xliv]

Beyond GDPR: ePrivacy and the New European Strategy for Data

In February 2020, the EU published "A European Strategy for Data." The continuous advancement and proven EU leadership in data protection is a driving force for the rest of the world.

The European Strategy for Data is sector-specific, e.g., healthcare, and provides for:

- Data can flow within the EU and across sectors.
- European rules and values, in particular personal data protection, consumer protection legislation, and competition law, are fully respected.
- The rules for access to and use of data are fair, practical, and clear, and there are clear and trustworthy data governance mechanisms in place; there is an open but assertive approach to international data flows based on European values.[xlv]

Through its focus on data sovereignty and supporting the privacy of people in its constituency, the European Union provides model guidance on newer technologies such as Decentralized Identity (DID) and Verifiable Credentials. Captured in large part in the Regulation on electronic identification and trust services (eIDAS Regulation), this regulation:

- ensures that people and businesses can use their own national electronic identification schemes (eIDs) to access public services available online in other EU countries;
- creates a European internal market for trust services by ensuring that they will work across borders and have the same legal status as their traditional paper-based equivalents.[xlvi]

The eIDAS Regulation is under consideration for an amendment that will further evolve the guidance available to include more information on digital wallets and their use.[xlvii]

Blockchain. The European Strategy for Data includes the evaluation of blockchain technology.

- New decentralised digital technologies such as blockchain offer a further possibility for both individuals and companies to manage data flows and usage, based on individual free choice and self-determination. Such technologies will make dynamic data portability in real-time possible for individuals and companies, along with various compensation models.[xlviii]

In addition, the French data protection authority, known as the National Commission on Informatics and Liberty (CNIL), has spearheaded work on "responsible use of the blockchain in the context of personal data" plus the potential privacy risks inherent in the technology.

The challenges raised by blockchains in terms of compliance with human rights and fundamental freedoms necessarily call for a response at the European level. The CNIL is one of the first authorities to officially address the matter and **will work cooperatively with its European counterparts to suggest a strong and harmonised approach.**[xlix]

Conclusion

Although it may be difficult to define privacy, the fundamental principles of Privacy by Design, depicted in Figure 5 above, create a well-defined foundation for understanding and implementing *Privacy and Compliance for Consumers*. This is why Privacy by Design is included in the GDPR and CCPA, and has significantly influenced subsequent privacy regulations and laws. By now, identity professionals have a clear picture of the interlinked dependencies between identity, privacy, and security. Security protects the data; how privacy is provided is based on business and risk policies. The silver lining for the daunting task of implementing privacy and compliance for consumers is that it may be viewed as a competitive advantage and well worth the extra effort.

Author Bio

Clare Nelson, CISSP, CIPP/E, AWS Certified Cloud Practitioner; is the CEO of ClearMark Consulting, specializing in business development and product strategy. Prior to that, she was VP Technology Alliances & Channel Sales for Identity Governance and Cloud Privileged Access Management leader Saviynt, responsible for AWS and Google Cloud partnerships. Clare's passion for cybersecurity includes her specializations in identity and privacy comprising: MFA, IGA, PAM, identity proofing, privacy-preserving authentication based on ZKP, identity theft, AML/KYC, and GDPR. Clare has held leadership positions at Novell, EMC2, Dell, and AllClear ID. She is a co-founder of C1ph3r_Qu33ns, an organization dedicated to cultivating and supporting the careers of women in cybersecurity. Clare is a second-generation yogi and technologist and has a degree in mathematics from Tufts University.

Change Log

Date	Change
2020-06-17	V1 published
2021-09-30	V2 published: Updated date of NY SHIELD act; added section on CIAM and Social Identity; added section title for CIAM and Workforce IAM; added Heather Flanagan as editor
2022-12-18	V3 published: Updated abstract; added notes re: threats to privacy in “What is Privacy?”; added information on eIDAS in “Beyond GDPR”

[i] “IDPro Body of Knowledge,” IDPro, https://www.idpro.org/body-of-knowledge/.

[ii] Cormack, Andrew, “Introduction to the GDPR (v2),” IDPro Body of Knowledge, 30 June 2021, https://bok.idpro.org/article/id/11/.

[iii] Hindle, Andrew, “Impact of GDPR on Identity and Access Management,” IDPro Body of Knowledge, 31 March 2020, https://bok.idpro.org/article/id/24/.

[iv] “Terminology in the IDPro Body of Knowledge,” IDPro Body of Knowledge, 30 September 2021, https://bok.idpro.org/article/id/41/.

[v] Solove, D., “The Myth of the Privacy Paradox,” SSRN e-Library, 24 February 2020, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3536265.

[vi] Narayanan, Arvind, and Vitaly Shmatikov, “Robust De-anonymization of Large Sparse Datasets,” The University of Texas at Austin, n.d., https://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf.

[vii] Narayanan, Arvind, Joanna Huey, and Edward W. Felton, "A Precautionary Approach to Big Data Privacy," 19 March 2015, https://www.cs.princeton.edu/~arvindn/publications/precautionary.pdf.

[viii] "'Anonymized' Data Can Never Be Totally Anonymous, says Study," The Guardian, 24 July 2019, https://cacm.acm.org/news/238352-anonymized-data-can-never-be-totally-anonymous-says-study/fulltext.

[ix] Ibid

[x] "Complete guide to GDPR compliance," Horizon 2020 Framework Programme of the European Union, https://gdpr.eu/.

[xi] "California Consumer Privacy Act (CCPA)," Office of the Attorney General, California Department of Justice, https://oag.ca.gov/privacy/ccpa.

[xii] "An act to amend the general business law and the state technology law, in relation to notification of a security breach," Senate Bill S5575B, The New York State Senate, 7 May 2019, https://www.nysenate.gov/legislation/bills/2019/s5575.

[xiii] "The California Consumer Privacy Act of 2018," Assembly Bill No. 375, Chapter 55, California State Legislature, 29 June 2018, https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375.

[xiv] "Data Protection Laws of the World," map, DLA Piper Intelligence, https://www.dlapiperdataprotection.com/

[xv] "Charter of Fundamental Rights of the European Union," Official Journal of the European Union, C 392/391, 26 October 2012, http://data.europa.eu/eli/treaty/char_2012/oj.

[xvi] "Treaty on the Functioning of the European Union," Official Journal of the European Union, C 326, 26 October 2012, ttp://data.europa.eu/eli/treaty/tfeu_2012/oj.

[xvii] United Nations, "The Universal Declaration of Human Rights," 1948, https://www.un.org/en/universal-declaration-human-rights/.

[xviii] "APEC Privacy Framework," International Association of Privacy Professionals (IAPP), n.d., https://iapp.org/resources/article/apec-privacy-framework/.

[xix] "The new EU ePrivacy Regulation: what you need to know," i-SCOOP, n.d., https://www.i-scoop.eu/gdpr/eu-eprivacy-regulation/.

[xx] "Richards, Neil, and Woodrow Hartzog, "The Pathologies of Digital Consent," SSRN e-Library, 11 November 2019, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3370433.

[xxi] Naughton, John, "'The goal is to automate us': welcome to the age of surveillance capitalism," The Guardian, 20 January 2020, https://www.theguardian.com/technology/2019/jan/20/shoshana-zuboff-age-of-surveillance-capitalism-google-facebook.

[xxii] Petters, Jeff, "Data Privacy Guide: Definitions, Explanations and Legislations," Varonis, 29 March 2020, https://www.varonis.com/blog/data-privacy/.

[xxiii] Maler, Eve, "Don't Pave Privacy Cow Paths: Retool Consent for the New Mobility" (video), Identiverse 2018, 26 June 2018, https://www.youtube.com/watch?v=eP5U2sA6EFk&t=254s.

[xxiv] Hulsen, L. Ten, "Open Sourcing Evidence from the Internet - The Protection of Privacy in Cvilian Criminal Investigations using OSINT (Open-Source Intelligence)", Amsterdam Law Forum, vol 12.2, 2020, https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/amslawf12&section=9.

[xxv] See for example Clausen, M.-L., 2021. Challenges of using biometrics in Yemen, DIIS: Dansk Institut for Internationale Studier. Retrieved from https://policycommons.net/artifacts/1526658/challenges-of-using-biometrics-in-yemen/2214896/ on 10 Nov 2022. CID: 20.500.12592/n9640f.

[xxvi] "Data Protection," European Data Protection Supervisor, n.d., https://edps.europa.eu/data-protection/data-protection_en

[xxvii] Westin, Alan F. "Privacy and freedom." Washington and Lee Law Review 25, no. 1 (1968): 166.

[xxviii] Cate, Fred H., Beth E. Cate, "The Supreme Court and information privacy," International Data Privacy Law, Volume 2, Issue 4, November 2012, p 255-267, https://doi.org/10.1093/idpl/ips024.

[xxix] Pernot-Leplay, Emmanuel, "Data Privacy Law in China: Comparison with the EU and U.S. Approaches," (blog post), 27 March 2020, https://epernot.com/data-privacy-law-china-comparison-europe-usa/.

[xxx] Member Economies map, Asia-Pacific Economic Cooperation, https://www.apec.org/About-Us/About-APEC/Member-Economies.

[xxxi] Sacks, Samm. "China's Emerging Data Privacy System and GDPR." *Washington, DC: Center for Strategic and International Studies* (2018).

[xxxii] Solove, Daniel J, "A Taxonomy of Privacy," University of Pennsylvania Law Review, Vol. 154, No. 3, January 2006, https://www.law.upenn.edu/journals/lawreview/articles/volume154/issue3/Solove154U.Pa.L.Rev.477(2006).pdf.

[xxxiii] Cavoukian, Ann, "Privacy by Design: The 7 Foundational Principles," www.privacybydesign.ca, n.d., https://www.ipc.on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf.

[xxxiv] "Privacy By Design," graphic, Aristi Ninja, n.d., https://aristininja.com/privacy-by-design/.

[xxxv] Cavoukian, Ann, "7 Laws of Identity: The Case for Privacy-Embedded Laws of Identity in the Digital Age," Information and Privacy Commission of Ontario, n.d., https://collections.ola.org/mon/15000/267376.pdf.

[xxxvi] Wadhwa, Vivek, "Laws and Ethics Can't Keep Pace with Technology," MIT Technology Review, 15 April 2014, https://www.technologyreview.com/s/526401/laws-and-ethics-cant-keep-pace-with-technology/.

[xxxvii] ISACA webinar, Robotic Process Automation (RPA) and Audit, March 19, 2020, https://www.isaca.org/education/online-events/lms_w031920

[xxxviii] "The UX Guide for Getting Consent," IAPP, n.d., https://iapp.org/store/books/a191a000002FUZKAA4/.

[xxxix] Schwartz, Karen D., "Data Privacy and Data Security: What's the Difference?" ITPro Today, 2 May 2019, https://www.itprotoday.com/security/data-privacy-and-data-security-what-s-difference.

[xl] "NIST Privacy Framework: A Tool for Improving Privacy Through Enterprise Risk Management, Version 1.0," National Institute of Standards and Technology, U.S. Department of Commerce, 16 January 2020, https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.01162020.pdf.

[xli] Ibid

[xlii] "NIST Privacy Framework: A Tool For Improving Privacy Through Enterprise Risk Management," Preliminary Draft, National Institute of Standards and Technology, U.S. Department of Commerce, 6 September 2019, https://www.nist.gov/system/files/documents/2019/09/09/nist_privacy_framework_preliminary_draft.pdf.

[xliii] "The Privacy Advisor," IAPP, Vol. 10, No. 10, December 2010, https://iapp.org/media/pdf/publications/Advisor_12-10_print.pdf.

[xliv] "White Paper: GDPR, CCPA, and Beyond: How to Comply with Data Privacy Laws and Improve Customer Trust," Akamai, n.d., https://www.akamai.com/us/en/campaign/assets/whitepapers/gdpr-ccpa-and-beyond-wp.jsp.

[xlv] "Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions," European Commission, COM(2020) 66 final, 19 February 2020 https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020DC0066&from=EN.

[xlvi] European Commission, "eIDAS Regulation," website, last updated 7 June 2022, https://digital-strategy.ec.europa.eu/en/policies/eidas-regulation.

[xlvii] European Commission, “Commission proposes a trusted and secure Digital Identity for all Europeans,” press release, 3 June 2021, https://ec.europa.eu/commission/presscorner/detail/en/IP_21_2663.

[xlviii] European Commission, COM(2020) 66 final, 19 February 2020.

[xlix] “Blockchain and the GDPR: Solutions for a responsible use of the blockchain in the context of personal data,” CNIL, 6 November 2018, https://www.cnil.fr/en/blockchain-and-gdpr-solutions-responsible-use-blockchain-context-personal-data.

Workforce IAM

An Overview of the Digital Identity Lifecycle (v2)

By Andrew Cameron and Olaf Grewe

Table of Contents

Abstract

A digital identity goes through several stages during its existence, from creation, through various modifications in response to different events, to inactivation or deletion. This article walks through the types of digital identities that must be managed, along with the various stages of a digital identity, describing the typical beginning-to-end lifecycle within or across multiple systems. The lifecycles outlined in this document are not meant to be comprehensive but should be applicable over most B2B, B2C, and B2E use cases.

Introduction to Digital Identity

A digital identity, for the purpose of this document, is defined as the combination of a unique identifier together with relevant attributes that uniquely identifies an entity. Depending on the complexity of the environment in which a digital identity is used, its lifecycle—from its inception to its closure—can be significantly more complicated than a simple create, read, update, and delete (CRUD) lifecycle.[i]

Depending on the type of identity (human such as Workforce or Customer, and non-human types such as System or Device), the lifecycle phases will differ. Enterprise IAM has typically been a well-established set of processes that provide the processes and governance capabilities to ensure only the correct people (via their accounts) have access to only the required applications (resources). Customer IAM has an entirely different set of requirements that represent value to a business due to the nature of its defining interactions with a customer. Poor or inefficient interactions with customers can have severe negative effects on a business. For these reasons, the different identity types will require separate systems and processes supporting them:

Identity Type	Description
Workforce	A workforce identity is one created to function in an enterprise context, which may include a Business-to-Business (B2B) and/or Business-to-Employee (B2E). Examples of these identity types will be Employees, Suppliers, Contractors, or other human identities that support the corporate workforce.
Customer	A customer identity type will usually function outside the enterprise context, enabling digital business between the owner of the customer identity and the enterprise. Typically, there will be multiple channels (Web, Mobile, IoT Device) of access to manage with a larger set of profile (identity attribute) data necessary to facilitate the interaction.
Device or System	Device identities typically are used to provide identification and representation on a digital network. System identities are used to authenticate services (e.g., applications or server-based processes) to a network.

Terminology

- Digital Identity – the combination of a unique identifier together with relevant attributes that uniquely identifies an entity.
- Journey-based Creation – The process that guides a customer through a series of interactions prior to establishing a digital identity. For example, capturing the minimum basic information needed from a customer to enable creation of an identity.
- Attributes - Key/value pairs relevant for the digital identity (username, first name, last name, etc.).
- Inter-organizational (Federation): An organization relies on another organization's digital identity and lifecycle management processes.
- Intra-organizational (Single Sign-On): A central digital identity, such as an account in a directory, is linked by downstream systems as authoritative for authentication.

Identity Lifecycles

For any lifecycle 'create' phase, a digital identity is created as a unique identifier in a system of record. It can be created either as part of a business process (workforce or device identity) or transparently as part of a user journey (customer identity).

Throughout its lifecycle, a digital identity enables digital transactions through all of its assigned accounts and the entitlements assigned to those accounts. Although a lifecycle is outlined as a continuum in this document, the reader should expect that:

- The digital identity lifecycle could be distributed across multiple technical solutions in most organizations.
- Some steps in the lifecycle (e.g., authenticate, use) will occur more frequently than others (e.g., merge, delete).

Workforce Identity

The workforce identity lifecycle is addressed through three principal business processes: Joiner, Mover, or Leaver. The **Joiner** processes cover all lifecycle phases that facilitate the creation of assets (identities, accounts, group memberships, etc.) to enable identification and access in an enterprise environment. The **Mover** process allows for changes or updates to identity status while still engaged in the enterprise environment and considers the necessary attestation processes to verify access permissions and entitlements. The **Leaver** process covers the series of steps that must occur when an identity is removed from access to the enterprise environment.

Figure 1 depicts the workforce IAM phases in the process:

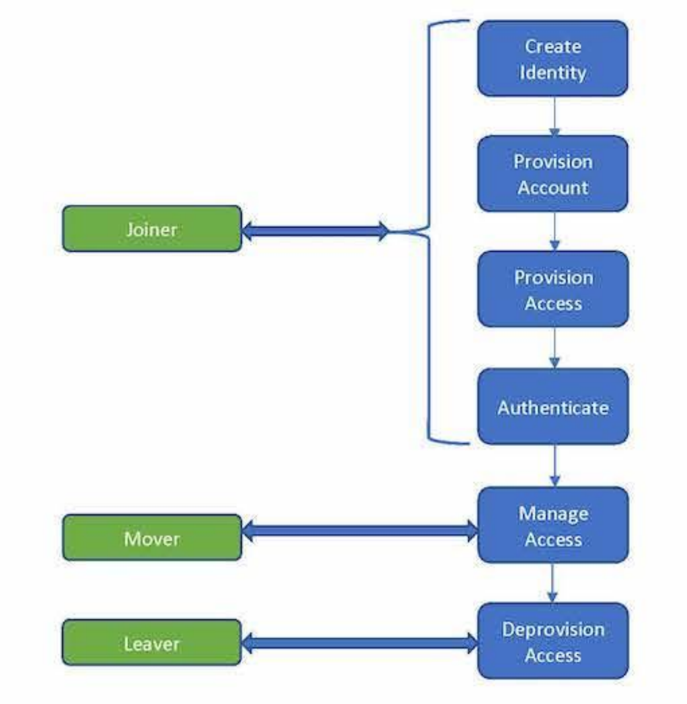


Figure 1 – Core IAM Processes

The following table describes the phases that support the workforce identity lifecycle:

Lifecycle Phase	Description
Create Identity	The creation of a workforce identity as part of a business process (employees, suppliers, etc.) is frequently combined with the collection of proof to establish a minimum set of attributes to be associated with the identifier. The creation of a digital identity may be automated (e.g., synchronized with an HR system event), especially when digital identities are generated at scale for various purposes, such as a merger or acquisition. Enrollment processes for workforce entities frequently involve other human entities (such as a line manager or delegated admin agent) validating the proof provided. In countries without an established national identity system (US, UK, AU, etc.), it can be required to provide multiple documents as proof (driver's license, passport, utility bill, bank card/statement) in lieu of a national identity document.
Provision Account	Create accounts in enterprise systems based on business rules and required access to resources.
Provision Access	Create entitlements by associating user accounts to objects that enable access to corporate resources in the required systems. Entitlements are generally represented by attribute values, group memberships, or organizational alignment. Business rules will define access to a resource based on enterprise entitlements.
Authenticate	Require a user account to validate a credential before allowing access to a network or resource.

Manage Access	Validate that the access that has been assigned an account and approving continued access to corporate resources. Access certification is a process that validates all current access and can be used to remove no longer needed access. The attestation process for verifying and access is a critical and often underestimated component of a mature IAM system. Digital identities are frequently subject to updates, primarily of their attributes. Less frequently, the identifier itself may change. An example is a digital identity for which the username is also used as the identifier (e.g., email address). A user may wish to change their username for various purposes, such as a name change due to a life event or a change of preferences. For an in-depth discussion, please refer to Ian Glazer's article, "Identifiers and Usernames."[ii] Frequently update the use cases describing workflow capabilities that address approval, step-up, or notification requirements. These are important controls to address identity take-over risks. Depending on the value of the digital identity for the organization, updates to digital identities may be subject to enrolment-type proofing.
Deprovision Access	Remove access to any or all corporate resources. The need to remove access could occur as a result of a Leaver process or a validation from an Access Certification. When a digital identity is not required anymore, it should be disabled in the system of record. This action implies not only disablement or deletion from a central directory but also downstream systems that maintain records associated with this digital identity as well as logging and auditing repositories. Only once the identifier used for this digital identity has been removed from all systems can a digital identity be considered genuinely deleted. A detailed discussion on the importance of account disable or removal given current best practices can be found in Andrew Hindle's article, "Impact of GDPR on Identity and Access Management."[iii]

Customer Identity

Customer IAM has evolved more recently to support the processes that govern consumers' User Experience as they interact with digital business. CIAM solutions have developed to provide companies with added value from the data they collect from customers as a result of the customers' experiences with corporate websites and services. Most customer experiences are described as part of a "User Journey," which represents the interactions

(Authentication, Registration, Profile Update) that a customer has when engaging with digital resources such as websites, mobile apps, or IoT interfaces.

The following diagram depicts the phases of the CIAM Lifecycle.

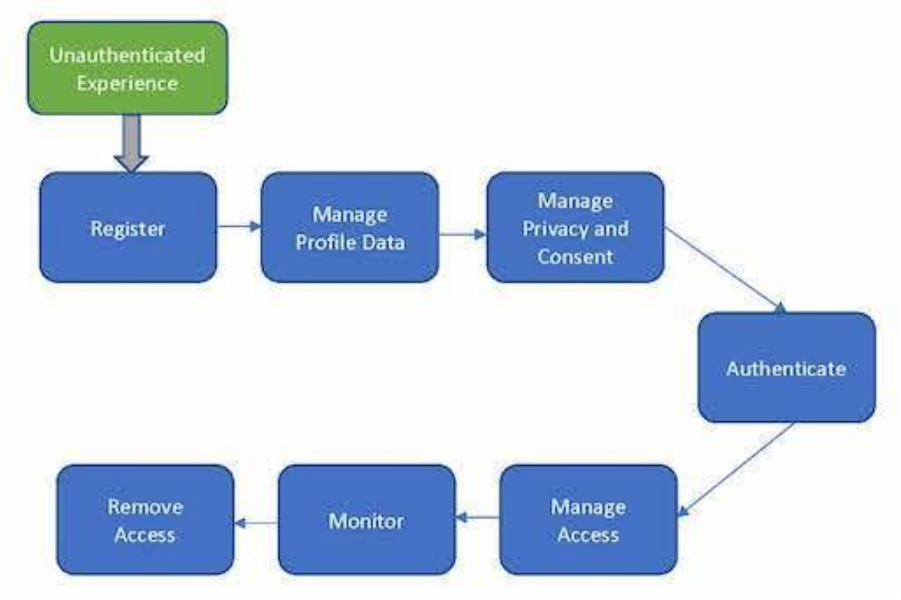


Figure 2 – The Customer Identity Lifecycle

The following table describes the phases that support the customer identity lifecycle:

Lifecycle Phase	Description
Register	The first part of the user journey is the creation of a customer identity through a registration process. This registration typically happens where a digital identity is required to enable an experience. Information is captured from a user as part of a user journey, and the user is allowed to consent to usage of the data provided. Registration interactions are typically a one-time interaction with the customer that concludes with a confirmation of the purpose of the flow (i.e. "Your account has been created"). Registration interactions can also be transparent to the user if enabled thru a federated identity such as a social account sign-in (i.e., "Sign in with your Facebook account to get registered"). Registration does not require mandatory attributes other than the linking steps in the user journey to the identifier. Depending on the nature of the digital transaction, customer identities may require assurance over several attributes. A key consideration here is the attributes used to establish ownership (or recovery) for a digital identity, either via human or non-human means.
Manage Profile Data	Each customer has a profile and managing the profile data involves a user experience that allows a customer to update their data across corporate resources (e.g., websites or mobile apps). This phase primarily applies to user journey-based digital identities. In order to enable digital services to resume user journeys, it is necessary to enhance the digital identities with attributes that are specific to the way the user accesses the service. Two common techniques are cookies or device fingerprinting. For an illustration of the latter, see the EFFs Panopticlick site.[iv]
Manage Privacy and Consent	The customer lifecycle must include a process that informs and enables the customer to invoke their rights around knowledge and consent of what can happen with their customer information.
Authenticate	As part of the workflow, the customer is required to validate their credential prior to accessing any customer services

Manage Access	The customer lifecycle will require managing access to business services based on customer interactions. The user may also choose to provide additional attributes. The service would typically allow the user to create a username and password to login after their current session has expired. At this stage, a service may be able to combine multiple identifiers created by different devices (mobile, desktop, laptop, etc.). At this stage, the digital identity is considered pseudonymous as there is no assurance over the attributes provided by the user.
Monitor	After the initial phases are complete, the customer lifecycle will move into monitoring, where the process of mining/collecting data about the customer and their experiences support a variety of business and consumer requirements occur. From a security perspective, monitoring data can be used to notify the customer of leaked credentials or other breaches of information. The business can also benefit by leveraging historical usage information of customer activity thru an analytics service.
Remove Access	Removal of customer access is typically done as a result of a customer request or based on some amount of inactivity measure.

Device or System Identity

A device or system identity is an evolving area in that devices are being enabled with increasing levels of technological capability, which increases the need to identify and manage them through a lifecycle. For example, cars have dozens of internal systems that require sophisticated management capabilities over the life of the vehicle identity. On the other end of the scale, some simple monitors can connect to a network and only provide a temperature value or some other basic information. All devices will need specific lifecycle phases to manage them based on their capabilities.

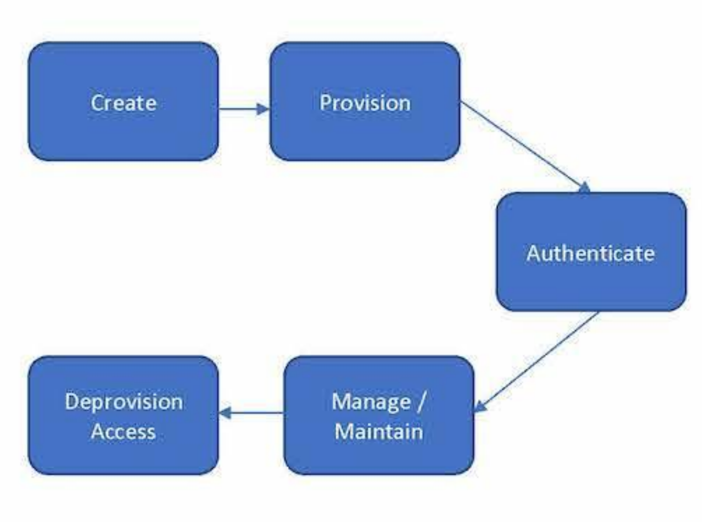


Figure 3 – The Device Identity Lifecycle

The following table describes the phases in a simple model that support the device identity lifecycle:

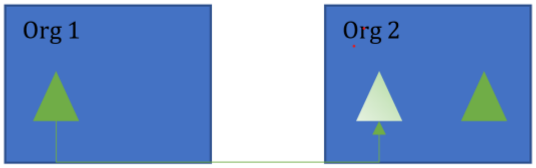
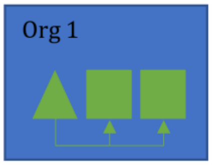
Lifecycle Phase	Description
Create	The first stage in the device or system lifecycle is to kick off the process of creating the identifier that will be assigned to the device or system.
Provision	When the identifier is assigned, the process of enabling the device or system to be recognized, monitored, and managed. Device provisioning is typically done using some sort of certificate or PKI infrastructure to ensure that only known devices can interact with corporate resources.
Authenticate	Device or system authentication typically is done using a PKI infrastructure that ensures that the connected device is known and allowed to interact with the network.
Manage / Maintain	Once the initial phases are complete, the device or system must be monitored to determine if any actions are needed to maintain the device. As an IT security best practice, credentials (passwords) associated with non-human identities should be rotated on a periodic basis to enable protection against brute force password-based attacks.

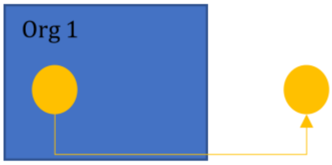
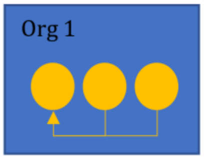
Deprovision Access	When the device or system is no longer in use (which may require different processes than workforce or customer digital identities to determine), remove access of the device or system from the system of record, disabling any access to the corporate network.

Other Digital Identity Relationships

Some digital transactions require an organization to establish relationships between digital identity issuers, also known as identity providers. These relationships may be with external partners (e.g., a B2B relationship) or across various enterprise applications (e.g., a single sign-on environment). In addition, digital identities may be related to other identities within an organization to establish delegation authority or to manage dual-control requirements. In all cases, relationships are typically managed either as attributes of the digital identity (e.g., identifiers for the allowed services) or as separate data points in a central directory (e.g., membership in an LDAP group).

Common types of relationships are:

Inter-organizational (Federation)	 Inter-organizational An organization relies on the digital identity and lifecycle management processes of another organization.
Intra-organizational (Single Sign-On)	 Intra-organizational A central digital identity, such as an account in a directory, is linked by downstream systems as authoritative for the purpose of authentication.

Inter-entity (Delegation)	 Inter-entity Delegation involves assigning a subset of authority from an identity in one business domain to an identity that resides in another business domain. In this example, business domain refers to defined boundaries that exist within or across an entity, which enables policy enforcement to occur. Examples of business domain include company, organization (within a company) or even work teams (within an organization). Authority is granted across domain boundaries for the purpose of enabling the transactions within the scope of a policy. Authority can be granted either explicitly or based on business rules (policies) defined at the domain level.
Intra-entity	 Intra-entity Either user-driven or out of organizational requirements, a relationship is established between multiple digital identities to identify a single human or non-human entity as the owner (see Enhance above).

Conclusion

The complexity of the digital identity lifecycle frequently becomes apparent only after a number of years and as more functionality gets added to systems. Therefore, it is advisable to approach life cycle requirements with a longer-term horizon and ensure user management capabilities are extensible.

Acknowledgements

The author would like to acknowledge Ian Glazer for articulating the progression of an identity from anonymous to pseudonymous and known. Dean Saxe contributed the classification of relationships. Jon Lehtinen, and Heather Flanagan contributed encouragement and suffered through early drafts of the article.

Change Log

Date	Change
2022-02-28	Updated definition of digital identity; clarified the use of the term 'lifecycle'; updated diagrams; updated Delegation description
2020-10-30	V1 published

[i] "Create Read Update Delete" ldapwiki.com, paged last modified 19 March 2020, https://ldapwiki.com/wiki/Create%20Read%20Update%20Delete.

[ii] Glazer, Ian, "Identifiers and Usernames," IDPro Body of Knowledge, 31 March 2020, https://bok.idpro.org/article/id/16/.

[iii] Hindle, Andrew, "Impact of GDPR on Identity and Access Management," IDPro Body of Knowledge, 31 March 2020, https://bok.idpro.org/article/id/24/.

[iv] "Panopticlic 3.0," Electronic Frontier Foundation, viewed 13 April 2020, https://panopticlick.eff.org/.

User Provisioning in the Enterprise

By Ian Glazer, Lori Robinson, and Mat Hamlin

Table of Contents

Abstract

User provisioning is the means by which user accounts are created and maintained in a system (e.g., database, SaaS app, operating system, etc.). When we say that a user-provisioning system maintains a user account, we mean everything from changes to attributes in the user account, changes of entitlements or privileges associated with the user account, locking and unlocking the user account, and even deletion of the user account. User provisioning is primarily an admin-time affair: a user account is created (or changed) based on an administrative action as opposed to a user's action at the time of resource use. This article explores the uses and components of a user-provisioning system and focuses mainly on situations where user accounts are maintained in central repositories, typically enterprise and workforce settings.

Introduction

Creating and managing user accounts is the bedrock of any IAM system. The process is generally referred to as user provisioning and is used to establish the entitlements a user is given to access restricted resources (applications, documents or databases) maintained by the organization. User provisioning processes not only create user accounts and assign entitlements but also maintains those user account entitlement through the detection of meaningful lifecycle events such as changes to job responsibility and the application of policies to ensure. User provisioning is often used to ensure the right people have access to the right systems in a timely fashion and with entitlement appropriate for their responsibilities.

Terminology

- **Authoritative source(s):** The system of record (SOR) for identity data; an organization may have more than one authoritative source of data in their environment.
- **Entitlement catalog:** A database of entitlements and their related metadata. The catalog includes an index of entitlement data pulled from business systems, applications, and platforms, as well as technical and business descriptions of the entitlements or their use
- **Identity lifecycle management:** A process that detects changes in authoritative systems of record and updates identity records based on policies.
- **Identity repository:** The identity repository is a directory or a database that can be referenced by external systems and services (such as authentication or authorization services).
- **Reconciliation:** The process of identifying and processing changes to users and user access made directly on target systems.

- **User provisioning**: the means by which user accounts are created, maintained, and deactivated/deleted in a system according to defined policies.

What is User Provisioning?

User provisioning is setting up the entitlements for users to the resources to which they need access. User-provisioning technologies are deployed across multiple industries, including healthcare, education, financial services, government, retail, manufacturing, technology, etc.

Functionality supported by user-provisioning technologies include:

Identity lifecycle management: An identity and its associated attributes are the basis for authentication and authorization decisions made in an environment. It is, therefore, essential that the identity record is maintained. Provisioning systems detect changes in authoritative systems of record (such as a Human Resources database/repository) and update the identity record accordingly.

User account provisioning: As the name suggests, a user-provisioning system's primary function is user account provisioning (and de-provisioning). User-provisioning technologies automate the creation, maintenance, and deactivation/deletion of user accounts on target systems according to defined policies.

Self-service and delegated administration: User-provisioning systems provide interfaces that allow users to request access to systems, manage passwords and update their data. Delegated administrators can perform similar tasks such as onboarding and off-boarding users, password changes, profile updates, and entitlement assignments on behalf of others.

Workflow: Provisioning systems employ workflow tools that allow for the automation of provisioning processes and approval workflows. Using automated approval workflows, business stakeholders can validate and approve proposed changes before they are applied to target systems. While many decisions to grant access are automated through policy, others may require human intervention.

Audit and Reporting: Provisioning systems log all identity lifecycle management, access policies, and user provisioning transactions and provide reporting mechanisms to extract the logged data.

A note about governance: User-provisioning systems are often packaged with identity governance capabilities such as access review and certification, risk analysis, and identity analytics. The combined user provisioning and identity governance solutions may be referred to as Identity Governance and Administration (IGA). This document focuses exclusively on user-provisioning functionality and does not include identity governance information. Similarly, password management, which may be packaged with provisioning solutions, is not covered in this document.

Business Drivers for Automated User Provisioning

Three primary business drivers justify the deployment of automated user-provisioning systems:

- **Operational efficiency**: The amount of administrative overhead associated with the manual creation and maintenance of user accounts is significant for medium- to large-size organizations. Without an automated process, it may be weeks before a user has access to the resources they need to perform their job duties or other tasks. User-provisioning systems automate the user account management process, reducing administrative overhead and improving time to productivity, resulting in operational efficiency.
- **Security:** Manual provisioning of user accounts may lead to security gaps such as overprivileged user accounts or orphaned accounts (active accounts assigned to inactive employees). Automated user account provisioning systems improve security by ensuring that user accounts and entitlements are provisioned according to policy and deprovisioned in a timely manner.
- **Compliance**: Various laws and regulations require organizations to demonstrate control over access to critical systems, resources, and data. User-provisioning systems enforce policy-based access controls and allow organizations to demonstrate the efficacy of these controls with reporting and attestation capabilities.

User Provisioning Logical Architecture

User-provisioning systems employ policies, workflows, and connectors to synchronize identity data from an authoritative system to an identity store and to provision user accounts to target applications.

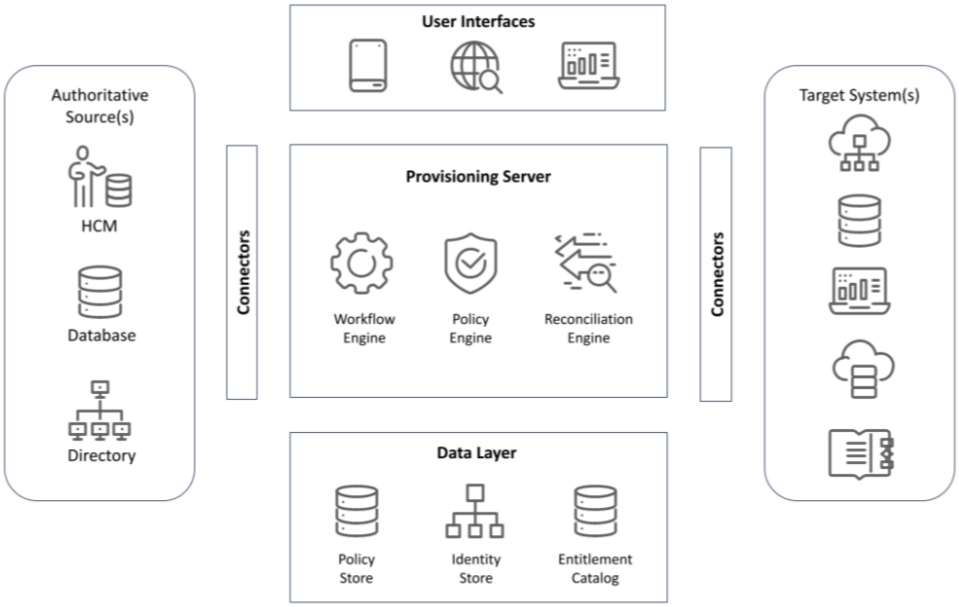


Figure 1: illustrates the standard architectural components of a user-provisioning ecosystem.

Authoritative source(s): The system of record (SOR) for identity data. The authoritative system publishes changes to the provisioning system. There may be more than one authoritative system in the environment. For example, in workforce use cases, the Human Capital Management (HCM) / Human Resources (HR) system may be the SOR for employee data, but contractor data may be stored in a procurement system.

Target system(s): Target systems subscribe to changes to identity records and are on the receiving end of the provisioning process. The provisioning system creates and manages user accounts and associated entitlements within the target system environment.

Connectors: The integration layer between the provisioning system and authoritative and target systems. There are various types of connectors: proprietary (application-specific connectors that communicate with app-specific APIs), generic (e.g., LDAP, JDBC, delimited text), or standards-based. See the standards section for more information on standards-based connectivity.

Provisioning server: The middleware layer responsible for data synchronization, mapping, and transformation; the application of business logic and access policies; and the orchestration of provisioning process flows. The provisioning server is comprised of the following functional components:

- **Account correlation rules:** correlates or matches disparate user accounts (in target or authoritative systems) with a single identity record and ensure that duplicate identities for a single person/entity are not created.
- **Data mapping rules:** maps and transforms data from the source context to the target context.
- **Account creation rules:** establishes standards for creating an identity record such as naming conventions, required attributes, password policy, location policy, etc.
- **Access policies**: determines access rights and entitlements that should be assigned to a user. See the Policies section for information on different types of access policies.
- **Workflow engine:** orchestrates the provisioning based on business processing logic and enables access request, approval, and review workflows.
- **Reconciliation engine:** discovers user accounts created directly in target systems (circumventing standard provisioning processes), ensures that the user account is in compliance with access policies, and correlates the user account with the individual's identity record.

Identity repository: Identity records are stored in an identity repository. The identity repository is a directory or a database that can be referenced by external systems and services (such as authentication or authorization services). The identity record includes attributes associated with the identity and a record of all user accounts associated with the identity.

Entitlement catalog: A database of entitlements and their related metadata. The catalog includes an index of entitlement data pulled from business systems, applications, and platforms. The entitlement data can be enriched with metadata such as risk scores and business-friendly descriptions of entitlements that can be displayed to users during access requests, access reviews, and certifications.

System configuration and audit store: A dedicated repository to hold information such as system configuration, identity mapping, policy, role definition, and workflow data. This repository may also serve as the store for audit logs.

User interfaces: User-provisioning systems include administrative, end-user, and delegated administration interfaces. Administrative interfaces are used for the set-up and configuration of the system. End-user and delegated administration interfaces are used for access requests, approval workflows, reporting, profile updates, etc. Provisioning systems typically include web-based interfaces that can be accessed from a pc or mobile device. While not standard, some provisioning vendors offer a mobile app for self-service and approval workflows.

Given their highly connected and interactive nature, user-provisioning systems must be open and extensible. The provisioning provider should provide open APIs, no or low code workflows, and generic connectors that allow for flexibility in the system.

User Provisioning Process Flow

User-provisioning technologies allow organizations to efficiently manage thousands of identities by capturing lifecycle events and ensuring that user accounts and their associated privileges are kept up-to-date and accurate. These processes reduce administrative overhead and improve security. That said, automated user account provisioning is a complex, multifaceted process that includes three distinct phases:

- **Event trigger**: A business event or a change to an identity that triggers a provisioning action
- **Policy administration**: Application of access policies that bind the identity to specific user accounts and entitlements
- **User account provisioning**: Creation, maintenance, deactivation, or deletion of user accounts in target applications

Event Trigger

The act of provisioning begins with an event. Such an event could be:

- The creation of a new employee in an HR system.
- A modification to an entry in Active Directory moving a person from one business unit to another.
- A ticket being created in an IT Service Management (ITSM) or Help Desk ticketing system.
- A person directly interacting with the provisioning system to request a change to a user account.

There are three primary types of events:

- Join
- Move
- Leave

Joiners, Movers, and Leavers (JML) are grist for the user provisioning mill. Managing JML processes becomes the work of an identity system. This work includes connecting the user-provisioning system to trigger sources and then constructing policies to be evaluated for each event type for each target system.

Join

The easiest way to think about the Join event is when a new employee joins a company. She needs her benefits and payroll set up along with user accounts in IT systems. In its purest sense, Join events are meant to create a net new identity and net new user accounts in IT systems.[i]

Move

When a person changes roles with an enterprise, she likely needs access to new business systems and to have access to her older ones removed. This is the purpose of the Move event. You can think of Move as a change in the relationship between the organization and person. They might change which business unit they report to, get promoted, or change their last name.[ii]

Leave

A person retiring is the simplest example of a Leave event. In such a case, the person's user accounts need to be deleted or at least locked to prevent further use in all target systems. Another example is when a contractor's project concludes and the contract ends. The story is the same; a Leave event triggers the user-provisioning system to remove their access.

User-provisioning technologies provide various mechanisms to capture JML events, including:

- **Automated provisioning:** The user-provisioning system "listens" for events from systems of record such as Human Resources, ITSM, or a directory.
- **Batch processing:** The provisioning system executes a regularly scheduled process that polls an authoritative source for changes and generates an output file.
- **Self-service request:** Today's user provisioning solutions include an end-user access request portal where end-users or managers can request access to specific systems and rights needed to perform their business responsibilities. The user or a delegated administrator updates the user profile or makes an access change request via the self-service interface.
- **Manual/Ticket:** In some instances, an organization may use a ticketing system or other manual process to notify the identity team of a change needed on the identity record. In this case, the identity administrator would update the identity record directly to trigger downstream policy and provisioning activities.
- **Reconciliation event:** Reconciliation is the process of identifying and processing changes to users and user access made directly on target systems. When an organization configures a user provisioning solution for centralized management of user access, that does not prevent changes from occurring directly on a target system. So, to ensure the consistency of user access and user attributes across the organization, the user-provisioning system will periodically ***reconcile*** what it knows about users and their access to a specific target system. This reconciliation is accomplished by gathering and comparing all user data on the target system (full reconciliation) or by processing known changes to user access based on a changelog or other time-based query. When changes or variances are identified in a reconciliation process, events are triggered and processed based on defined

policies. The result of reconciliation could be to synchronize changes from the target system to other systems, or it could be to roll back any locally applied changes that occurred outside of the user provisioning solution.

Policy Administration

In the past, organizations have managed users' access to target systems in an ad hoc manner; given the complexities of the enterprise environment, this is no longer viable. They need documented rules to determine who should have access to which target systems; furthermore, they need to control what kind of entitlements and privileges people have in those target systems. This is the role of policies in a user-provisioning system. Instead of leaving the details to an administrator to determine which groups a user account ought to be a member of, a policy can describe which groups are required, optional, and even forbidden for people to be a member of.

A policy can be thought of as a way to bind groups of people to groups of target systems with groups of related access (entitlements, privileges, etc.) In this way, there are always two components of user provisioning: Who and What. The Who portion of the policy describes the inclusion criteria for which people the policy will apply. For example, all full-time employees, contractors, and finance people are all examples of the Who portion of a policy. The What portion of the policy describes the user accounts and associated entitlements and privileges a person gets. The What can be very coarse-grained, for example, the creation of a user account in all target systems, to very fine-grained, as when this specific entitlement and these two specific privileges are created in the target system.

Different kinds of policies use different combinations of Who and What to help identity practitioners govern access. While the overall topic of policies can be extremely broad, for the purposes of this article, let's focus on four kinds of policies:

- Birthright
- Role-based
- Segregation of duties
- Workflow approvals

It is important to note that a user provisioning process won't have just a single policy for a target system or event. Policies can be combined and applied to multiple target systems and triggering events.

Birthright

There are specific systems and entitlements that often broad swaths of the organization need; this kind of access is considered a birthright. Examples of such policies include:

- All full-time employees need email, calendar, collaboration, and file sharing.
- Everyone in the Finance department needs at least minimal access to the financial reporting system.

- Interns need access to the ‘Intern Team Excellence’ collaboration channel.

Birthright policies can be thought of as defining access that is fundamental to certain kinds of people who have a relationship with the organization. Such access does not need additional scrutiny, review, or approval; simply by being a person who matches the criteria of the policy (such as being a member of the Finance department) that person is allowed to have and will get user accounts in certain systems with specified levels of access (as managed by their user account’s associated entitlements and privileges.) More often than not, birthright policies grant coarse-grained access to target systems; that is to say, they might only give someone a user account in an email system but not necessarily access to specific distribution groups. Birthright policies are most commonly applied as part of a Join event and typically occur by assigning one or more business roles. Birthright events can also happen as a part of a Move event, specifically when a person moves from one business function to another. For example, when a person is hired into the Accounting division within an organization, they’d receive birthright access to things like email and the productivity suite and basic access to critical accounting systems. When that person then transfers to Corporate Strategy, they lose access to the account systems, gain access to budget forecasting systems, and continue to retain access to email and productivity tools.

Role-based

Because an organization can have many business functions and thus lots of different business responsibilities as well as tens of thousands of individual entitlements in their systems, there is no way to manage who gets access at an individual level. Trying to do so would quickly lead to the management of tens of millions of combinations of people and privileges. User-provisioning systems attempt to bring order to this chaos by using roles to aggregate people and entitlements into more manageable policy components.

Much is made of roles in identity management.[iii] Roles can come in a variety of flavors; this article focuses on business roles and technical roles. A business role is a way to aggregate people who share the same business responsibilities. For example, a retail banking organization might have a business role called “Teller” and use it to describe the access appropriate for people who work as tellers. The second kind of role we’ll need to understand for this article is a “Technical” role. A technical role is a way to aggregate the entitlements and privileges required within one or more target systems to perform a task. For example, the same retail bank could have a technical role called “Check and Update Balances,” which gives user accounts in their systems the ability to check and update savings account balances.

A role-based policy in a user-provisioning system uses business and technical roles to govern access for more specific sets of people, entitlements, and privileges in target systems than birthright policies do. For example, a user-provisioning system might have a birthright policy that gives all full-time employees access to email. An additional role-based policy might grant Tellers access to a specific mailing list and shared drive.

Segregation of Duties

Stemming from the fallout of the WorldCom and Enron accounting scandals, the Sarbanes-Oxley Act had a profound impact on business practices.[iv] These impacts made their way to user provisioning. As part of compliance activities, organizations looked to their user provisioning policies to not only grant access to people but also prevent “toxic combinations” of access. A toxic combination of access is one in which a person has privileges that could enable some form of fraud, such as the ability to create a new vendor and issue a payment to that vendor. This combination of access would allow a bad actor to create a fictitious company in the financial system and then divert monies to that company. Another application of a toxic combination policy is to prevent anyone who isn’t a system administrator from having system admin or highly privileged entitlements.

If roles are a way of describing what someone should have, then segregation of duty (SoD) policies are a way of describing what they must not have. Such policies are typically evaluated when a provisioning event is triggered so new toxic combinations are not introduced into target systems and existing ones are detected and remediated.

Workflow Approvals

Workflow approvals are an essential component of the policy management toolbox. Organizations with a mature provisioning deployment may only auto-provision 70-80% of access using birthright rules, roles, or SoD. So how does the remaining 20-30% of access get provisioned? The answer is self-service access request and approval workflows.

Workflow approvals are used when a human needs to make a policy decision. If a rule or role is not available, the provisioning system invokes a workflow process that routes an access request to a designee for approval. For example, an employee may make a self-service access request that is routed to a line manager for approval. The workflow approval process applies a layer of control and documents the access policy decision.

User Account Provisioning

Once a provisioning event has been triggered and policy evaluated to determine what user account attributes, entitlements, and privileges need to be set or changed, then that information needs to make its way into the target system to affect the change to the user account stored locally there. How the necessary changes are made in the target system is the act of provisioning. Provisioning can be accomplished in two primary ways:

- Automated
- Manual

Automated

Automated user account provisioning is the process of creating and maintaining a user account in the target system using automated processing. To automate the user account provisioning process, the target system must provide a user management API or other means for the user-provisioning solution to systematically create, manage, and deactivate/delete user accounts.

Automated user account provisioning in target systems is the ultimate goal of user-provisioning technologies, but it is not without challenges. Each target system is an island. The user-provisioning system must maintain connections to the various target systems, which can be a heavy lift.

Manual

Manual provisioning requires human intervention to affect the change to the user account in the target system. This intervention often takes the form of the details of a user-provisioning event being sent to a team or a person who takes that information and manually keys it into the target system, using the target system's unique user management interfaces. The information required could be sent via email or work ticket.

Manual provisioning introduces humans into a critical step of user provisioning, creating two specific risks. First, the person who manually works with the target system to create and change user accounts, by definition of the work she does, is a highly privileged user. It is a good practice to minimize the distribution of such privileges, but sometimes it is necessary. The second risk, manual provisioning, introduces is the possibility of human error. The person might misread or mistype an attribute, entitlement, or privilege, thus incorrectly setting the user account in the target system. While that might result in a minor annoyance, such as misspelling a user's name, it might also lead to the assignment of incorrect privileges or even a toxic combination of entitlements.

It is fair to ask why manual provisioning is needed or wanted, given such risks. Manual provisioning is needed because not all target systems have APIs to which automatic provisioning connectors can connect. That homegrown general ledger system running on an extremely old operating system is an example of such. Another example is situations in which the target system is actually managed by a managed service provider and the identity team does not have direct access to that service provider. In that case, the change to a user account needs to be sent to the managed service provider via an email or ticket to trigger them to make the necessary change.

Manual provisioning is wanted because automation isn't worth the effort. Consider an application with very few users, entitlements, or changes required, or all three. An identity team may decide that it is not worth deploying (or possibly building) an automatic provisioning connector but instead choose to accept the human cost and risk of manual provisioning. It is a best practice to apply automated provisioning to high volume (lots of users), high velocity (frequent changes to user accounts), and high value (mission-critical, financial material, etc.) systems. Conversely, it is not a best practice to automate every single system in the enterprise because, eventually, the costs to maintain connectors are simply not worth it.

The Role of Standards

The identity industry recognized that the proliferation of proprietary user management APIs would lead to a lack of automated provisioning and make it difficult for organizations to mitigate the risks inherent in manual user provisioning. Starting with Directory Service Markup Language in 1999, followed by Service Provisioning Markup Language[v] (SPML) in 2003, and finally followed by System for Cross-domain Identity Management[vi] (SCIM) in 2011, the industry has produced standards. The latest version of SCIM, version 2, has had significant uptake and, as of the second half of 2021, signs that the standards community is interested in making further enhancements. The fact that there have been at least three different standards with multiple versions is a testament to both the challenge of building a viable standard and the changes in the application development world.

For a user provisioning standard to be considered successful, it requires adoption from both user-provisioning system providers and application vendors. This “it takes two” challenge had thwarted mass adoption, especially in the era of on-premise software. The era of cloud computing and SaaS has seen a marked increase in the number of service providers willing to use SCIM v2 as well as user-provisioning technology providers. If the reader’s IT organization is building custom applications, it is worth investigating the implementation of SCIM v2 in those apps to facilitate automated user provisioning, especially in high-volume, high-velocity, and high-value applications.

Why is User Provisioning Challenging?

User-provisioning systems have been on the market for 20+ years. In that time, they have garnered a reputation for being difficult and expensive to deploy, and many organizations have found it challenging to realize a return on investment. Why?

User provisioning is similar to data integration technologies in many ways. Like data integration technologies, user-provisioning systems aggregate and synchronize data to many different systems and services in the environment. Each new connection adds a new level of complexity. The process of onboarding a single application to a provisioning environment requires an understanding of the application’s user management APIs and authorization construct, deployment of a connector, configuration of access policies, and implementation of policies and procedures for managing users (e.g., rules against creating or managing users directly within the application). Adding one application can be complex; consider the complexity when you have hundreds or even thousands of applications in your environment.

Another aspect that can be difficult is data quality issues in the SOR. Ideally, there is one authoritative SOR, but this is not always the case. Data collisions may happen when information is coming from multiple authoritative sources. Also, the administrators and users of the SOR may not understand the downstream effects of insufficient and poor-quality data. For example, they may not populate certain fields, enter inaccurate data, or delay event triggers. This all has implications for the provisioning system's ability to accurately update identity records and access entitlements. Managing the data quality process can be taxing on identity practitioners.

Another common challenge is policy definition. Identity practitioners are responsible for configuring access policies (rules/roles), but they don't own access decisions. The line of business in partnership with audit, legal, governance, risk management and compliance (GRC), etc., own access decisions. The effort to collect this data for provisioning policy and role definition is a significant undertaking.

Last but not least, maintaining the entitlement catalog can be a difficult task. A single organization may have hundreds of thousands, if not millions, of entitlements. The effort to collect entitlements and metadata should not be underestimated.

While the advantages of automated user account provisioning are well understood, deployments can be challenging. Identity practitioners should obtain executive support and set proper expectations, build a structured process for onboarding applications (e.g., dedicated resources, intake forms/surveys, automation, etc.), and set clear key performance indicators that show continued progress.

The Next Generation, Hybrid-Approach to Provisioning

While user-provisioning technologies have been around for quite some time, user-provisioning technologies were developed at a time when the IT environment was much more contained. Target systems and systems of record were located on-premises, and users were primarily employees accessing resources onsite.

Cloud, mobile, work from home, and various other initiatives have changed the dynamics of user provisioning. Now authoritative and target systems are hosted in the cloud (and on-premises), and external users are accessing internal resources.

In traditional security models, provisioning is an admin-time function (users are pre-provisioned into systems by an administrator). Contrast this with authentication and authorization technologies that are run-time functions that occur at the point the user is logging into the system. Security technologies like SIEM, DLP, and threat detection kick in once the user is in session.

Modern security models that include remote access, cloud computing, and zero-trust security principles require a new approach to user provisioning: a just-in-time (JIT), least privilege approach. This modern approach to provisioning means that rather than pre-provisioning users at admin-time, users are provisioned at run-time and they are given the minimal privileges necessary to complete a task. Consequently, organizations are deploying a hybrid approach to provisioning that includes a mix of traditional API-based, SAML/OIDC-based, and SCIM-based connectors.

A hybrid provisioning architecture allows organizations to pre-provision to a set of applications (typically on-premises) but use OIDC, SAML, and SCIM-based connectors to enable JIT provisioning (primarily used in cloud and SaaS product use cases).

Author Bios

Ian Glazer

Ian Glazer is the founder and president of Weave Identity – an advisory services firm. Prior to founding Weave, Ian was the Senior Vice President for Identity Product Management at Salesforce. His responsibilities include leading the product management team, product strategy, and identity standards work. Earlier in his career, Ian was a research vice president and agenda manager on the Identity and Privacy Strategies team at Gartner, where he oversaw the entire team's research. He is a Board Emeritus and the co-founder of IDPro and works to deliver more services and value to the IDPro membership, raise funds for the organization, and help identity management professionals learn from one another. During his career in the identity industry, he has co-authored a patent on federated user provisioning, co-authored and contributed to user provisioning specifications, and is a noted blogger, speaker, and photographer of his socks.

Lori Robinson

Lori Robinson is the Vice President of Enterprise Identity Product Management at Salesforce where leads a team responsible for Salesforce's enterprise identity management program. Before joining Salesforce she was the VP Product and Market Strategy at SailPoint. She also served as the Managing Vice President and Analyst at Gartner where she covered the identity governance and administration, privileged access management, and consumer IAM markets. Lori is a recognized industry thought leader, speaker, and publisher. She is passionate about advancing opportunities for women in IT and has led various user groups, round tables, and events for women in identity.

Mat Hamlin

Mat Hamlin is the Vice President of Product Management for Platform Identity at Salesforce. His responsibilities include leading the product management team, product strategy, and innovation for the Identity Services on the Salesforce core platform. Prior to Salesforce, he served in multiple product management roles at SailPoint, Oracle, and Sun Microsystems, focusing on User Provisioning, Access Governance, and Enterprise Role Management.

Change Log

Date	Change
2022-06-03	V1 published

[i] More on Joiner, Mover, and Leaver is available in Cameron, A. & Grewe, O. (2022) "An Overview of the Digital Identity Lifecycle (v2)", *IDPro Body of Knowledge* 1(7). doi: https://doi.org/10.55621/idpro.31

[ii] Changes to name or place of residence can be just as impactful as a change in job role or reporting structure.

[iii] More on roles is available in McKee, M. K., (2021) "Policy-Based Access Controls", *IDPro Body of Knowledge* 1(4). doi: https://doi.org/10.55621/idpro.61 and Koot, A., (2020) "Introduction to Access Control (v3)", *IDPro Body of Knowledge* 1(6). doi: https://doi.org/10.55621/idpro.42.

[iv] Corporate Finance Institute, "Top Accounting Scandals: A recap of the top scandals in the past," n.d., https://corporatefinanceinstitute.com/resources/knowledge/other/top-accounting-scandals/ (accessed 17 May 2022).

[v] For the sake of transparency, one of the authors of this article contributed to the SPML v2 standard and apologizes for the mistakes he didn't know he was making at the time.

[vi] For the sake of transparency, one of the authors of this article contributed to the SCIM v2 standard and is proud of that work.

Standards, Regulations, and Laws

Laws Governing Identity Systems (v2)

Thomas J. Smedinghoff

Table of Contents

Abstract

Identity systems and their participants are governed by a myriad and complex set of laws, regulations, and contractual requirements. This article offers a high-level overview of the legal environment that governs identity systems, focusing on three different levels of legal rules: General Law, Generic Identity System Law, and Individual Identity System Rules.

Introduction

What are the legal rules that govern identity systems? What obligations do those rules impose on the participants involved?

The reality is that identity systems and their participants are governed by a myriad and complex set of laws, regulations, and contractual requirements, and the obligations they impose are not always clear. To make sense of it all, it is best to focus first on the legal environment that governs identity systems.

Terminology

- Consumer Protection Law - laws and regulations that are designed to protect the rights of individual consumers and to stop unfair, deceptive, and fraudulent business practices.

- Contract Law – laws that relate to making and enforcing agreements between or among separate parties.

- Fraud Law – laws that protect against the intentional misrepresentation of information made by one person to another, with knowledge of its falsity and for the purpose of inducing the other person to act, and upon which the other person relies with resulting injury or damage.

- Identity Theft Law – laws governing crimes in which the perpetrator gains access to sensitive personal information belonging to the victim (such as birth dates, passwords, email addresses, driver's license numbers, social security numbers, financial records, etc.), and then uses this information to impersonate the victim for personal gain, such as to commit fraud, establish credit in the victim's name, or access the victim's accounts.

- Privacy Law - laws that regulate the collection, use, storage, and transfer of personal data relating to identified or identifiable individuals.

- Tort Law - the body of law that covers situations where one person's behavior causes injury, suffering, unfair loss, or harm to another person, giving the injured person (or the person suffering damages) a right to bring a civil lawsuit for compensation from the person who caused the injury. Examples include battery, fraud, defamation, negligence, and strict liability.

The Identity System Legal Environment

At a high level, the legal environment that governs the operation of any identity system consists of three different levels of legal rules, categorized as follows:

- Level 1: General Law: The first level is law that applies generally to all business and personal activities. This law covers a wide variety of subjects and is not written with identity systems in mind, although it is frequently applied to identity system activities where appropriate. Examples of general law that might affect the operation of an identity system include contract law, tort law, privacy law, warranty law, and consumer protection law.

- Level 2: Generic Identity System Law: The second level of legal rules consists of law written specifically to govern identity systems generally. Level 2 identity management laws typically apply to all identity systems within a jurisdiction and are often relatively high level in nature. At present, however, very few such Level 2 laws exist. Examples of such generic identity system law include Virginia's Electronic Identity Management Act[i] and the Draft Provisions on the Cross-border Recognition of IdM and Trust Services[ii] being developed by the UN Commission on International Trade Law (UNCITRAL). In many jurisdictions, Level 2 law for identity systems does not yet exist.

- Level 3: Individual Identity System Rules: The third level of legal rules consists of the set of system-specific rules written to govern the operation of a particular identity system. These rules provide the technical, business, and operational specifications and rules for the identity system, specify the rights and responsibilities of the participants and govern the relationships between the various parties. They can be quite detailed but apply only within the confines of the identity system they were written to govern.

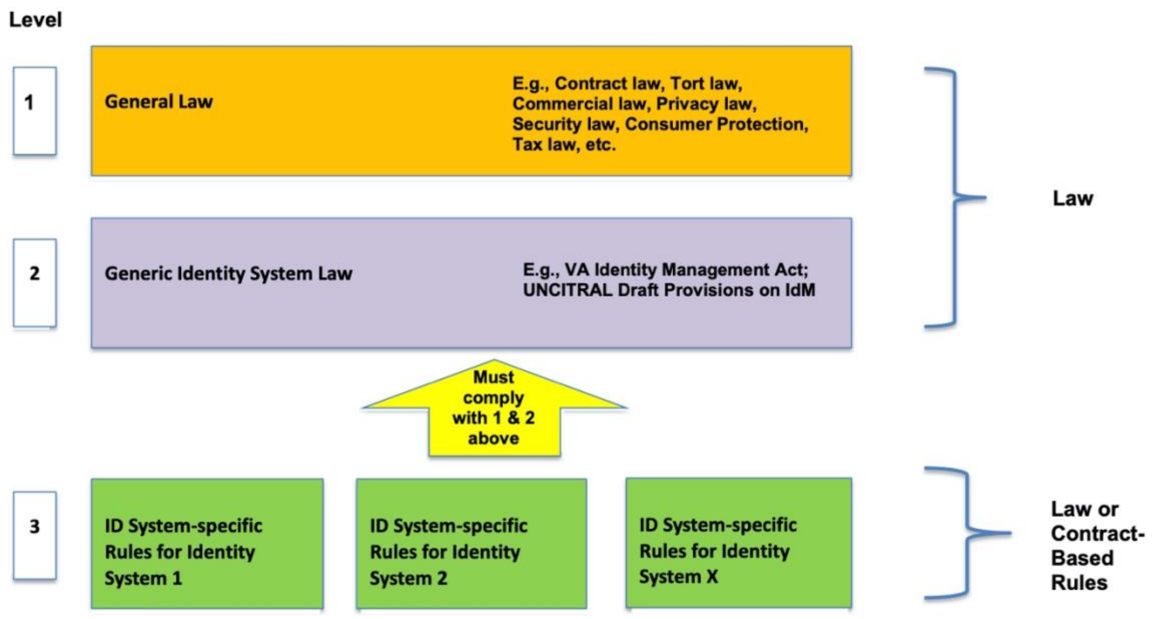
  For private sector identity systems, these legal rules are typically contract-based, are often referred to as a trust framework or system rules, and apply only to those system participants who have contractually agreed to be bound to them. Examples include the SAFE Identity Trust Framework (previously the SAFE-BioPharma Trust Framework),[iii] the Sovrin Governance Framework,[iv] and the SecureKey Concierge Trust Framework.[v]

  For government identity systems, these Level 3 legal rules are often embodied in a law or regulation enacted by the government and thus automatically apply to all those who participate in the identity system. Examples include the eIDAS Regulation in the European Union,[vi] the Identity Documents Act in Estonia,[vii] and the Aadhaar Act in India.[viii] In some cases, however, government identity systems also use contract-based trust frameworks, such as the Trusted Digital Identity Framework (TDIF)[ix] for the Australian national federated identity system.

The Level 3 portion of the legal environment for any identity system is under the control of the developers of that identity system (government or private sector). That is, the operators of a private sector identity system are free to make up the Level 3 system rules and design them in the manner best suited to meet the goals of that specific identity system. However, where such rules are contract-based, they will apply only to the participants that agree to be bound by them, and they may be supplemented (and in some cases overruled) by existing laws and regulations at Levels 1 or 2. In other words, the Level 3 rules designed for any specific identity system must comply with existing law – a challenge made all the more difficult for identity systems that cross jurisdictional boundaries.

The structure of this identity system legal environment
is summarized on the diagram below.

Three Levels of Rules Govern Identity Systems


Level
1
General Law
E.g., Contract law, Tort law, Commercial law, Privacy law, Security law, Consumer Protection, Tax law, etc.
2
Generic Identity System Law
E.g., VA Identity Management Act; UNCITRAL Draft Provisions on IdM
Law
Must comply with 1 & 2 above
3
ID System-specific Rules for Identity System 1
ID System-specific Rules for Identity System 2
ID System-specific Rules for Identity System X
Law or Contract-Based Rules

This structure of the identity system legal environment is very similar to that which governs a credit card system (such as Amex®, Discover®, MasterCard®, or Visa®). Each credit card system is governed by Level 3 system rules developed by the operator of that system (e.g., the MasterCard Rules[x] and the Visa Core Rules and Visa Product and Service Rules[xi]). Those rules provide the technical, business, and operational specifications for the specific credit card system and govern the relationships between the various parties. They are made binding on the parties that participate in the system (e.g., credit card holders, merchants, issuing banks, processors, etc.) by contract.

Those Level 3 credit card system rules and the associated contracts are also governed by: (1) Level 1 general law (e.g., the law of contracts, the law of negligence, etc.), and (2) Level 2 generic credit card system law written to regulate all credit card systems (e.g., Regulation Z[xii] in the US). Like the legal environment governing identity systems, this combination of Level 3 system rules and contracts and Level 1 and 2 law forms the legal environment in which each credit card system operates.

The Legal Rules Governing Identity Systems

Level 1 – General Law

Currently, most law applicable to identity systems is general law (Level 1). Typically, this law was written for a purpose completely unrelated to identity management (e.g., tort law, contract law, warranty law, privacy law, etc.) and without considering how it might apply to identity systems. In fact, in many cases it was written before the concept of identity systems even existed. And in some cases, the law developed over hundreds of years via common law and court decisions. Nonetheless, such general law often applies to identity system-related activities, often in ways that were unanticipated at the time of its original adoption.

Identity systems primarily deal in information. Thus, the Level 1 law that applies to identity systems will typically include those laws that address various aspects of transactions involving information. This primarily includes law governing the following aspects of information:

-- Collection, Use, and Transfer of Identity Information

Identity information about individuals is personal data, and identity system processes typically involve the collection and processing (by an identity provider, attribute provider, or its agents) and disclosure (to a relying party) of such personal data about a subject. Thus, ***privacy*** laws will regulate the collection, storage, use, and transfer of identity information and will have a major impact on all identity system participants and all identity system transactions. This may include, for example, imposing limits on what data may be collected, requirements regarding notices of collection practices, limits on the use that may be made of such data, and restrictions on the transfer of such data to third parties and/or across country boundaries.

-- Accuracy of Identity Information

A key concern of all participants in an identity system relates to the accuracy and reliability of the identity information they are communicating or relying upon. Inaccurate identity data can cause a variety of problems for persons who rely on that data, as well as liability for those who provide it.

Laws governing providing false or incorrect information, whether intentionally or negligently, will be relevant in the evaluation of the rights, obligations, and liabilities of the participants in identity systems, including identity providers, attribute providers, and data subjects.

Key among them are ***fraud*** laws and ***identity theft*** laws. Fraud involves a representation of fact (or material omission of fact) that is intended to deceive another to their material detriment. Identity theft occurs when a party acquires, transfers, possesses, or uses someone's personal information in an unauthorized manner, with the intent to commit, or in connection with, fraud or other crimes.

Even in the absence of fraud, the tort of ***negligent misrepresentation*** can create liability for communicating false information. This occurs where the information is intended for the guidance of others in their business transactions, but the information provider did not exercise reasonable care in determining the accuracy of the information prior to the communication. Thus, in certain circumstances, an incorrect assertion of one or more identity attributes might qualify as a negligent misrepresentation.

This tort of negligent misrepresentation creates a duty to exercise reasonable care or competence to verify facts and creates liability for incorrect representations made without exercising reasonable care about the accuracy of the facts asserted. However, it does not make the supplier of information (e.g., the identity provider) a guarantor of the accuracy of an identity assertion. Generally, the information provider does not have liability for inaccurate or "false" information unless the provider failed to exercise reasonable care in obtaining or communicating the information.

To the extent that incorrectly communicated identity information damages the reputation of the data subject, the tort of ***defamation*** may also be relevant. Defamation involves a false or disparaging statement of fact about a person that is published to a third party causing the person to suffer harm. It is possible that incorrect identity or attribute assertions could be considered defamatory in certain situations. For example, asserting an inaccurate attribute – e.g., age, medical information, sexual orientation, political affiliation, or employment -- might be considered defamatory in certain cases where the named person suffered harm as a result.

The accuracy or reliability of identity attribute information communicated to a relying party by an identity provider or attribute provider may also be governed by ***warranty*** law. A warranty is an assurance, promise, or guaranty by one party to another party that facts or conditions are true and may be relied upon by the other party.

A warranty may be either express or implied. An *express warranty* arises from specific statements made by one party to another. Such statements may be made in writing, such as in a contract or advertisement, or may be made orally, such as by a sales representative. For example, an identity provider’s published processes may include a warranty regarding the quality of the information it provides to relying parties.

An *implied warranty* is an unspoken, unwritten promise created by law that arises from the nature of the transaction and the inherent understanding by the recipient rather than from the express representations of the provider. Implied warranties are based upon the common law principle of “fair value for money spent.” Thus, for example, a court could conceivably conclude that identity providers make implied warranties regarding the reasonableness of the processes they used to collect and verify identity attribute data.

Finally, it is important to note that some privacy laws also regulate the accuracy of personal data. The EU GDPR, for example, requires that personal data maintained by data controllers (such as identity providers) must be “accurate and, where necessary, kept up to date” and that “every reasonable step must be taken to ensure that personal data that are inaccurate … are erased or rectified without delay.” Article 5(1)(d). In addition, it provides that “The data subject shall have the right to obtain from the controller without undue delay the rectification of inaccurate personal data concerning him or her.” Article 16.

-- Availability, Retention, and Deletion of Identity Information

In the case of identity systems where an identity provider, relying party, or other identity system participant retains data about a data subject, the availability, retention, and deletion of such identity information can be regulated by a variety of Level 1 laws.

Privacy law (e.g., GDPR and the California Consumer Privacy Act (CCPA)[xiii]) often regulates the availability of personal data (and hence identity data) to the data subject. In particular, such laws often impose on identity providers a duty to provide individual data subjects with access to the data it has collected about them, as well as information regarding the purposes for which it collects and processes such data, and the recipients or categories of recipients to whom the data are disclosed

Numerous laws also impose ***data retention*** obligations on companies regarding their corporate records. These laws may apply to and require both identity providers and relying parties to retain certain identity data for a particular period of time.

Finally, however, ***privacy*** laws (such as the GDPR) may impose limits on the retention of personal data. And increasingly, privacy laws (such as GDPR and CCPA) grant data subjects are right to request that data about them be deleted or erased.

-- Security of Identity Information and Processes

Many ***data security laws*** and regulations impose obligations on companies with respect to the security of personal data and other information in their possession or under their control. To the extent that a participant in an identity system is collecting, using, storing, or transferring personal data, such data security laws may have a significant impact on its obligations and potential liability. This is particularly true for identity providers and relying parties.

Data security laws are sometimes incorporated into privacy laws, but regardless of form, they generally impose two key obligations: (1) a duty to *provide reasonable security* for personal data, and (2) a duty to *disclose breaches* of security of personal data to the persons affected and to regulators. Although not written specifically to address identity system activities, such laws will undoubtedly apply to the personal data used by identity systems as well.

Level 2 – Generic Identity System Law

The application of existing general law to identity systems is often not a good fit, frequently ambiguous, and in many cases leads to arguably inappropriate results. This is further complicated by the fact that the Level 1 laws applied to identity systems can vary considerably across jurisdictions. Thus, there have been several attempts to address these concerns.

Some jurisdictions have proposed, and some have enacted, legislation or regulations expressly governing all identity systems within their jurisdiction. However, there is not yet agreement on the desirability or goals of such generic legislation, much less on how to achieve them. Key questions yet to be resolved include whether such legislation should be designed to: (1) simply remove legal barriers (actual and perceived) to identity systems, (2) encourage and assist the development of identity systems, or otherwise help establish the “trust” and the “predictability” needed by parties engaged in online identity transactions, or (3) regulate and control identity systems, such as by protecting the privacy of personal information, ensuring the security and trustworthiness of identity transactions, or imposing or limiting the liability of identity providers.

At present, very little Level 2 law exists. Nevertheless, some noteworthy efforts to develop Level 2 law governing identity systems include the following:

Virginia. The state of Virginia became the first US state to adopt Level 2 identity legislation by enacting the Virginia Electronic Identity Management Act in 2015. That legislation is focused primarily on the issue of liability. To do that, it provides for the creation of a Virginia Identity Management Standards Advisory Council, which was tasked with developing Identity Management Standards. Identity providers and trust framework operators that comply with the requirements of those Identity Management Standards are then granted immunity from civil liability. In other words, the Virginia Act provides a safe harbor from liability for identity providers and trust framework operators.

UN Commission on International Trade Law (UNCITRAL). In the Spring of 2015, both the American Bar Association Identity Management Legal Task Force, and a group of EU countries (Austria, Belgium, France, Italy, and Poland, with support from the EU Commission), submitted proposals to UN Commission on International Trade Law (UNCITRAL) regarding identity management legislation. Those proposals recommended that UNCITRAL undertake a project to develop "a basic legal framework covering identity management transactions, including appropriate provisions designed to facilitate international cross-border interoperability." UNCITRAL has since agreed to move forward with such a project.[xiv]

UNCITRAL provides an international forum capable of developing a harmonized set of globally accepted law governing identity management. Such law can be adapted domestically by individual countries to promote a universal approach to identity management law and can be extended globally (to facilitate cross-border identity transactions) through an international treaty or convention.

In September 2019, UNCITRAL produced the second version of its Draft Provisions on the Cross-border Recognition of IdM and Trust Services. Issues currently being considered include the:

- Rights and responsibilities of various identity system roles
- Determination of the reliability of identity systems
- Liability of identity providers
- Legal recognition of identity credentials.
- Cross-border recognition of identity credentials.

Level 3 – Individual Identity System Rules

Both Level 1 and Level 2 law provides general rules applicable to all identity systems. But because each identity system is unique, it also requires its own tailored set of more detailed rules to govern its operations.

In fact, having predictable and enforceable rules designed to ensure that it functions properly and is trustworthy is key to any identity system. Unique system rules (e.g., a trust framework) will ideally provide such a structure to govern the operation of an identity system, much like the Visa or MasterCard rules (including the payment card industry data security standard or PCI-DCSS) that govern credit card systems.[xv] Such rules include the technical specifications and operational rules and requirements necessary to make the system functional and trustworthy and the legal rules that define the rights and legal obligations of the parties and facilitate enforcement where necessary.

These individual identity system rules are the Level 3 law that governs an identity system. For private sector identity systems, these rules typically take the form of a so-called trust framework and are made enforceable against the various system participants by contract. Accordingly, those rules must comply with any restrictions at Levels 1 and 2 law.

In the case of public sector identity systems (such as a national ID system), these rules usually take the form of legislation or regulations adopted by the government to govern the system. Many countries, including most notably Estonia and India, have adopted laws to govern their specific national ID systems. In some cases, a country may establish an identity system based on a set of rules that participants voluntarily agreed to by contract. The Australian Trusted Digital Identity Framework (TDIF), and the UK GOV. UK Verify program takes this approach.

Regardless of whether an identity system is public or private, the issues addressed by the Level 3 system rules/trust framework often include the following:

- technical specifications that will govern the system
- rights and obligations of participants in each system role
- data subject registration and enrollment processes
- identity verification process requirements
- credential issuance requirements
- authentication process requirements
- rules governing reliance by relying parties
- data security requirements (over and above requirements of applicable law)
- privacy requirements (over and above requirements of applicable law)
- audits, assessments, and certification requirements
- allocation of liability risk among roles
- termination rights and obligations
- dispute resolution
- enforcement of rights and obligations

Where such rules are embodied in laws or regulations issued by a government, they are of course binding on all system participants by force of law. But in the case of a trust framework (typically used in a private-sector system), the system rules are binding on the participants only to the extent they agree by contract to be bound to comply with the rules.

In all cases, however, the Level 3 law is comprised of system rules written for a specific identity system, and thus its applicability is limited to that system.

Author Bio

Thomas J. Smedinghoff is Of Counsel at Locke Lord, LLP, and Chair of the American Bar Association Identity Management Legal Task Force. He can be reached at Tom.Smedinghoff@lockelord.com

Change Log

Date	Change
2021-06-30	Editorial updates

[i] Code of Virginia - Chapter 50. Electronic Identity Management Act. 2015. https://law.lis.virginia.gov/vacode/title59.1/chapter50/.

[ii] "Draft Provisions on the Cross-border Recognition of IdM and Trust Services," revision A/CN.9/WG.IV/WP.160, United Nations Commission on International Trade Law, last revised 16 September 2019, https://uncitral.un.org/sites/uncitral.un.org/files/media-documents/uncitral/en/wp-160-e.pdf.

[iii] "Greater Security via the SAFE Identity Trust Framework," accessed 18 May 2021, SAFE Identity, https://makeidentitysafe.com/trust-framework/.

[iv] "Sovrin Governance Framework," accessed 18 May 2021, Sovrin, https://sovrin.org/library/sovrin-governance-framework/.

[v] "SecureKey Concierge Trust Framework," accessed October 10, 2019, SecureKey, https://securekey.com/resources/trust-framework-securekey-concierge-in-canada/.

[vi] eIDAS Regulation[EU]: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2014.257.01.0073.01.ENG

[vii] Identity Documents Act [Estonia]: http://www.unhcr.org/refworld/docid/4728ab1b2.html

[viii] Aadhaar Act https://uidai.gov.in/images/targeted_delivery_of_financial_and_other_subsidies_benefits_and_services_13072016.pdf.

[ix] "Trusted Digital Identity Framework," accessed 18 May 2021, Australian Government Digital Transformation Agency, https://www.dta.gov.au/our-projects/digital-identity/trusted-digital-identity-framework.

[x] "Mastercard Rules," accessed 18 May 2021, Mastercard, https://www.mastercard.us/en-us/about-mastercard/what-we-do/rules.html.

[xi] "Visa Core Rules and Visa Product and Service Rules," 15 October 2013, accessed 18 May 2021, Visa, https://usa.visa.com/dam/VCOM/download/merchants/visa-international-operating-regulations-main.pdf.

[xii] "§ 1026.1 Authority, purpose, coverage, organization, enforcement, and liability," 12 CFR Prt 1026 (Regulation Z), Consumer Financial Protection Bureau, accessed 18 May 2021 https://www.consumerfinance.gov/rules-policy/regulations/1026/.

[xiii] "California Consumer Privacy Act of 2018," Title 1.81.5, Section 1798.100, Part 4 of Division 3, State of California, accessed 18 May 2021, https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5.

[xiv] UNCITRAL, "Colloquium on Identity Management and Trust Services," 21-22 April 2016, accessed 18 May 2021, https://uncitral.un.org/en/colloquia/electronic_commerce/2016.

[xv] PCI Security Standards Council, Standards Overview, website, https://www.pcisecuritystandards.org/standards/.

An Introduction to the GDPR (v3)

By Andrew Cormack, Chief Regulatory Adviser at Jisc

To comment on this article, please visit our GitHub repository and submit an issue.

Table of Contents

Abstract

The General Data Protection Regulation (GDPR) applies to any processing (including collection, storage, or sharing) of data relating to identifiable (including by serial numbers, IP addresses, etc.) individuals who are physically in Europe. This scope may well cover international or online Identity and Access Management (IAM) activities, as well as all IAM activities actually conducted in Europe. All such processing must conform to seven principles: lawfulness, fairness & transparency; purpose limitation; data minimisation; accuracy; storage limitation; integrity & confidentiality; accountability. Individuals have rights of information; subject access; rectification, erasure & restriction. Processing must be for one of six legal bases: contract, legal obligation, vital interests, public interests,

legitimate interests, or consent. Each basis has its own requirements; some confer additional rights on individuals.

Introduction

The *General Data Protection Regulation (GDPR)*,[i] which came into force in all EU member states on May 25, 2018, applies when processing 'any information relating to an identified or identifiable natural person'.[ii] The inclusion of 'identifiable' makes it much broader than most privacy laws: IP addresses, MAC addresses of personal devices, account numbers, and even unique patterns or combinations of attributes may be sufficient to bring an activity within its scope. 'Processing' is not limited to digital formats: personal information prepared for, or derived from, digital processing is covered, as well as the content of any structured filing system. The range of activities covered is similarly wide: including 'collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or deletion'.[iii] Since the GDPR covers all individuals physically in Europe – there is no citizenship or similar requirement – it is very likely to apply to the international or online activities of organisations elsewhere in the world, as well as to all organisations in Europe.

IAM activities are likely to be regulated by the GDPR; however, effective IAM may make it easier for organisations to comply with the law's requirements. The behaviour it prescribes is increasingly expected, not only in Europe, but in the increasing number of countries subscribing to the Council of Europe's Convention 108.[iv] Within Europe there are significant fines for contravention of the GDPR, but following its principles should have benefits for the reputation and efficient operation of organisations anywhere in the world.

This article is not a complete guide to the GDPR but covers those aspects most relevant to IAM. It first describes the general principles and obligations that apply to all personal data processing; then examines the permitted legal bases for processing and the specific obligations and rights associated with them. Finally, examples show how IAM activities can help organisations conform to the GDPR's requirements.

Terminology

- **General Data Protection Act (GDPR).** Formally, Regulation 2016/679 of the European Union, in force May 25, 2018. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679
- **Personal Data.** Defined in Article 4(1) of the GDPR: "'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;". Note: "natural person" (human) is used to distinguish from companies and other corporate entities that are "legal persons".
- **Processing.** Defined in Article 4(2) of the GDPR: "'processing' means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation,

structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction”. Note that even this long list of activities is not exhaustive: other activities may also fall within the definition of “processing”. Additional rules, in Article 22, apply to “automated individual decision-making, including profiling”. These generally have the effect of strengthening the rights of information and objection described later and may limit the use of automation for some high-impact decisions.

- **Special Category Data (SCD).** Categories of data that are regarded as particularly sensitive, so subject to additional regulation. Defined in Article 9(1) of the GDPR as “personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person’s sex life or sexual orientation”; Article 10’s “personal data relating to criminal convictions and offences” requires similar treatment, so is normally considered as another category of SCD.
- **Data Controller.** Defined in Article 4(7) of the GDPR: “‘controller’ means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data;”.[v] This article uses the term “organisation” as a synonym for “data controller”, since organisations involved in IAM will normally be data controllers.
- **Data Processor.** Defined in Article 4(8) of the GDPR for situations where an organisation processes personal data solely on the instructions of others. A Data Processor must not determine the purposes of processing, for example by processing in its own interests, or, beyond limited technical choices, the means of doing so. Data Processors are regulated by Article 28: in particular they must have a contract with the Data Controller that covers all the subjects listed in Article 28(3). Data Processors are excluded from some, but not all, of the liabilities and duties of Data Controllers.
- **Data Subject.** Defined in Article 4(1) of the GDPR (see “Personal Data” above) as the formal term for the human to whom personal data relates. This article uses the term “individual” as a synonym for “data subject”.

Rules for Personal Data

The GDPR places most of its obligations on organisations that “determine[…] the purposes and means of the processing of personal data” (Art 4(7)): these organisations are referred to as *Data Controllers*. Some organisations may process data solely on behalf of others – not determining the purposes and means – these are known as *Data Processors* and have fewer obligations. Since IAM systems are likely to act as data controllers, their main obligations are described here. The fundamental obligations on all data controllers are to act in accordance with seven principles, and to satisfy obligations to, and rights of, individuals (“*data subjects*”) whose information they process.

Principles (Art 5)

According to GDPR Article 5, the following principles apply to all processing of personal data:

- **Lawfulness, Fairness, Transparency**: all processing must be covered by one of the six legal bases set out in the GDPR (see below) and must not breach other laws; it should not be deceptive, any activities that individuals might be surprised by should be explained and justified as must any adverse effects on individuals; organisations should be open about their processing, in particular through the rights to information and subject access described below.
- **Purpose Limitation**: the purposes for which information is processed must be clearly stated; existing information may only be used for new purposes if, either, the new purpose is compatible with the existing ones (roughly summarised as 'not surprisingly different'), or it is required by law, or each individual has given consent to the new purpose. IAM systems should be designed to serve a single purpose and any proposals to re-use their data for other purposes should be reviewed for compatibility with that purpose and with the information provided to users.
- **Data Minimisation**: the data and processing must be relevant to the purpose, sufficient to achieve it ("adequate"), but not excessive. Well-defined IAM systems should contribute to data minimisation: for example, federated systems can reduce disclosure by using opaque identifiers ("pseudonyms") that allow an individual to be recognised when they return to a system, without identifying them. IAM systems should be designed to collect, use and disclose the minimum personal data required for each function. If a function can be delivered with anonymous or pseudonymous data, then it should be. This is the basis for Data Protection by Design, discussed in GDPR Article 25.
- **Accuracy**: personal data must be accurate and up to date. Although individuals have the right to correct errors in their data (see "right of rectification" below) organisations should not rely on them doing so as the sole, or even principal, way to ensure accuracy. IAM systems that act as a single source of truth for their organisations should make accuracy significantly easier to achieve; those that do not should be accompanied by appropriate policies, processes and workflows to ensure that their information is, and remains, accurate.
- **Storage [time] Limitation**: personal data must not be kept for longer than needed for the stated purpose(s). Before collecting personal data, organisations should know, and declare, how long they will keep it for, either in relation to a fixed time period (e.g., 'six months'), or a known event (e.g., 'until you leave'). Organisations should have processes to ensure their stated retention periods are implemented; at the end of them data should be deleted or anonymised. The purposes of archiving, research, and statistics may allow personal data to be kept for longer, but subject to specific conditions in both European and national laws.
- **Integrity and Confidentiality**: organisations must use appropriate technical and organisational controls to protect the security of personal data. What is appropriate will depend on the sensitivity of the data and the purpose: it is likely to change both as new protective technologies and approaches become available and as new threats and risks become apparent. The GDPR imposes specific obligations if there is a breach of security, which are described below. IAM systems should help both by holding their own personal data securely, and as a component of the access control systems used to prevent unauthorised access to personal data elsewhere in the organisation.

- **Accountability**: organisations must be able to demonstrate that they are complying with the principles and other requirements of the Regulation. This will normally require both documentation showing that these principles and requirements were considered in the design of the system, and audit logs (which themselves may contain personal data) confirming that normal operations and responses to events such as breaches and any exercise of individual rights were, in fact, conducted in accordance with them.

Obligations and Rights

Three groups of "rights" apply to all processing of personal data except where limited exceptions, set out in the specific Articles, apply. The first group creates an obligation on organisations towards all those whose information they process; the second and third require organisations to have systems to handle requests from individuals who exercise their rights:

- **Rights to Information**: to support the above Principles, organisations are required to provide at least a minimum set of information to all those whose personal data are processed: who the organisation is, what data are being processed, why, for how long, whether automated decisions are involved; any other organisations or further processing involved; how to exercise your rights. Article 13 applies when data are collected directly from the individual; Article 14 when an organisation obtains personal data from another source (including public sources).
- **Subject Access Right**: individuals have a general right, under Article 15, to ask and be told whether their data are being processed, what data, why, for how long, whether automated decisions are involved; the source of the data and any recipients; how to exercise their rights. In addition, if this can be done without affecting the rights of others, the individual has a right to receive a copy of their own data. Determining what to release, and when, can be complex, especially when the requester's identity may be uncertain. IAM systems built around guidance from regulators[vi] can reduce the risk of error or fraud.
- **Rights of Rectification/Erasure/Restriction**: Article 16 ("rectification") entitles individuals to correct inaccurate personal data, including to add additional information. Article 17 ("erasure") entitles individuals to have their personal data deleted if there is no lawful basis for it to be kept. This might arise, for example, when excessive information is held, if it has been kept beyond its retention time, or, if it was being processed on the basis of consent (see below) when that consent has been withdrawn. Article 18 ("restriction") entitles an individual to block further processing of their data (including deletion) while a rectification or objection right is being processed, or as an alternative to erasure if the individual needs the data for a legal claim. IAM systems that provide a single point of truth and control should make it easier to implement these rights.

Legal Bases for Processing

To be lawful, any activity that involves processing personal data must be covered by one of the six legal bases set out in Article 6 of the GDPR. Note that the basis applies to a particular processing activity, not to a dataset. As illustrated in the example below, an IAM system may involve several different legal bases. While IAM professionals should probably not be determining the Legal Bases on behalf of their organisations, they need to be aware of the implications of that choice.

Various types of personal data – including race, ethnicity, and health – are considered higher risk and processing must be for one of the purposes set out in Article 9, as well as having an Article 6 basis. The requirements on processing these types – known as *Special Category Data* – are often set in national, rather than European, legislation. IAM systems that process them should therefore consult lawyers familiar with the relevant national schemes. Similarly, although the GDPR highlights the extra risks involved in children's personal data, the specific additional requirements – including the age below which someone is considered a child – are largely set at national level, so are not covered here.

Each of the Article 6 bases imposes additional conditions on processing, both by its definition and, in some cases, by explicit additions. Several of the bases also create additional obligations for organisations processing personal data and/or rights for individuals whose personal data are processed. The following sections describe these legal bases; here they are set out in the likely order of preference for organisations, rather than that in which they are listed in the legislation; those at the bottom of the list are significantly more onerous.

Necessary for the Performance of a Contract

Five of the legal bases begin "necessary for…". Regulators have confirmed that this means there must be no less intrusive way to achieve the purpose.

The inclusion of "performance of" indicates that there must be a particularly close link between the processing and the subject of the contract; the individual whose data are processed must also be a party to the contract. However, the term "contract" is likely to be widely interpreted, covering many situations where parties have made an agreement, even without a formal contract document. If stopping processing would make that agreement impossible to fulfil, then "necessary for contract" may well be an appropriate basis. This is likely to apply to many IAM systems, for example those provided for internal use by an employer or educator. Even for stand-alone IAM systems – so long as there is a direct relationship between the individual and the IAM provider – using "necessary for contract" may be a useful way to distinguish the minimum data and processing without which the service cannot function from optional data that the system can use but does not need. The latter should use the basis of "consent" described below. The European Data Protection Board's Guidelines clarify that ancillary functions including service improvement, fraud prevention and online behavioural advertising are likely to need a different legal basis[vii].

Where personal data are processed on this basis, the GDPR introduced a Right to Portability (Article 20) covering data "which [the individual] has provided". This right may therefore cover only a subset of the information available under the general Subject Access Right, though the information must be provided "in a structured, commonly used and machine readable format". So far, Regulators have only provided high-level guidance on this right,[viii] including suggesting that CSV might fulfil the format requirements, so further developments are likely.

Necessary for Compliance with a Legal Obligation

Where a European or Member State law requires an organisation to process personal data, this is likely to be the appropriate legal basis. It is possible that this might apply to some national IAM schemes, and those in regulated industry sectors, but otherwise it is unlikely to be relevant.

Necessary in Order to Protect Vital Interests

This legal basis may apply when there is a threat to life or serious injury. We should hope that it is not relevant to our IAM systems!

Necessary for the Performance of a Task Carried out in the Public Interest

This legal basis is typically used where a law permits processing for a public interest task but does not require it. Since national, and other statutory, IAM schemes will normally be subject to a legal requirement (see "legal obligation" above), rather than a permission, it seems unlikely to be relevant to IAM systems.

This basis gives individuals the Right to Object to processing, as described under "legitimate interests" below.

Necessary for the Legitimate Interests of the Controller or a Third Party

Whereas the first four bases cover specific situations defined in law the last two ("legitimate interest" and "consent") are more flexible and are therefore subject to more onerous requirements to protect individuals. This Legitimate Interests basis requires not just that the processing be necessary to achieve a specific purpose (the "interest") but also that that interest be "legitimate" and, uniquely, that the benefits of processing not be overridden by its risks to individuals. A processing activity may be necessary for a legitimate interest, but still be unlawful if it cannot satisfy this balancing test.

Legitimate interest will, however, often be the most appropriate legal basis for multi-lateral IAM, for example where identity assertions are provided to external organisations ancillary to a contract for some other purpose. Organisations participating in federations – whether as identity providers, service providers, attribute authorities, or otherwise – are unlikely to know enough about the user's reason for making a particular request to know whether it is necessary for a contract or, conversely, a situation where the individual is able to give free consent. Rather than trying to communicate that information among multiple parties or establishing a mesh of contracts among them, it is often simpler to consider the interest of each individual organisation in providing the service that the individual – by initiating an authentication or authorisation process – has requested of them.

This basis can only be used if "such interests are not overridden by the interests or fundamental rights and freedoms of the [individual] which require protection of personal data" (Article 6(1)(f)). Before an IAM organisation considers releasing (or requesting) information on this basis, it must therefore consider what risks might arise to the individual as a result of that disclosure. The mention of "fundamental rights and freedoms" indicates that risks beyond just data protection should be considered. Although this might appear onerous, the process can often be simplified, and implemented in the form of attribute release policies, by considering the types of data involved and what is known about the entities that will receive the information. Releasing a low-risk attribute to an organisation that has committed (or is required by its own applicable laws) to only use such data for service provision might be considered an acceptable risk, given that the individual must first have chosen to request federated authentication to that organisation's services.

When using the legitimate interests basis, each individual has a "Right to Object" under Art.21. The legal requirement is to consider whether the organisation has "compelling legitimate grounds" for continuing the processing, in which case it may do so. In practice, since IAM systems should, in any case, only be processing the minimum information necessary to provide their service to users, an objection is effectively a request to stop using those parts of the service that rely on Legitimate Interests. An organisation might, therefore, respond to such a request by checking that that is, indeed, the individual's intention.

Consent

The only legal basis that does not contain the word "necessary" is that the individual has given consent to processing. However, this is subject to significant conditions – in Article 7 and Recitals 32, 42 & 43 – which are likely to make consent inappropriate for much of the processing involved in IAM. Consent must be indicated by "a clear affirmative act establishing a freely given, specific, informed and unambiguous indication of the [individual's] agreement"; it must be possible to withdraw consent at any time, as easily as it was given; consent will not be valid "if the [individual] has no genuine or free choice or is unable to refuse or withdraw consent without detriment". Consent might be used where an IAM system can contain additional information, or support other processing, that is not necessary for its core function (for example nicknames), but in this case the individual has an absolute right to have that additional information removed, or the extra processing terminated, at any time.

In addition, consent sought by an employer, public authority, or other organisation with similar power over the individual is presumed not to be free. Consent must not be sought as a condition of providing a service. Organisations relying on consent must be able to demonstrate that it was obtained in accordance with these conditions. As for "contract" above, the Right to Portability applies to information obtained using consent.

Summary

The "necessary" bases – usually either contract, legitimate interest, or legal obligation – are more suitable for the information necessary to maintain the relationship between the individual and the IAM system. With these, the organisation does not have to worry whether lawful consent was obtained, nor that it might be withdrawn on a whim. Consent should be reserved for information that the IAM system can handle but does not need: circumstances that are much more likely to satisfy the requirements for it to be valid. Consent, according to the UK's Data Protection Regulator, should be an offer to the individual to enter into a deeper, more trusting, relationship.[ix]

International Transfers

Any transfer of personal data from a country within the European Economic Area to one outside (commonly referred to as an "export") requires its own legal basis. The full list of possible bases can be found in Articles 45-49. In practice, and unlike the previous Data Protection Directive, it will usually be possible to use the same legal basis for international IAM operations as those within Europe: regular transfers of personal data (for example between a customer organisation and a non-European IAM supplier) should normally be covered by a contract including one of the sets of Standard Contract Clauses;[x] occasional, ad hoc, low-risk transfers should be able to use the legitimate interests basis; consent may be used where the individual is free to choose whether or not their personal information are transferred. Arrangements for international transfers are subject to change: for example both the original US Safe Harbor scheme and the Privacy Shield that replaced it have been declared invalid by the European Court of Justice; the latter case ("Schrems II") also added new obligations for exporting organisations using the Standard Contract Clauses: new versions of the Clauses were issued by the European Commission in June 2021.[xi] Organisations operating international IAM systems should be aware of developments.

Security

As well as requiring organisations to take proactive measures to protect the security of personal data, Article 33 of the GDPR introduces significant reporting requirements when an organisation becomes aware of a "breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed". The wide definition of "breach" and the inclusion of "accidental" means that organisations should be particularly careful when designing, testing, and documenting processes that may alter, delete, or disclose data. All such breaches must be reported to the Regulator unless they are "unlikely to result in a risk to rights and freedoms of natural persons". Loss of an encrypted memory stick, while the decryption key remains secure, is often given as an example of a breach that may not need to be reported. The expectation is that such reports will be sent within 72 hours: if not, then a satisfactory explanation for the delay must be included. Where a breach is likely to involve a "high risk" to individuals' rights and freedoms, then a notification to affected individuals is required under Article 34.

The GDPR recognises in Recital 49 that the ability to detect, contain, and remedy security breaches is an important part of keeping data secure. Indeed, it has been suggested that failure to do so may itself be a breach of Article 33.[xii] Processing of personal data such as access and activity logs required for those purposes is recognised as a legitimate interest (so permitted, subject to the balancing test). Such logs must, of course, be held and processed securely. IAM can play a significant role in mitigating security breaches, by disabling compromised accounts quickly and effectively; its logs may also provide early warning when an organisation is under attack.

To meet the GDPR's tight timescale for understanding and reporting breaches, organisations must plan, prepare, resource, and practice how they will respond to security incidents. This could include assessing which types of breach of the IAM system would require notification to regulators, individuals, or neither, as well as identifying and establishing contact with the internal and external partners whose help would be required.

IAM Examples

The following examples show ways that IAM systems can support the GDPR.

Example 1: Outsourced Office Systems

John works at a small business, which has contracted with a cloud service provider to run its HR and office software services. As agreed in that contract, the service provider subcontracts the operation of email and document sharing to Google. John's employer enters the information necessary for his employment role into a series of webforms; the service provider sets up the necessary accounts and document permissions. John's personal data is processed on the basis that it is necessary for his contract of employment; only the information required to set up his email and document account is passed to Google.

In this example, John is the Data Subject and his employer is the Data Controller. Provided they only use information to provide the contracted services, the service provider and Google are Data Processors. If either were to use data for their own purposes – for example, to display customised adverts – then they would be Data Controller for that processing and be required to fulfil all the Data Controller's obligations.

Example 2: Federated Access Management

Janet is a professor at the University of Erewhon. The university has a central IAM system containing the details of all staff required for them to do their jobs. This information is stored and processed on the legal basis that it is necessary for Janet's contract of employment with the university: without doing so, it would be impossible to perform that contract. The IAM system acts as a single point of truth, so ensuring that information is up to date throughout the university and that any correction requests can be easily implemented.

The IAM system also allows Janet to store optional information, such as her personal interests, that will appear on her staff webpage. Since she can add, change, or remove these at any time, without affecting her work, the appropriate legal basis is consent.

The university is also a member of an Authentication & Authorisation Infrastructure (AAI) Federation. When Janet accesses a website of another Federation member (for example, a journal publisher), she can choose to log in with her university credentials. A wide variety of organisations are Federation members since – with the university taking responsibility for providing verified information and ensuring its users' good behaviour – this allows them to receive and process considerably less personal data, in accordance with the data minimisation principle. Janet needs to access some of these for her work, but others may be just for personal interest. Since neither the university nor the sites wish to work out which sites are necessary for contract and which accessed with free consent (where Janet needs to access a site for work, her consent cannot be free) they both use the legal basis that the processing is necessary in their legitimate interest in helping Janet access the information she wants.

The legitimate interests basis requires the university to balance the risks of releasing information against the benefits. Since the federation agreement requires members only to use authentication and other attributes for the purposes of service provision and personalisation, and not to attempt to identify pseudonymous users, the university assesses that there is very little risk in releasing a unique opaque identifier and Janet's status as a member of staff to any Federation member; it has therefore configured its systems to release that information by default when a user requests a federated login. This is sufficient both for Janet to access online journals, and to verify her entitlement to a staff discount at the local health club.

The Federation has defined a class of services that are specifically designed for Research and Education use, and that require a name and email address in addition to the opaque identifier and status. This additional requirement is mentioned in the services' privacy notices. Although this disclosure involves a slightly higher risk, the university is satisfied that this is justified by the greater benefit; such services will therefore receive the additional information by default. This allows Janet to use discussion groups and virtual research environments in her field.

Where services ask for more information, the university will perform an individual assessment of the benefit and risk. This may indicate that additional measures, such as a bilateral contract or the free consent of each individual, are required to reduce the risk of the disclosure.

In this example, Janet is the Data Subject. Both the university and the service provider are Data Controllers, since the service provider chooses which services to offer to Janet.

Author Bio

Andrew Cormack is Chief Regulatory Adviser at Jisc. He has been involved in the technical and policy development of federated identity systems in the UK, Europe, and globally for more than fifteen years. He has spoken and written extensively on how digital technologies can be used to improve privacy and data protection and, more recently, on the application of the GDPR to them. His publications can be found at https://orcid.org/0000-0002-8448-2881 and his blogs at https://community.jisc.ac.uk/blogs/regulatory-developments.

Change Log

Date	Change
2021-06-30	Updated based on https://github.com/IDPros/bok/issues/42, https://github.com/IDPros/bok/issues/41
2022-09-30	Updated information on The EU Standard Contract Clauses

[i] "EU General Data Protection Regulation (GDPR): Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)," OJ 2016 L 119/1.

ii GDPR Art.4(1)

[iii] GDPR Art.4(2)

[iv] "Council of Europe Data Protection website," Council of Europe, accessed October 10, 2019, https://www.coe.int/en/web/data-protection/home.

[v] Note that some public authorities are excluded from GDPR, notably institutions of the European Union itself and law enforcement and national security agencies when performing those tasks. These will normally be subject to other legislation that applies the same principles: for example, Regulation 2018/1725 for EU bodies and Directive 2016/680 for law enforcement.

[vi] See, for example, the UK Regulator at https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-of-access/

[vii] European Data Protection Board, "Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects," Version 2.0, 8 October 2019, https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines-art_6-1-b-adopted_after_public_consultation_en.pdf.

[viii] "Guidelines on the right to "data portability"," revision (wp242rev.01), European Commission, last modified October 27, 2017, https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611233.

[ix] "When is consent appropriate?" Information Commissioner's Office, accessed October 10, 2019, https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/consent/when-is-consent-appropriate/#when3.

[x] "Standard Contractual Clauses: Standard contractual clauses for data transfer between EU and non-EU countries," European Commission, accessed October 10, 2019, https://ec.europa.eu/info/law/law-topic/data-protection/data-transfers-outside-eu/model-contracts-transfer-personal-data-third-countries_en.

[xi] European Commission, "Standard Contractual Clauses (SCC)," website, https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/standard-contractual-clauses-scc_en (accessed 27 September 2022).

[xii] "Guidelines on Personal data breach notification under Regulation 2016/679," Article 29 Data Protection Working Party, last revised and adopted on February 6, 2018, https://ec.europa.eu/newsroom/article29/document.cfm?action=display&doc_id=49827

Impact of GDPR on Identity and Access Management

Andrew Hindle (CIPP/E, CIPM, CIPT)

Table of Contents

Abstract

This article examines the implications of the General Data Protection Regulation ("GDPR", "Regulation") on Identity and Access Management ("IAM") process and system design. It introduces organisational and technical good practices that may help ensure demonstrable compliance with the Regulation as well as improve user experience and customer trust.

Although the focus here is on the GDPR, the approaches described may, by extension, also help in complying with data protection legislation in other geographies including (for example) the California Consumer Privacy Act ("CCPA"), or the Brazilian General Data Protection Law ("LGPD").

A Word to the Reader

We assume at least a basic knowledge of data protection and privacy - in particular, the GDPR[i] and the basic principles outlined in the OECD privacy guidelines[ii]. Even if you are not a security or privacy officer in your organization, understanding the rules will help you have better conversations with your privacy peers.

The privacy regulation landscape is evolving rapidly. Hence, the advice given here cannot be comprehensive and is neither intended nor should be considered as a substitute for legal advice. Whilst a good awareness of and sensitivity to privacy considerations is important for the digital identity professional, the majority of professionals are unlikely to be privacy lawyers. As with any area of regulation, it is always best to seek professional advice if at all uncertain.

Throughout the article, specific 'good technical practice' advice will be underlined; this same advice is also collated into a separate section at the end of the article as a checklist to follow for good IAM practices.

Introduction

Privacy conventions, regulations, and laws have been in existence for much longer than most people realise.[iii] As far back as 1948, the United Nations General Assembly enshrined a right to privacy in Article 12 of the Universal Declaration of Human Rights.[iv] In 1980, the OECD issued its "Guidelines on the Protection of Privacy and Transborder Flows of Personal Data,"[v] which were significantly revised and updated in 2013 (q.v.).

Individual countries and broader trading blocs continue to evolve their data protection and privacy regulations, in part to account for the evolution of technology. It is not uncommon for regulatory frameworks to lag behind technological developments, but changes will continue to be made in light of the impact that the Internet, connected devices, artificial intelligence, and genomics bring. To add to the complexity, the greater global mobility of individuals suggests that the local changes to data protection and privacy regulation impact changes in other jurisdictions. Aside from the changes in Europe, recent years have seen updated privacy regulations emerge in Brazil, Singapore, the Philippines, China, and parts of the United States, to name a few.

This evolution of regulation is important. When viewed through the lens of an ever-changing and evolving regulatory landscape, we can see the GDPR (and other modern privacy regulations) as a set of tools that can help us build better systems, not just as a set of checkboxes that we need to mark off.

Even if you work for an organisation that does not directly do business with Europe, certain elements of the GDPR have a global impact. Other privacy regulations may contain similar provisions; although it would be unwise today to plan for global harmonisation of these regulations, there are increasing commonalities between geographies. Recognize, too, that the GDPR applies equally to any data about individuals, whether it is data within a company about its employees or data about external individuals such as customers. In other words: the GDPR really does affect everyone.

The Regulation includes[vi] a Data Protection by Design requirement. Leaving aside the specific need to comply with the Regulation, these are fundamentally good design principles. They help mitigate business risk (e.g., the less data you have, the less interesting you are to attack, and the less impact any attack will have), and they help reduce administrative overhead and wasted effort (e.g., the less data you have, the less likely it is that you will have duplication or contradictory records).

By definition, since the GDPR is concerned with personal data, these principles have significant implications for how we design and implement systems that use such data, including IAM systems and processes. Indeed, without an IAM foundation which itself complies with the Regulation, it’s simply not possible for a final product to be compliant. (Though note that even if your IAM systems and processes themselves are Regulation-ready, you still need to ensure that your final service is compliant as well!)

The rest of this article will explore the principal considerations teams should have when developing IAM projects that can comply with the needs of the Regulation.

The Regulation applies to the physical representation of data (such as on paper) as much as it does to digital data. We’ll focus here on digital information, but we’ll make reference where appropriate to some specific implications (for example, in the areas of debugging, management reporting and so on.)

We’ll start with some general observations, including commentary on your project team’s composition and project structure. Then we’ll focus on the four stages that data - including Personal Data - goes through during its lifecycle: create, read, update, and delete.   For each of these stages, we’ll reference some of the specific areas of the GDPR that apply and identify some architectures, tools, and techniques which can help. Where relevant, we’ll note differences that might apply if you are a ‘data controller’ or a ‘data processor’ - but for the most part, the impact of these differences is more likely to be at the business/legal level, rather than the technical level. We’ll finish with a summary of key takeaways that can also be used as a quick aide-memoire for future projects or team induction.

Terminology

- Data Mapping – “a system of **cataloguing what data you collect**, how it’s used, where it’s stored, and how it travels throughout your organization and beyond.”[vii]

- Data Protection Officer (“DPO”) – An individual who must be appointed in any organization that processes any data defined by the GDPR as sensitive.[viii] The DPO is responsible for “Working towards the compliance with all relevant data protection laws, monitoring specific processes, such as data protection impact assessments, increasing employee awareness for data protection and training them accordingly, as well as collaborating with the supervisory authorities.”(See GDPR Articles 35, 37, 38, and 39 for more detail)

- Personal Data - Personal data are any information which are related to an identified or identifiable natural person.[ix] (See GDPR Article 4 (1) for more detail.)

- Data Protection by Design - data protection through appropriate technology and organizational measures.[x] See GDPR Article 25 for more detail.

General Observations

Collaboration with Privacy Advisors

With the principles of Data Protection by Design and Default, as defined by Article 25 of the GDPR, in mind, perhaps the most critical action you can take is to ensure that privacy requirements are considered at the very start of any project.

The GDPR requires many (but not all) organisations to have an appointed Data Protection Officer (“DPO”). Larger organisations may have a team of privacy advisors. If you’re leading a project that involves data about individuals, it is your responsibility to make sure your privacy colleagues are involved. Make sure you involve the relevant people in your project at the very earliest stage. Remember that they may not know about your project unless you tell them! Even if you are not, don’t be afraid to ask who is involved from a privacy standpoint, and then develop a working relationship with your advisor.

Your privacy specialist (if you do not have an experienced DPO) may not have a deep technical background, so you'll need to make sure you are providing them with the information they need in a format that makes sense to them so that they can provide complete and comprehensive advice.   To make conversations more productive and efficient, consider doing additional privacy reading or even investigating publicly recognised qualifications or certifications from relevant privacy trade bodies or other institutions as part of your ongoing professional development.

Raising Concerns/Whistleblowing

If you are worried that your project or your organisation isn't taking privacy seriously enough, or if you think you've identified an issue that leaves you out of compliance with the GDPR, or - in the worst case - an actual data breach, make sure you know the right channels through which to report this. Larger organisations should have well-established reporting/escalation mechanisms and are also likely to have whistleblowing policies and processes which you can use as a last resort. Smaller organisations may not, and so you'll need to use your best professional judgement to work out how to most effectively raise concerns. **Do** keep good records of any such conversations but **do not** include specific examples of Personal Data when reporting issues if it can be avoided, lest you make a data breach worse (or unwittingly turn a potential incident into an actual data breach!).

Finally, if your organisation has a DPO, remember that the GDPR imposes quite strict requirements on the independent relationship and reporting lines of the DPO.[xi] These can be helpful reassurances if you find you need to escalate.

Personal Data – Definition and Mapping

The GDPR has a very broad definition of what is considered to be personal data: "'personal data'' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person."[xii]

It is important, then, to be equally broad in your approach. Make sure you fully understand across the entire project when data might be considered to fall into the category of personal data. Remember that even if you are dealing with aggregated or pseudonimised data, the Regulation will still apply if that data can be 're-identified'.

Consider a process known as "data mapping" at the start of a project to discover and map out what personal data might be used where and how and monitor and update this data map through the entire lifecycle of the project. Data mapping is "a system of **cataloguing what data you collect**, how it's used, where it's stored, and how it travels throughout your organization and beyond."[xiii] Such a process is often part of a data protection impact assessment ("DPIA") but it can also be helpful in the overall design of your IAM architecture, so it's never wasted effort.

Individual tracking technologies

The use of individual tracking technologies including, but not limited to, cookies, is (at the time of writing) requires more than just the GDPR to consider if a service or website includes their use.[xiv] Each EU member state is responsible for issuing its own guidance in line with this directive, which has led to some important divergence in this area; case law is still evolving.

Some cookies can, however, fall under the GDPR - if, for example, they contain information which could be used (perhaps in conjunction with other data) to identify an individual. Hence, although perhaps not a core consideration for the IAM practitioner, it is worth being aware of and alert to potential issues in this area.

Special Circumstances

The guidance given in this article is intended to be generally applicable to all IAM projects. Some projects, however, will have particular circumstances which merit additional care and consideration. Some of these circumstances include children, law enforcement, and certain special category data. In all cases, practitioners should seek independent legal guidance.

Special Category and Other Sensitive Data

The GDPR makes special provision for certain more sensitive types of data, including (but not limited to) race, sexual orientation, political and religious affiliation, health related data and biometric data.[xv] If you are handling special category data in these areas, you will need additional safeguards.

Children

If your project involves data about Children, you will also need to take special care. The GDPR defines a Child as being below the age of 16 - but note that individual countries may (and some, including the UK, have) lower this limit to the age of 13 or below[xvi].

Law Enforcement and Personal Data

The use of personal data by law enforcement agencies is the subject of separate directives, regulations and laws; these are not considered in this article.

Automated Processing, Machine Learning, and Artificial Intelligence

The automated processing of personal data is a special circumstance in its own right, but one which merits a particular level of attention. Automated processing can include everything from machine learning ("ML"), algorithmic processing, the use of blockchain technologies, or artificial intelligence ("AI"). AI/ML, in particular, is a fast-moving field; developers, ethicists, lawmakers, and regulators worldwide and still trying to gauge the complete scope of what might be possible. It is already clear that AI/ML systems can infer or deduce 'facts' about individuals that were not part of the original data set (we often see this in user profiling). It's also clear that an original data set might not itself contain personal data (by definition) but can do once processed.

It's beyond the scope of this article to explore this area in any depth; the GDPR, however, imposes quite stringent requirements in the case of Automated Decision-Making and Profiling.[xvii]

Greenfield/Brownfield[xviii]

GDPR itself makes no distinction between greenfield applications and the refactoring of existing – 'Brownfield' – applications to bring them up to a state of compliance. From a practical standpoint, the former affords an easier path to using modern standards and techniques and is often less encumbered with legacy integration/support requirements.

Recognise, however, that GDPR compliance may drive you to review all the applications in your project's working environment. In some cases, there will be (more or less) simple technical or procedural solutions to achieve compliance. In others, however, you may need to revisit and revise the original business objectives in the light of the Regulation.

Proxy/Delegated Access

This article makes a general assumption that a data subject will be providing and/or accessing data about themselves. That said, there is naturally a variety of cases where someone might quite legitimately be accessing data on behalf of a third party.

In such circumstances, it is crucial to establish and apply appropriate mechanisms of authentication, identification, and authorisation (as recommended variously later in this article) both for the original data subject and for their proxy, along with delegation consent. In some circumstances, consent can arise via legal instruments such as a power of attorney, a court order, or similar. Establish whether this is a requirement for your use-case and design processes accordingly. You should also strongly consider maintaining a record of delegation consent and other authorisation actions where applicable. Standards such as UMA[xix] and Consent Receipt[xx] may help in this regard.

Backups

Having a reliable mechanism to secure your data in the event of a disaster is not only good general practice, it is also, essentially, required by the GDPR. Remember, though, that a poorly designed backup mechanism can potentially put you at greater risk of a breach. Ensure that data in any backup is protected with strong encryption and with other tools including, but not limited to, privileged access/user management – though be aware that these protections can complicate the restoration process. Certain sectors or applications may also require a physical or paper-based backup mechanism. Whilst this is likely to be outside of the immediate scope of responsibility of the digital identity professional, do bear in mind that the GDPR requirements apply equally to data in physical form. Backups also introduce additional complexity in the area of retention.

Data Journey

Step One - Create

The first stage of our data journey - 'create' - starts at the moment you set out to collect personal data. Do not confuse this with the moment you write the data into a database (or other storage mechanism). Before you even request data from (or about) the data subject, you need to clearly understand and communicate to them what you are collecting and why, along with outlining their data subject rights. These are most commonly expressed via a privacy notice that uses clear and plain language - and you should at least ensure that the notice accurately reflects the way your system actually works!

Depending on the lawful basis for processing the relevant data, you may need to obtain the consent of the data subject for you to collect and process their information. How you obtain consent will differ from project to project, depending on what data is being collected and what it is being used for. Your privacy advisor can provide guidance.

From an audit perspective, consider keeping a record of that consent and/or providing your data subject with a record for themselves - evolving standards such as the Consent Receipt may be applicable here. Do remember, though, that any such receipt or record may itself contain Personal Data!

Create Minimally

If Data Protection by Design and Default formed the first guiding principle for your project, then your second guiding principle should be that of Data Minimisation. Data minimisation is good practice irrespective of compliance requirements: the less you collect or process, the less you have to protect and manage over time. It is also one of the 7 principles established by the GDPR for the handling of personal data.[xxi]

The bottom line: When collecting data from a data subject, collect and keep **as little data as you possibly can** in order to meet your requirements. Similarly, for indirect data about a data subject, such as browser fingerprints,[xxii] collect and keep as little data as you possibly can. This means you need to have a good understanding of the business rationale for the project, so that you are clear about the justification and so that you can help your colleagues on the business side meet their obligations: it's always helpful to ask *why* a given piece of information needs to be collected.

Remember that the GDPR considers data in the aggregate. Consider whether there is any possibility of data your project is collecting being combined with other data the organisation holds in such a way as might result in identification of the individual (see also 'read' below). Avoid repeat collection of data that your organisation already holds about an individual. Aside from being a frustrating experience for the user, this also results in duplicate and/or conflicting records, which can cause problems with data accuracy, subject access requests, deletion, and other areas of the Regulation. If you have a large and disparate data map, consider using data discovery or meta-directory tools to help with visibility and consolidation.

Bear in mind that you may be collecting *implicit* or inferred data, which may also qualify as personal data: IP addresses, for example, or system analytics. These will need handling with the same diligence as data you *explicitly* request from or on behalf of a data subject. Even if this data is collected and used on a transient basis, it still needs handling correctly.

Consider also creative ways to limit the amount of data you collect. Besides simply collecting less, an organization might use an attribute service for answers to questions such as 'is the data subject over the age of 18', instead of collecting and storing the subject's date of birth, or requiring them to disclose credit card information. Technologies which can provide evidence that an authority has knowledge of certain information without revealing the information itself — zero-knowledge proof[xxiii], for instance — is also worth investigation. Be aware, however, that existing legal requirements may not yet take such technologies into account.

As noted earlier, this article mainly considers the impact of the GDPR on digital identity. However, the moment of data collection/creation is often where paper-based processes occur. Even if these are not your direct concern, it's no bad thing to make sure you understand how any paper records are being processed.

Possibilities for Federation

Having dealt with the basics, you now need to ask an important question: does your use-case actually need full user account creation. There is a tendency - born out of years of experience - to gravitate towards this as the first port of call in any identity project. Yet, in many cases, it's unnecessary; or it's something that only becomes needed later in the customer journey. Established standards like SAML or OpenID Connect support transient identity federation; this is often all you need. In such a case, you are only handling personal data (if at all) for a brief period of time, and so the normal data minimisation principles and precautions for data in transit may be sufficient. (use the most current version of TLS, plus additional specific data encryption as necessary)

If you do need a user account for technical reasons — session data persistence, for example — can it be made essentially 'impersonal' though the use of (for example) pseudonymous federation? Pseudonymisation allows for the user identity to be matched, using an identifier that cannot easily be associated with a known individual. Take care in this case, however: it can be possible to combine data in such a way as to re-identify the information, so defeating the purpose of pseudonymisation. Pseudononymous data is still considered personal data, and as such it must be considered against the requirements of the GDPR.

Storing Data

If you do find you need to persist data — whether pseudonymously or not — you will need to think about where and how you store the data. The usual protections for data at rest are important. Use appropriate encryption techniques and keep these under routine review: cryptography is an area of rapid development (particularly given the advent of quantum cryptography and the evolution of 'quantum-safe' algorithms and techniques).   You should also ensure that the right processes are in place to keep supporting systems, applications, and libraries up to date and patched.

Other GDPR requirements notwithstanding, modern application design patterns will almost certainly lead you to provide an API for handling your personal data. In such cases, access to such APIs must be protected, ideally using a protocol such as OAuth; you could also consider using an API gateway. We'll come back to API protection again later in our data journey.

If you are considering a storage solution using a distributed ledger, you should take extra care. There is now clear consensus that storing personal data directly in such a ledger is not good practice. Some solutions under development today may avoid this particular pitfall, but it is still worth bearing in mind, particularly if you are building your own. Until this area of technology is more stable, the best advice is to proceed with caution; to keep such projects under regular periodic review, even after deployment; and to ensure you have a well-documented and easily implemented way to reverse out of using the ledger-based solution, should that become necessary.

Using a cloud-based data or user store may have benefits from a risk management and privacy perspective. Ensure that you work with your privacy team so that your privacy notice accurately reflects the relationship between you and your provider.

Location of Data Storage

The GDPR does not itself impose requirements of data territoriality – that is, it does not require that data be stored in a particular geography – though regulations in other jurisdictions do. You should, at the very least, develop a flexible architecture that will allow you to segregate data on a regional basis should that become necessary — although bear in mind that this could mean collecting additional personal data which you might otherwise not need.

With that said, the GDPR **does** have requirements around the transfer of data outside of the European Union (i.e. to a “Third Country”). The transfer of personal data to any Third Country must always be a significant concern in the context of GDPR, and – although solutions can certainly be devised – this is an area of ongoing regulatory development. You will need careful discussion with your privacy adviser to make sure this is being handled correctly.

Step 2 – Read

Any and all access to the personal data you hold must be kept secure. At the most basic level, this means ensuring that you minimise any such access. If you are not already doing so, consider deploying a Privileged Access/User Management solution where applicable. You should also ensure that even those authorised privileged users, including database and systems administrators, cannot get access to personal data in clear form - even accidentally. Remember that **any** unauthorised access to personal data constitutes a potential data breach. Such a breach may be more or less severe and have greater or lesser consequences… but it is still a breach.

In order to provide useful functionality, whilst avoiding a potential data breach, be sure to use secure modern methods to authenticate and authorise your users, both internal and external. Use multiple factors of authentication; consider FIDO authenticators; avoid SMS as a factor; consider modern authorisation standards (and products which support them), including established protocols like XACML, newer standards like User-Managed Access ("UMA") and emergent approaches such as Transactional Authorization.

Note that 'authentication' is not necessarily the same as 'verification'.   You may not need to establish the user's actual physical identity to any level of assurance in order to safely satisfy their request. However, where some level of assurance to a real-world identity is required, remember to treat any data used to verify the identity of the user with an appropriate level of security.

If you are pre-populating client-visible forms, be especially careful that such data is only displayed to the correctly authorised user, and that it cannot be cached across the visits of different users.

Modern application design patterns will likely mean that you have an API for 'read' operations. As noted earlier, any such API must be properly protected. Consider also adding additional program- or system-level protections: for example, protecting against multiple sequential reads by requiring additional authorisation or by imposing a total read limit or a repeat-time restriction.

Be conscious of other systems which may have access to personal data - security applications (especially ML or AI-driven solutions) and data mining tools, for example. Make sure such systems don't have unauthorised or unnecessary access to personal data in the clear, and be aware that in some cases, such access might constitute automated decision making or profiling (as referenced earlier).

Consider also unintended consequences. If you have a reporting tool which (for example) generates an Excel spreadsheet of data which can then be emailed, consider (a) whether all the PII needs to be in there; and (b) whether you can provide protection in some automated way upfront (for instance - by automatically creating an encrypted sheet, rather than relying on the user to have it do that for themselves), to help reduce the risk of an accidental breach further down the line.

Data Subject Access Request and Data Portability

The subject of the personal data has the right, under GDPR, to access the personal data you hold about them.[xxiv] This presents an obvious breach risk. If you are handling a response to a data subject access request, or if you are designing a system to be used for such a case, then you must be particularly careful to ensure that you correctly authenticate and/or verify the user; that they are properly authorized; and that the data you are sharing does not itself contain the personal data of other data subjects.

You are also required under GDPR to provide all the subject's personal data in a machine-readable format for data portability. The same security considerations apply in this case.

Somewhat perversely, in order to help satisfy some of these requirements, you may need to collect (or infer) more personal data than you might prefer, although you should always be careful not to collect more data that you absolutely need. For example: you may need to establish what country a given user lives in, is in, or is a citizen of, in order to establish what legislation applies! Depending on your system design, you can perhaps avoid storing this information and instead request it in real-time when the decision needs to be made (and verify it as needed).

Data Breach Reporting

Breach reporting is a special case in the context of 'read': if you are required to report a breach or a potential breach, you must ensure that you do not send personal data as part of your breach report. If you have automated breach or security reporting tools, make sure these tools don't accidentally create or worsen a breach by including personal data in their reporting. Consider also the use of privacy software solutions that can help search across data sets securely.

Step 3 - Update

GDPR mandates that data subject should be easily able to correct any personal data you hold about them. Make sure your system has such a mechanism. User self-service solutions can be particularly helpful in this regard, as long as they are appropriately easy to find and to use. Again, proper authentication and - in some cases - verification is crucial to mitigate against a potential accidental breach.

It is worth noting that this ‘update’ requirement of the Regulation may have implications for distributed ledger-based solutions. In particular, you should establish whether such a solution will allow for the rectification of a historical record in the ledger (or on the chain). Simply marking the historical record as ‘no longer active’ is unlikely to be sufficient.

Step 4 - Delete

In some instances, the GDPR provides the data subject with a right to request that the data you hold about them be deleted. You will need to make sure you have a straightforward way to do this - and that this mechanism is secured against accidental or deliberate misuse with appropriate safeguards including necessary levels and methods of authentication and authorisation.   Consider maintaining audit logs for such transactions (bearing in mind that you will want to keep the actual personal data out of the log record), and potentially having a time-limited ‘roll-back’ mechanism in the event of an error.

The Regulation also requires that data be stored only for the period it is actually needed. Business requirements, informed by privacy needs, will dictate the length of the retention period; but you will need to design your system such that data can be easily expunged at the end of this period.   Consider maintaining a separate record indicating when the data in question was originally created and running an automated task either to report on the data which has reached its retention date (hence flagging it for manual deletion) or to remove it directly.

For large and/or brownfield deployments, you may need to run a discovery process in order to establish what data you actually hold about a given data subject. There exists a variety of software solutions that can facilitate this.

As with ‘Update’, If you have an API (or other facility) which can perform data deletion - and especially if you allow for bulk delete - make sure you protect against misuse. For instance: add an additional (even a manual) check before a bulk delete or require additional authorisation for requests exceeding a certain number of records. You should also ensure you have a way to routinely back-up data and to restore in the event of a mistake (or a deliberate attempt to corrupt data), and consider forcing a backup via your API code before the delete process runs. Recall that retention of any such backup copies must be limited.

Conclusion

GDPR - and other modern data protection and privacy legislation and regulation - means we have to take extra care in designing, developing, and maintaining our IAM solutions. In particular:

- Collect only the data we need
- Only keep it for as long as it is needed
- Look after it when it is in our care
- Make sure it can only be accessed by those who should have access
- Make sure it can be appropriately updated
- Dispose of it safely when it is time to do so

We already have the tools we need to do this, but we need to be careful to apply those tools in the right way and to ensure that business owners aren't asking us to do things we shouldn't be doing:

- Only create accounts if absolutely necessary; use federation (SAML; OpenID Connect) or other transient or non-identifying information where we can (User Info; Zero-Knowledge Proofs)
- Authenticate users, preferably with strong and multiple factors of authentication (FIDO)
- Authorise users, preferably with modern protocols (XACML and UMA)
- Protect APIs (OAuth)

Much of what we need to do isn't new, and much of it has always been good practice. It's just not necessarily been standard practice or even top of the list for projects. New privacy regulations give us the opportunity to do things the right way.

Your IAM Project Checklist

- Ensure that privacy requirements are considered from the very start of a project, and routinely re-evaluated through the lifetime of the application
- Involve the relevant people (people who represent organizations consuming the IAM data as well as those serving as sources of truth for your IAM data, together with your privacy peers) in your project at the very earliest stage.
- Do keep good records of any conversations around potential data breaches but do not include specific examples of Personal Data when reporting issues.
- Map what, where, and how personal data might be used; this will be valuable input to a more complete Data Protection Impact Assessment (DPIA)
- If you are handling special category data as defined by GDPR and/or your local or sectoral privacy regulations, you will need additional safeguards.
- If your project involves data about Children, you will also need to take special care.
- Ensure that your organization's or service's privacy notice accurately reflects the way the system actually works!
- Collect the consent of the data subject for you to collect and process their information.
- Keep a record of that consent and/or providing your data subject with a record for themselves.
- Explore the Consent Receipt specification and emerging implementations.
- Collect as little data as you possibly can (data minimization).
- Avoid repeat collection of data.
- Consider using data discovery or meta-directory tools to help with visibility and consolidation.
- Explore zero-knowledge proof technologies and implementations and investigate whether such solutions should form a part of your deployment
- Instead of creating an account, consider instead using transient identity federation and/or single sign-on. If account creation cannot be avoided, consider using pseudonymous federation and/or single sign-on to reduce the amount of identifiable personal data you hold.
- Use the most current version of TLS plus additional specific data encryption as necessary.
- Use appropriate encryption techniques and keep these under routine review.
- Keep supporting systems, applications, and libraries up to date and patched.
- Protect access to APIs that handle personal data, ideally using a protocol such as OAuth.
- Storing personal data directly in a distributed ledger is not good practice.

- Develop a flexible architecture that will allow you to segregate data on a regional basis.
- The transfer of personal data to any Third Country (as defined in the Regulation) must always be a significant concern.
- Access (physical and digital) to the personal data you hold must be kept secure.
- Consider deploying a Privileged Access/User Management solution.
- Ensure that even those authorised privileged users, including database and systems administrators, cannot get access to personal data in clear form.
- Use multiple factors of authentication; consider FIDO authenticators; avoid SMS as a factor; consider modern authorisation standards (and products which support them), including established protocols like XACML, newer standards like UMA and emergent approaches such as Transactional Authorization.
- Be careful that Personal Data is only displayed to the correctly authorised user, and that it cannot be cached across the visits of different users.
- Be particularly careful to ensure that you correctly authenticate the user and that they are properly authorized.
- Avoid storing personally identifiable information, and instead request it in real-time when the decision needs to be made (and verify it as needed).
- If you discover a breach in your system, do not send personal data as part of your breach report.
- Make sure your system has a self-service mechanism to support the correction and/or deletion of a user's personal data.
- Consider maintaining audit logs for such transactions (bearing in mind that you will want to keep the actual personal data out of the log record).
- Consider maintaining a separate record indicating when the data in question was originally created and running an automated task either to report on data which has reached its retention date (hence flagging it for manual deletion) or to remove it directly, in line with your privacy policy and notice
- Check before a bulk delete and require additional authorisation for requests exceeding a certain number of records.
- Ensure you have a way to routinely back-up data and to restore in the event of a mistake (or a deliberate attempt to corrupt data), and consider forcing a backup via your API code before the delete process runs.

[i] For an overview, read the IDPro Body of Knowledge GDPR article. The full text of the regulation can be found at https://eur-lex.europa.eu/eli/reg/2016/679/oj

[ii] Organisation for Economic Co-operation and Development, “The OECD Privacy Framework,” 2013, https://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf

[iii] An overview of the history of the GDPR can be found at https://cloudprivacycheck.eu/latest-news/article/a-brief-history-of-data-protection-how-did-it-all-start/

[iv] United Nations, “The Universal Declaration of Human Rights,” 1948, https://www.un.org/en/universal-declaration-human-rights/

[v] Organisation for Economic Co-operation and Development, “OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data,” 2013, http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm

[vi] See Article 25 of the Regulation

[vii] KJ Dearie, “What is Data Mapping? The Importance of Data Mapping for GDPR Compliance,” Termly, 30 October 2018, https://termly.io/resources/articles/gdpr-data-mapping/

[viii] “GDPR: Data Protection Officer,” Intersoft Consulting, https://gdpr-info.eu/issues/data-protection-officer/

[ix] “GDPR: Personal Data,” Intersoft Consulting, https://gdpr-info.eu/issues/personal-data/

[x] “GDPR: Privacy by Design,” Intersoft Consulting, https://gdpr-info.eu/art-25-gdpr/

[xi] See in particular Article 38 of the Regulation.

[xii] Article 4 of the Regulation

[xiii] Dearly, “What is Data Mapping? The Importance of Data Mapping for GDPR Compliance,” https://termly.io/resources/articles/gdpr-data-mapping/

[xiv] European Parliament, Council of the European Union,
“Directive 2009/136/EC of the European Parliament and of the Council,” November 2009, http://data.europa.eu/eli/dir/2009/136/oj

[xv] See Article 9 of the Regulation.

[xvi] See Article 8 of the Regulation.

[xvii] See Article 22 of the Regulation.

[xviii] Greenfield is a term used to describe a project with no prior work to constrain its development. Brownfield, in contrast, refers to projects with predetermined limitations based on having to work in an existing platform or under pre-existing constraints.

[xix] Specifications and Auxiliary Documents, User Managed Access Working Group, Kantara Initiative, https://kantarainitiative.org/confluence/display/uma/Specifications+and+Auxiliary+Documents

[xx] Lizar, Mark and David Turner, eds. “Consent Receipt Specification,” Consent & Information Sharing Working Group, Kantara Initiativehttps://kantarainitiative.org/file-downloads/consent-receipt-specification-v1-1-0/

[xxi] “The Principles,” Information Commissioner’s Office Guide to the General Data Protection Regulation (GDPR), https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/.

[xxii] Katarzyna Szymielewicz, Bill Budington, “The GDPR and Browser Fingerprinting: How it Changes the Game for the Sneakiest Web Trackers,” Electronic Frontier Foundation, 19 June 2018,

https://www.eff.org/deeplinks/2018/06/gdpr-and-browser-fingerprinting-how-it-changes-game-sneakiest-web-trackers

[xxiii] "Zero-knowledge proof," Wikipedia, last updated 24 January 2020, https://en.wikipedia.org/wiki/Zero-knowledge_proof.

[xxiv] "Art. 15 GDPR Right of access by the data subject," Intersoft Consulting, https://gdpr-info.eu/art-15-gdpr/

Review - ISO/IEC 24760-1:2019

"IT Security and Privacy - A framework for identity management - Part 1: Terminology and concepts," *International Organization for Standards*, Technical Committee ISO/IEC JTC 1, Subcommittee SC 27, May 2019, https://www.iso.org/standard/77582.html.

Reviewer: Corey Scholefield

Abstract

This review offers insight into the first part of the ISO standard for Identity Management, ISO/IEC 24760-1:2019, which covers terminology and concepts.

Review

ISO/IEC 24760-1:2019 provides an introduction to the vocabulary of the identity management space, with definitions of key terms in common usage within the community. Under review is the 2nd edition of the document, revised for 2019-May.

As stated in the introduction to the document:

> *The goal of this document is to specify the terminology and concepts for identity management, in order to promote a common understanding in the field of identity management.*

According to its abstract:

> *It is applicable to any information system that processes identity information.*

The document supports the goal by offering brief definitions of community-standard terms, such as:

- *identity*
- *attribute*
- *identifier*
- *principal*
- *identity-proofing*

While the tone of the document is slightly academic, the definitions themselves:

- are written using terms familiar to English speakers;

- include other terms that appear in the document, with convenient links to access their definitions easily;
- include some examples to illustrate the usage of the term or to illustrate the concept.

This document only includes terminology, concepts, and brief definitions and outlines. It is intended to be used as reference material.

The authors have made some effort to ensure that these definitions can be applied to a broad set of use cases, i.e., definitions of identity for use within human and non-human (device) contexts. This treatment keeps some of the coverage at a high level, causing the supporting examples to be quite helpful for providing a real-world abstraction of some concepts. That being said, this reader would have appreciated a few more examples to help support some of the definitions.

The article only contains one figure, which is supportive of the concept it depicts. The document could be improved by using more illustrations to outline concepts.

This document intends to provide authoritative definitions of terms and concepts, so other documents probably use this one as a reference document. The bibliography section is excellent and provides links to many other foundational documents in the contemporary identity-management space. Many of those references are freely available for download.

A reader who needs a basic introduction to the common terms included in this document will find this material very helpful, as the terminology is very relevant in contemporary identity-management conversations.

The seasoned reader will also find this a useful reference document but may also wonder about omitting terms such as *persona*, *account*, or *authorization*. It could be that these terms might not fall within the strict scope of identity management that the authors wished to cover in this document. Instead, those terms may fall under the category of access management, a connected but separate body of information security knowledge.

The document does not support any treatment of identity in a social science concept, so the definitions should be taken as they apply to identity management in technical use cases.

Review - ISO/IEC 24760-2:2015

IT Security and Privacy - A framework for identity management - Part 2: Reference architecture and requirements

“IT Security and Privacy - A framework for identity management - Part 2: Reference architecture and requirements,” *International Organization for Standards*, Technical Committee ISO/IEC JTC 1, Subcommittee SC 27, June 2015, https://www.iso.org/standard/57915.html.

Reviewer: George Dobbs

Abstract

This is a summary of what is in ISO/IEC 24760-2:2015, one of the core ISO standards on IAM, along with an opinion on its suitability for use by the identity practitioner.

Review

This document is formal in nature and provides a rigorous model of an identity management system and a notion of what should be include in the design for that system. Those without architectural background may find the approach to be too academic, but if you are looking to add a degree of rigor to your plans, read on.

After the preliminaries the text jumps into a set of viewpoints that provide a minimally acceptable documented design; these are a context view and a functional view. The document then describes these views in terms of definitions, concerns, and models.

Moving on from viewpoints, the text takes up the two mandatory views in some detail. For the context view the text elaborates on stakeholders, actors, context model, use case model, and compliance and governance model. The elaboration regarding stakeholders may be useful as it identifies some of the stakeholders that are often forgotten, such as regulatory bodies, and the rarely mentioned consumer/citizen representative or advocate.

The text not only lists the set of stakeholders to consider; it also identifies their concerns. The text distinguishes between stakeholders and actors although there is significant overlap in the lists. Where stakeholders have concerns, the actors have responsibilities and, in some cases, provide capabilities, both of which are listed. The diligent reader may want to study the text of the actor section carefully as it is quite precise and conveys a lot of concepts in a small space.

The text moves on to the use cases. The text provides a simple use-case as an example, then describes several more specific classes of use-case (employee, employer, principal, and device). Interestingly enough, the text does not call out specific use case for Customer/Citizen, although it does for Employee. Instead these concepts are included in the Principal use cases section. Additional examples are given in Annex B. These examples should be useful. For a practitioner fluent in universal modeling language (UML) the diagrams and use-case section should be straightforward. For others, this may be harder hill to climb.

The context view is rounded out with a short section on what should be included in a compliance and governance model. This section provides a checklist.

Next up is the functional view. This section lists interactions expected between the actors and architectural elements in the system. It covers ten processes including maintenance of identity information, access to identity information, winding up with less common processes such as identity authority discovery and publication of identity information (under a policy). Again, these are tersely worded but should provide useful checklists to the practitioner. The functional components are laid out as a UML diagram in Annex C, which brings in a couple of new items such as “Trust Root”.

Before moving on to the requirements section, the text outlines 4 scenarios. The scenarios are used to determine the trust relationships that are needed. This brief section encourages the architect to design for confidentiality, integrity and trust needed by each scenario. There is very little detail provided. For instance, the federation scenario is described in abstract terms but there is no mention of the common notions of identity provider or relying party. But it does encourage the architect to at least consider what scenarios are desired, helping to establish requirements.

The main text wraps up with a listing of requirements, both functional and non-functional. This is an excellent source to use in establishing the requirements for a system.
In addition to Annexes A - C, the document provides Annex D, which elaborates on selected business processes including consent management, credential lifecycle management, configuration management (in a federation), policy management, and principal’s life cycle management.

Overall the document is very formal and structured, which enables consumption of the core foundation concepts of Identity Management. Access management is referenced only in a reflexive mode – to control the access to identity information itself. The more general access management may be covered in another ISO/IEC document: ISO/IEC 29146, *Information technology — Security techniques — A framework for access management.* This may lead the reader to a point of frustration if looking for details about authorization and authentication.

The reviewer finds this document to be valuable for the identity practitioner who is confronted with developing a new identity system or evaluating the current state of an identity system in order to mitigate gaps and shortcomings. It provides a structured framework of concepts that can be used to inform such work. That being said, it is a text that requires the reader to bring significant powers of mind and experience to the reading. The presentation does not cover access management; it focuses entirely on identity management. It refers to another ISO/IEC document for access management. There are a few other off-document references, but this reviewer feels those can be skipped without affecting the understanding too much. This document is appropriate for those seeking to build or revise a robust identity system and are seeking to compare their own thoughts to the work of others in order to gain assurance that a complete design has been produced.

Review – ISO/IEC 24760-3:2016

Information technology — Security techniques — A framework for identity management — Part 3: Practice

"IT Security and Privacy - A framework for identity management - Part 3: Practice," *International Organization for Standards*, Technical Committee ISO/IEC JTC 1, Subcommittee SC 27, August 2016, https://www.iso.org/standard/57915.html.

Reviewer: Espen Bago

Abstract

The document reviewed here is the third and final part of the ISO/IEC 24760 standard, focusing on "Practice", which in the abstract is described as providing *"guidance for the management of identity information and for ensuring that an identity management system conforms to ISO/IEC 24760-1 and ISO/IEC 24760-2".* Parts 1 and 2 covers "*Terminology and concepts"* and "*Reference architecture and requirements".* ISO/IEC 24760-3 is in its first edition, dated 2016-08.

Review

An important note here is that this review is written looking exclusively at Part 3 without having detailed knowledge of the prior parts. Based on the references within Part 3 to the other parts, this document is not intended to be used in isolation, but since each part is licensed and sold separately, reviewing it in isolation from the other parts may give an indication of its individual worth.

ISO/IEC 24760-3 states its own purpose as to *specify relevant concepts, operational structures and practices that may enable the required assurance and control for use of both identity information and identity management systems*. The implication is that this document provides good practices for identity management, with the main target audience being those who are starting an identity initiative or need to better control an ongoing initiative of this sort.

This intention of providing practices for achieving central and typical goals within identity is laudable, and it is something often searched for by practitioners. But this document fails to deliver on the promises due to several factors, the most important ones being inconsistency of structure and inconsistency of content in each section.

The core of the ISO/IEC 24760-3 are the 12 pages about risk mitigation for identity, identifiers and identity information, auditing and about control objectives and controls, with this last section on controls and objectives taking up the main part. These sections list advice (practices) for different parts of the work necessary when setting up and maintaining identity management systems, and when extracting that information, there is plenty of useful information that could read as a checklist of advice and suggestions.

The challenge is that getting to that useful information and extracting it, is hard due to the convoluted setup in subsections that are difficult to follow, especially since the subsections do not consistently contain the same level or detail of information. Thus it is unnecessarily challenging to understand the given practices either as a whole or to find the relevant, sought after practice for a given situation. Additionally, when found, such information tends to be very simplistic or high level. As an example, the section auditing an identity management system mainly states that audits should be done, and that their purpose should be to validate that the system functions in accordance to its requirements and policies.

A future revision of this standard would benefit from simplifying its section structure, with emphasis on making it clearer what it is trying to express. Possibly, since most of the information is very high level in nature, a format closer to a checklist might also be beneficial.

As it stands now, this standard is most accessible to the most experienced practitioners, since they are better equipped to navigate the document. But these practitioners are also those least in need of the information, since they normally already know most of the practices from experience. Most practitioners new to the area would struggle putting the current (2016-08) version ISO/IEC 24760-3 to use for the stated purpose.

There are no figures in the main body of the document, which seems reasonable as the practices described do not lend themselves to be easily visualized.

Apart from the aforementioned core of the document, half of the ISO/IEC 24760-3 are taken up by two annexes. The reason these are not so far reviewed as being core, is that nothing in the text refers to them, and they are not directly related to anything in the preceding text. Put simply, these annexes of 16 pages out of the total 38 appear out of place, giving the impression that they were included to reach a certain page length.

That being said, the two articles in the annexes are well written and cover interesting areas. Had they been directly relevant to the stated purpose of the standard, the annexes would be enough to warrant a recommendation of the whole document.

For reference, the annexes, including descriptive figures and diagrams, cover practices for federating identity (or potentially rather access) management systems - annex A - and a breakdown of what attribute-based credentials are and how they can be used for authentication. Anyone needing either specific information on setting up federated systems, or working with attribute-based credentials, would probably find this document worth perusing.

Introduction to OAuth 2.0

By Bertrand Carlier

To comment on this article, please visit our GitHub repository and submit an issue.

Table of Contents

Abstract

This article introduces a widely deployed protocol named OAuth 2.0 (**O**pen **Auth**orization 2.0, commonly referred to as OAuth2). It is used extensively by large social media service providers and many other web-based Internet services today.

About OAuth2

In a nutshell, this standard protocol aims to allow access from a **client application** (a website, a mobile application, an Internet-connected device, etc.) to a **protected resource** (e.g., an API), possibly on behalf of a **resource owner** (e.g., the end-user). It can be associated with several transport protocols but has been very popular to secure REST web services.

This article will focus on the current published standards; work is underway in the [OAuth working group]() in the IETF to update some of this material. For more information on how OAuth came about and its relationship with other authentication protocols, see Pamela Dingle's IDPro Body of Knowledge article, "Introduction to Identity - Part 2: Access Management."[i]

OAuth2 can be considered a three-step protocol:

1. Get an access token
2. Use the access token
3. Validate the access token

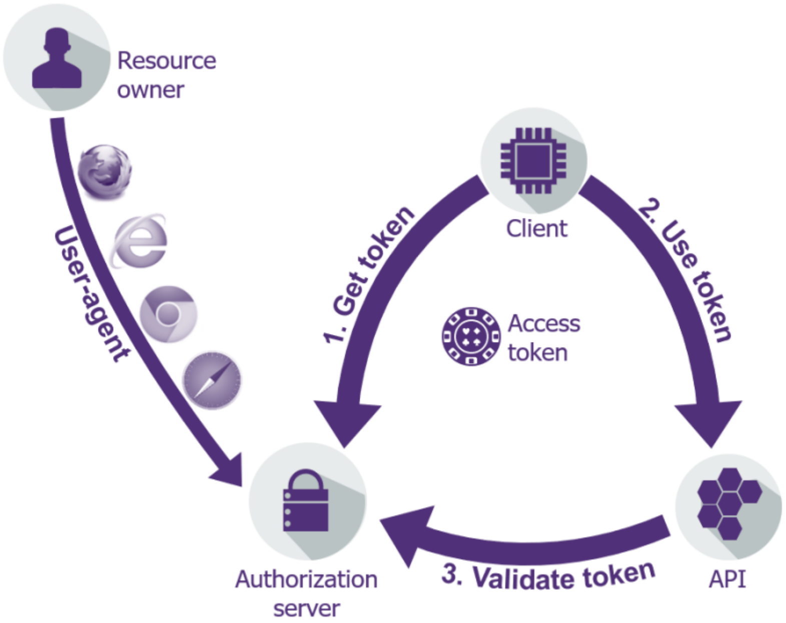


Figure 1: High-level diagram of OAuth2 flows

When looking into the OAuth2 specification space, you are quickly surrounded with many documents, making it difficult to determine the easiest path to follow.

Let's see where to start the journey and where to head.

Terminology

Term	Definition
Client	A client application consuming an API
Protected Resource	An API in the OAuth2 terminology
Resource Owner	An end-user using the client application and granting it access to the protected resource
Authorization Server (AS)	The OAuth2 server is able to authorize a client, issue tokens, and potentially validate tokens
Scope	A string designating a (part) of a protected resource that a client is authorized to access
Bearer token	A token whose possession is sufficient to enable access to a protected resource
Sender constrained token	A token whose possession is not sufficient to enable access to a protected resource (additional proof of identity by the client application is required)
Access token	The OAuth2 token that allows a client to get access to a protected resource
Refresh token	The OAuth2 token that allows a client to renew an access token when it is expired without the user’s presence

Where to start

OAuth2 is defined through a series of IETF RFC documents that each describe a specific aspect, use case, or profile of use of the protocol.

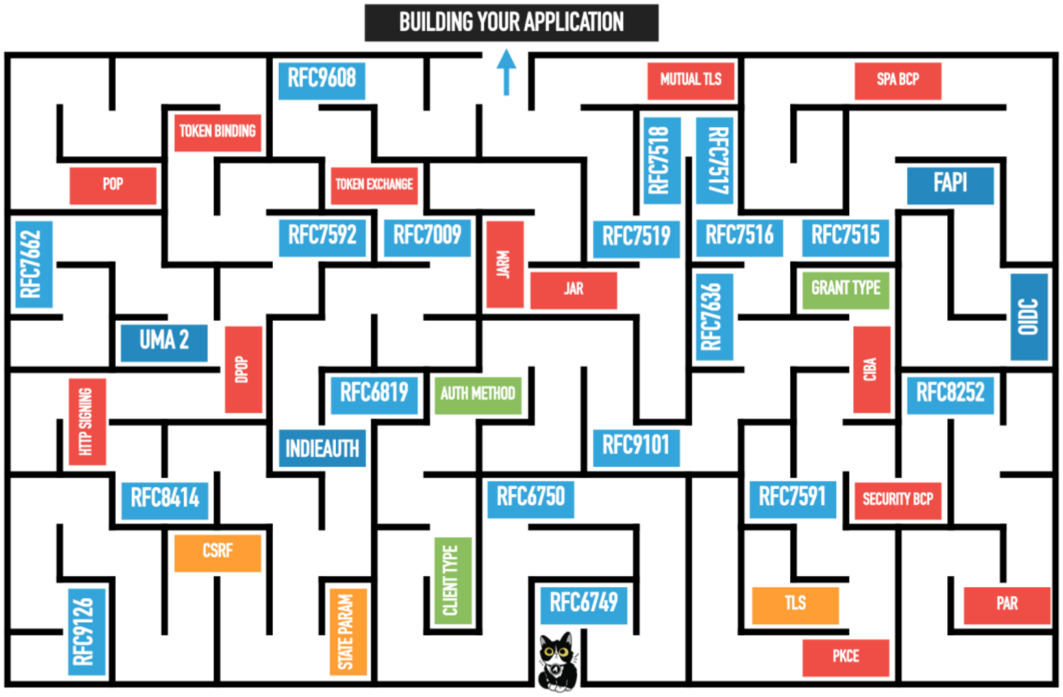


Figure 2: An artistic rendering of OAuth and related standards, courtesy of Aaron Parecki

Everything starts with two RFC documents:

- RFC 6749 - The OAuth 2.0 Authorization Framework defines four ways for a client application to obtain a token from an authorization server (two of those are now deprecated). Those are called flows or authorization grants.[ii]
- RFC 6750 - The OAuth 2.0 Authorization Framework: Bearer Token Usage defines the way for a client application to use a token in a subsequent request to a protected resource.[iii]
- Later on, different documents would help with the validation step:
  - RFC 7662 - OAuth 2.0 Token Introspection defining token introspection against the authorization server, which can be used to verify token validity and extract data from the token.[iv]
  - or RFC 9068 - JSON Web Token (JWT) Profile for OAuth 2.0 Access Tokens defining a JWT profile for the access token.[v]

Let's use this breakdown to see what OAuth2 offers.

Get a Token:

This step can be seen as a two-step process: first, the client must be authorized for an access token, then the client will perform a token request.

- As mentioned above, of the four initial ways to obtain a token, two are deprecated following OAuth2.1 (currently draft):
  - Resource Owner Password Credentials, which encouraged an anti-pattern of sharing end-user credentials with the client application
  - Implicit flow, which made extensive use of the browser's front channel and therefore introduced security issues
- The two recommended flows remaining are the following:
  - **Authorization code flow** is the recommended way to obtain a token when a resource owner is present and needs to authenticate first and then consent to delegate access for the client application to the protected resource. This flow uses redirections within a user-agent, typically the user's browser, as well as a back-channel request to eventually obtain the OAuth2 Access Token.
    There is a first step to authorize the client to get an access token and then a second step where the client actually gets the token.
    An additional protection to the original Authorization Code flow is now recommended in order to tighten the security of OAuth2 authorization and deliver the Access Token to the legitimate client that initiated the request. The name of this additional protection is PKCE (for Proof Key Code Exchange, pronounced "pixie," as defined in RFC 7636) and is considered a good approach to handle public clients.[vi]
  - **Client credentials** aim to authenticate the client application only to deliver the access token (in that case, the AT is not linked to an end user's identity but only to the client application identity). This flow is suited for application-to-application access.

Use the Token

This step aims to use the access token while calling the protected resource.

RFC 6750 describes how an access token should be conveyed to a protected resource. In a very brief summary, and in order of preference, the token should be passed as:

- An HTTP header as a bearer token (Authorization: Bearer <access token>)
- A POST parameter
- A GET parameter (aka Query String parameter)

Validate the Token

Finally, the protected resource receiving a token needs to check the token's validity. This token validation was, for a long time, left to implementations to define how to proceed:

- The token format is not specified and can be anything from a randomly generated opaque string acting as a reference token to a quite frequently witnessed JWT signed value token (RFC 7519), but it can be anything that would fit the designers of any given implementation.[vii]
- If the token is opaque to the client as per the RFC, no specific instructions are defined regarding how the protected resource should validate it. It relies on an out-of-band and beyond-the-scope-of-the-specification process to agree between protected resource and authorization server on how to validate a token: digital signature validation and possibly decryption of a self-contained token (see RFC 9068 for standardization of this approach using JWT as the token format) or introspection of a reference token against an Authorization Server (AS) endpoint (see RFC 7662 for standardization of this approach).

It is generally recommended to rely on one of those two documents to help with interoperability between the protected resource and the authorization server

Beyond the Basics

This section of the article now gives additional details on more aspects of the OAuth2 protocol and additional specification documents.

Scopes

OAuth2 does not allow a client application to access any resource without restriction once it has an access token. An authorization request and, ultimately, the issued token holds a scope (which is a list of space-delimited, case-sensitive strings) that will allow the protected resource to determine if the authorization was indeed given to access it.

Get a Token (Also)

A few additional ways to obtain an access token were later provided through additional specifications:

- SAML profile and JWT profile will allow the delivery of an access token in exchange for, respectively, a SAML token or a JWT token issued for a specific end-user or crafted by the client application itself in order to authenticate itself.[viii]
- Device flow will allow Internet-connected devices to retrieve an access token even if they can't display a browser or are input-constrained.[ix] This flow will rely on the end-user using another device (e.g., a browser on a smartphone) to complete part of the sequence.
- Token exchange will enable an access token to be issued in exchange of any other security token and will provide guidelines to correctly implement delegation or impersonation.[x]

Tokens

Until now, only the access token was mentioned. It is the core token that OAuth2 provides to client applications. This token is generally a bearer token, meaning that any entity that gets access to it can use it to access the protected resource. This characteristic has several security implications:

- The protected resource cannot be sure that the client application currently requesting access is the same one that initially obtained the token
- The end user that may have had to be authenticated to allow the token to be generated may not be present anymore

Access tokens, therefore, can have different characteristics to mitigate those implications:

- Time-limited tokens. The specification recommends that the access token has a limited lifetime.
- Sender-constrained tokens. Recent specifications (mTLS, DPoP, etc.) allow that access tokens can be bound to the initial client application using various mechanisms, generally involving proof-of-possession of a cryptographic key both at the token request and at the token usage and that the token is linked to that key material (through a public key thumbprint for instance).[xi] As a consequence, a sender-constrained token can only be used by the application that requested the token. It is worth noting that while approaches like DPoP can protect against a stolen token, they do not protect against a stolen client ID/secret for a client_credential grant.

OAuth2 also defines the concept of a **refresh token** issued by the Authorization Server and shared with the client app. This refresh token will allow the client app to request a fresh AT (e.g., once it expires) and potentially a refreshed refresh token without having to involve the end-user, for instance. This can be used to maintain a decent UX in a single-page application (SPA) or to allow for offline access when the user is not present anymore, but the client app needs access to the protected resource.

Discovery

In order to help clients dynamically register against an authorization server or programmatically get information about the authorization server features and level of support, some discovery and dynamic registration specifications are also available:

- Client dynamic registration (RFC 7591)[xii]
- Authorization Server Metadata (RFC 8414)[xiii]

Beyond OAuth2

Now that most OAuth2 specifications have been introduced, you can easily imagine how difficult it can sometimes be to navigate through them and ensure one’s implementation is solid and secure. OAuth2 working group members created additional documents to help:

- RFC 6819 - OAuth 2.0 Threat Model and Security Considerations[xiv]
- OAuth 2.0 Security Best Current Practice (currently draft)
- OAuth 2.1 (currently draft) is a minor but important revision to the standard that incorporates security best practices
- RFC 8252 - OAuth 2.0 for Native Apps for best practices around native application clients on different platforms[xv]
- OAuth 2.0 for Browser-Based Apps (currently draft) for best practices around Single Page Applications

OAuth2 is also a foundation upon which other protocols were developed, the most known among these being OpenID Connect.

- OpenID Connect, as described in the specification, is a "simple identity layer on top of the OAuth 2.0 protocol."[xvi] Contrary to OAuth2, which focuses on authorization delegation, OIDC focuses on authentication. It introduces another token (**ID Token**), which is shared between the Authorization Server (or OpenID provider) and the client (or Relying Party). This token is a JWT formatted token. It conveys information about the authenticated identity through standard-defined claims and information about the authentication itself (time of authentication, method used, etc.).
- User-Managed Access 2.0 is another protocol defined on top of OAuth2 (as a new authorization grant).[xvii] It introduces additional tokens, but most importantly, it does introduce a new player in the picture: the **requesting party,** which can be different from the resource owner (in OAuth2, the resource owner is the requesting party).

Additional Reading

For additional information on implementing OAuth2, these resources may be of assistance:

- Richer, Justin, and Antonio Sanso. 2017. *OAuth 2 in Action*. Manning.
- Parecki, Aaron. 2018. *OAUTH 2.0 Simplified*. Lulu.com.

Author

Bertrand Carlier is a senior manager in the Cybersecurity & Digital Trust practice at Wavestone consultancy with 20 years of experience. He has been leading major Identity & Access Management projects, working with many client companies in a variety of industries.

He is devoted to promoting and encouraging the use of open standards and has done so through leading projects and talks at various international conferences.

He likes nothing more than to tackle the newest problems in the Identity and Access Management space: API & microservices security, IAM of Things, AI for IAM and IAM for AI, and, of course, the longstanding problem of "how to cope with both the legacy and the ever more shiny (and accumulating) new toys?"

[i] Dingle, P., (2020) “Introduction to Identity - Part 2: Access Management”, *IDPro Body of Knowledge* 1(2). doi: https://doi.org/10.55621/idpro.45

[ii] Hardt, D., Ed., "The OAuth 2.0 Authorization Framework", RFC 6749, DOI 10.17487/RFC6749, October 2012, <https://www.rfc-editor.org/info/rfc6749>.

[iii] Jones, M. and D. Hardt, "The OAuth 2.0 Authorization Framework: Bearer Token Usage", RFC 6750, DOI 10.17487/RFC6750, October 2012, <https://www.rfc-editor.org/info/rfc6750>.

[iv] Richer, J., Ed., "OAuth 2.0 Token Introspection", RFC 7662, DOI 10.17487/RFC7662, October 2015, <https://www.rfc-editor.org/info/rfc7662>.

[v] Bertocci, V., "JSON Web Token (JWT) Profile for OAuth 2.0 Access Tokens", RFC 9068, DOI 10.17487/RFC9068, October 2021, <https://www.rfc-editor.org/info/rfc9068>.

[vi] Sakimura, N., Ed., Bradley, J., and N. Agarwal, "Proof Key for Code Exchange by OAuth Public Clients", RFC 7636, DOI 10.17487/RFC7636, September 2015, <https://www.rfc-editor.org/info/rfc7636>.

[vii] Jones, M., Bradley, J., and N. Sakimura, "JSON Web Token (JWT)", RFC 7519, DOI 10.17487/RFC7519, May 2015, <https://www.rfc-editor.org/info/rfc7519>.

[viii] Campbell, B., Mortimore, C., and M. Jones, "Security Assertion Markup Language (SAML) 2.0 Profile for OAuth 2.0 Client Authentication and Authorization Grants", RFC 7522, DOI 10.17487/RFC7522, May 2015, <https://www.rfc-editor.org/info/rfc7522> and Jones, M., Campbell, B., and C. Mortimore, "JSON Web Token (JWT) Profile for OAuth 2.0 Client Authentication and Authorization Grants", RFC 7523, DOI 10.17487/RFC7523, May 2015, <https://www.rfc-editor.org/info/rfc7523>.

[ix] Denniss, W., Bradley, J., Jones, M., and H. Tschofenig, "OAuth 2.0 Device Authorization Grant", RFC 8628, DOI 10.17487/RFC8628, August 2019, <https://www.rfc-editor.org/info/rfc8628>.

[x] Jones, M., Nadalin, A., Campbell, B., Ed., Bradley, J., and C. Mortimore, "OAuth 2.0 Token Exchange", RFC 8693, DOI 10.17487/RFC8693, January 2020, <https://www.rfc-editor.org/info/rfc8693>.

[xi] Campbell, B., Bradley, J., Sakimura, N., and T. Lodderstedt, "OAuth 2.0 Mutual-TLS Client Authentication and Certificate-Bound Access Tokens", RFC 8705, DOI 10.17487/RFC8705, February 2020, <https://www.rfc-editor.org/info/rfc8705> and Fett, D., Campbell, B., Bradley, J., Lodderstedt, T., Jones, M., and D. Waite, "OAuth 2.0 Demonstrating Proof of Possession (DPoP)", RFC 9449, DOI 10.17487/RFC9449, September 2023, <https://www.rfc-editor.org/info/rfc9449>.

[xii] Richer, J., Ed., Jones, M., Bradley, J., Machulak, M., and P. Hunt, "OAuth 2.0 Dynamic Client Registration Protocol", RFC 7591, DOI 10.17487/RFC7591, July 2015, <https://www.rfc-editor.org/info/rfc7591>.

[xiii] Jones, M., Sakimura, N., and J. Bradley, "OAuth 2.0 Authorization Server Metadata", RFC 8414, DOI 10.17487/RFC8414, June 2018, <https://www.rfc-editor.org/info/rfc8414>.

[xiv] Lodderstedt, T., Ed., McGloin, M., and P. Hunt, "OAuth 2.0 Threat Model and Security Considerations", RFC 6819, DOI 10.17487/RFC6819, January 2013, <https://www.rfc-editor.org/info/rfc6819>.

[xv] Denniss, W. and J. Bradley, "OAuth 2.0 for Native Apps", BCP 212, RFC 8252, DOI 10.17487/RFC8252, October 2017, <https://www.rfc-editor.org/info/rfc8252>.

[xvi] Sakimura, N., Bradley, J., Jones, M., de Medeiros, B., Mortimore, C. “OpenID Connect Core 1.0 incorporating errata set 1,” OpenID Foundation, November 2014, https://openid.net/specs/openid-connect-core-1_0.html.

[xvii] Maler, E. (ed.), Machulak, M., Richer, J. “User-Managed Access (UMA) 2.0 Grant for OAuth 2.0 Authorization,” Kantara Initiative, January 2018, https://docs.kantarainitiative.org/uma/wg/rec-oauth-uma-grant-2.0.html.

IAM Architecture and Solutions

Introduction to IAM Architecture (v2)

By Andrew Cameron and Graham Williamson

To comment on this article, please visit our GitHub repository and submit an issue.

Table of Contents

Abstract

This article explores several conceptual architectures and how they enable IAM solutions across the enterprise. IAM touches all aspects of an organization's IT environment; whether it's the HR system, email system, phone system, or corporate applications, they all need to interface to the IAM environment. Whether it is by supporting the enforcement of user provisioning rules or validating the access of non-corporate users, IAM will always play a role in making IT operations efficient and secure. An architectural approach will help an organization achieve a consistent and comprehensive IAM solution.

Note: IDPro® does not endorse a particular architecture framework. IAM practitioners will face many different approaches and must adopt the model that best suits their organizations.

Introduction

Identity and Access Management (IAM) touches all aspects of an organization's IT environment. Whether it is the human resources (HR) system, email system, phone system, or corporate applications, each system needs to interface to the IAM environment. IAM will always play a role in making IT operations efficient and secure, by supporting the enforcement of user provisioning rules, as an example, or validating the access of non-corporate users. An architectural approach to developing IAM systems will heighten the organization's probability of achieving a consistent and comprehensive IAM solution.

If the organization maintains an enterprise architecture (EA), any IAM solution they deploy must adhere to the enterprise models and be reflected in the organization's EA artifacts. This article provides a basic approach for IAM professionals to consider whether or not there is an EA in place.

Terminology

- **Access Management**: the use of identity information to provide access control to protected resources such as computer systems, databases, or physical spaces.
- **Architecture**: a framework for the design, deployment, and operation of an information technology infrastructure. It provides a structure whereby an organization can standardize its technology and align its IT infrastructure with digital transformation policy, IT development plans, and business goals.
- **Architecture Overview**: describes the architecture components required for supporting IAM across the enterprise.
- **Architecture Patterns**: identifies the essential patterns that categorize the IT infrastructure architecture in an organization and will guide the deployment choices for IAM solutions.
- **Enterprise Architecture:** an architecture covering all components of the information technology (IT) environment

- **Identity Governance and Administration (IGA):** includes the collection and use of identity information as well as the governance processes that ensure the right person has the right access to the right systems at the right time.

Acronyms

- AP – Application Portfolio
- BPMn – Business Process Mapping notation
- BSA – Business System Architecture
- EA – Enterprise Architecture
- HTTP – HyperText Transfer Protocol
- IA – Information Architecture
- IAM – Identity and Access Management
- IDaaS – Identity-as-a-Service
- IGA – Identity Governance and Administration
- JSON – file structure for the communication of data attributes
- MFA – Multi-factor Authentication
- PABX – Private Automatic Branch Exchange
- PAP – Policy Administration Point
- PDP – Policy Decision Point
- PEP – Policy Enforcement Point
- PIP – Policy Information Point
- RBAC – Role-based Access Control
- RESTful API - architecture for a programming interface defining how HTTP methods are to be used
- SAML – Security Assertion Markup Language
- SCIM – System for Cross-domain Identity Management
- SSO – Single Sign-On
- TA – Technical Architecure
- XML – eXtensible Markup Language - a file structure for the communication of data attributes

IAM Architecture Overview

IAM professionals must have a vision for the IAM environment that satisfies corporate requirements. Each IAM project must build towards the desired target state. An architectural approach will enable the IAM professional to plan, design, and deploy IAM solutions that are both coordinated and integrated; and combine to form a comprehensive IAM environment that meets corporate stakeholders' current and projected needs.

Identity management within an enterprise touches virtually all systems in use within the organization. Systems, in this context, comprise computer systems that staff and business partners use in the performance of their job responsibilities and physical access systems, such as a requirement to show an identity pass to gain access to a restricted area. Staff includes contractors; they are typically managed through a different system (many HR systems only accommodate employees) but need access to many of the same corporate systems as employees. Including non-human accounts should also be considered; most organizations have service accounts for machine access to systems. As more automation is incorporated into company operations, access control for sensors or bots should be incorporated in the IAM environment. Including non-human entities in the architecture allows the enterprise to manage their access control in a manner consistent with all other accounts; IAM professionals should consider these entities should during the system development planning process.

It is the task of an IAM practitioner to ensure that, wherever and whenever identity information is used within an enterprise, the information is collected and used in a properly designed environment that ensures efficiency, protects privacy, and safeguards integrity. Applying an architectural approach, i.e., developing project requirements within a structured framework, will significantly raise the likelihood that an IAM project will be completed consistently and comprehensively with a controlled impact on stakeholders.

There are four levels that the IAM practitioner should consider when developing a solution architecture:

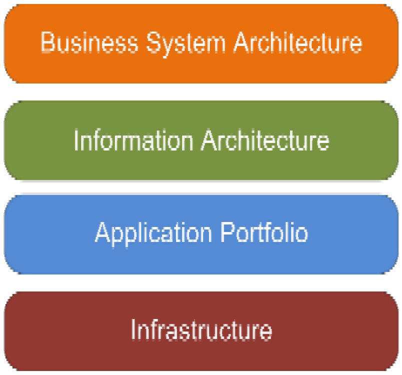


Figure 1: Generic Enterprise Architecture Framework

Business System Architecture (BSA)

Mapping business processes for the collection, usage, and eventual deletion of identity data will greatly assist in understanding the breadth of the IAM task. While BPMn is typically used for business process mapping, the IAM practitioner should adopt whatever tool is typically used in their company.

Considering IT architecture at the business level will facilitate a more holistic approach that considers the identity requirements of all connected systems and ensures consistency in naming conventions. It will also reduce the probability of an IAM project running over budget or over time (a common occurrence when a system owner who has not previously been consulted hears about an IAM project and adds unanticipated requirements).

Information Architecture

It is important to map the identity data elements required by the various applications to the IAM collection, management, and governance systems. This mapping will ensure no application is 'left behind' when the IAM systems are re-developed. A useful tool is an 'entity-relationship diagram' that maps each attribute collected to each system that requires it. The Information Architecture (IA) should drive consistency between connected systems (e.g., should Firstname, Middle Initial, and Lastname be used, or should Common name, Lastname be used). It should also help define roles (e.g., is this role for a Payroll Clerk or a Financial Officer). The IA should nominate attribute authority (e.g., which system is the authority for phone numbers). Best practice is for the IAM system to be the 'source of truth' for identity information in the company (sometimes called the 'book of record') because it is typically bad practice for source systems (HR, PABX, etc.) to be queried for data attribute lookups.

The IA becomes the vehicle for 'identity data orchestration.' It is the master plan for the collection and use of identity data within an enterprise.

Application Portfolio

An inventory of applications to be included in the IAM project should be conducted.[i] How current are they? Are any of the included applications under development? Will the IAM project materially change how each application interacts with the IAM environment? For instance, if an API gateway is being deployed for access to IAM attributes, any application redevelopment should migrate from existing authentication mechanisms to the gateway operation.

A company's Application Portfolio (AP) becomes an inventory of corporate applications. The record for each application should identify the system owner, type of application (web app, client-server, mainframe, etc.), and its reliance on the IAM environment. Some applications will expect the IAM system to pass authenticated sessions to it. In contrast, others will require user attributes so that it can determine the authorization that a user has to application functionality. The AP should identify the level of integration between each relying application and the IAM system. Web applications will likely pass user requests and responses via HTTP headers. In other scenarios, client-server applications may use an API, while cloud applications may use a SAML request or, if it maintains its own data repository, the SCIM protocol.[ii]

The AP becomes an important record for an organization because it facilitates the planning required as applications are updated.

Technical Architecture

The Technical Architecture (TA) describes, among other things, the technical environment to be supported by the IAM environment. This description will involve understanding the patterns used within the company. Most organizations will have “n-tier” web services and hybrid cloud patterns, but there might still be client-server patterns and potentially mainframe hub-and-spoke patterns. Each additional pattern to be supported will increase the complexity and cost of the project. Often IAM environments with older infrastructure leave out support for legacy technology due to cost considerations, but this fragments the IAM task. Properly constituted, a cost/benefit analysis for deploying legacy connectors will typically be successful.

The TA impacts the IAM environment because different solutions are required for different patterns. For example, a web services pattern will mandate a single sign-on (SSO) environment capable of supporting RESTful APIs and SAML assertions and passing identity attributes in JSON arrays or XML files. An on-premise Windows environment, as another example, will typically use the Kerberos authentication protocol from an AD infrastructure or an LDAP directory. A cloud environment will often require a SAML operation or an Identity-as-a-Service (IDaaS) offering, whereas an older directory should be supported via a connector from the IAM infrastructure.

Additionally, corporate security policy may create requirements that require certain technical decisions. For instance, a requirement to maintain full control and authority over the data and infrastructure may require hosting the entire identity management stack on premises.

Architectural Approach

It is an unfortunate fact that many IAM (identity and access management) projects exceed their scheduled time and budget. The usual reason for this is a misunderstanding of the extent of the project and the systems impacted. The project team tends to focus just on the task at hand, e.g., installing the IAM software package, without realizing that IAM systems within an enterprise touch virtually all other systems in use within the organization. These other systems might include a birthright system such as email, an administrative system such as the Financial Management system, or an operational system such as an Enterprise Resource Management system.

In some circumstances, the change caused by an IAM project will be minimal, with a limited impact on resources. In other cases, the change will be significant, impacting both infrastructure and personnel across the organization. An architectural approach will ensure that a solution architecture is developed for each IAM project to understand the extent of the work required and effectively plan for the change it will generate.

An IAM practitioner's task is to ensure that, wherever and whenever identity information is used within an enterprise, the information is collected and used in a properly designed environment that ensures efficiency, protects privacy, and safeguards integrity.

For organizations with an EA, understanding how information is collected and used should be quite easy, as it is fundamentally a part of how the systems are deployed. For other organizations, the environment will be a “greenfield,” allowing the IAM practitioner to develop their own architectural approach.

Architecture Patterns

At the Technical Architecture level, a “pattern” approach is useful to understand the supported technology within an organization. For instance: what is the predominant server infrastructure – is it Linux or Windows or both? What server operating system versions are supported? Are VMs used? What is the support for cloud infrastructure – public, private, hybrid? Is AWS, Azure, or Google Cloud supported? Can the scale required for customer IAM be accommodated? For IoT devices – how does the IoT platform integrate with the corporate environment?

The TA will define the computer system “patterns” to be supported by the IAM environment within an organization. For young companies, this will be web-based patterns, either “2-tier” or “n-tier.” Increasingly managed cloud environments are being engaged, potentially with a micro-services approach. But for mature organizations, there will typically be legacy applications with a client-server pattern, or even a mainframe ‘hub and spoke’ pattern, with PCs running terminal emulator software.

The IAM environment must support the selected patterns and ensure a managed approach that adheres to the organization’s governance and cybersecurity policy.

Host

There are few mainframe systems left in service, with notable exceptions in the banking industry and some government installations. The IAM environment will often be required to synchronize to an older data store to support a mainframe system.

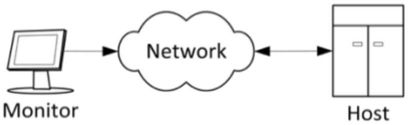


Figure 2: Mainframe application accessed from a monitor

Client-Server

Client-server environments can present a complex support requirement since many such systems maintain their own identity database to provide fine-grained access control to system functionality. Redevelopment of a client-server application to externalize access control decisions to an authentic authorization server can be a way to harmonize access policies across an organization.

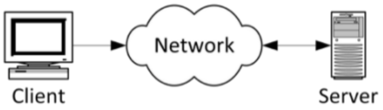


Figure 3: Client application access a backend server

N-tier

The most common on-premise application environment these days is an "n-tier" web services infrastructure. While there are many variants, a user accessing the front-end web server will be redirected to an authentication service, usually supporting SSO, with an authentication token passed back to the application in an HTTP header. If the application requires user authentication, the IAM system should set user entitlements as part of the initial provisioning activity when a user joins the organization.

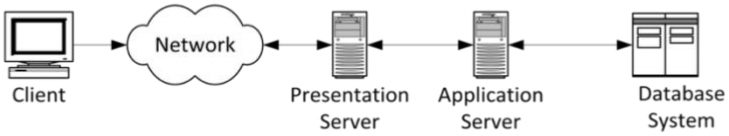


Figure 4: Common web-services model

Hub & Spoke

Hub and spoke systems are typically only in large transaction processing systems. Often the only IAM touchpoint is access control for DevOps staff via a privileged access management system.

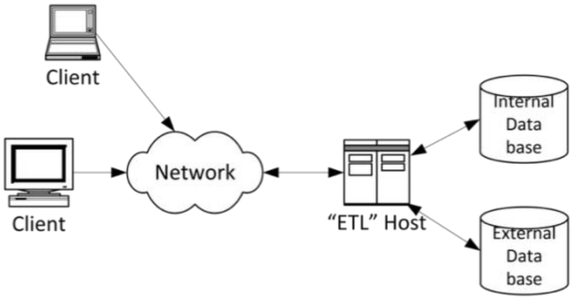


Figure 5: Common data service configuration

Remote Access

Increasingly remote access to corporate systems must be supported. The authentication server must accommodate the required access control mechanisms, from basic LDAP lookups for password accounts to sophisticated MFA environments capable of elevating authentication levels to suit application security requirements. The provisioning task in such environments requires maintaining one or more identity provider services within the enterprise.

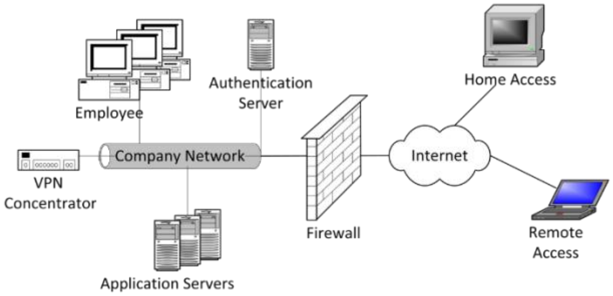


Figure 6: Typical enterprise network access model

Hybrid Cloud Identity

A key indicator of effectiveness in an IAM Architecture is how complexity is managed across the IAM components in the environment. Today, most organizations are leveraging cloud infrastructure platforms in some capacity, either private clouds provided by their technology partners or public clouds such as AWS, Azure, or Google. This raises the issue of how to establish identity as a common control plane between the on-premises environment and the cloud infrastructure. IAM is a critical component of a hybrid IT architecture. Hybrid IAM allows organizations to establish a common credential that can be enabled for access to resources in either on-premises or cloud environments.

The hybrid cloud example assumes an existing 'source of truth' to which all enterprise users authenticate; this is typically Active Directory. With the Hybrid IAM pattern, authenticated on-premise users will have access to on-premise, public cloud, or other external services that support common identity standards such as OpenID Connect or OAuth.

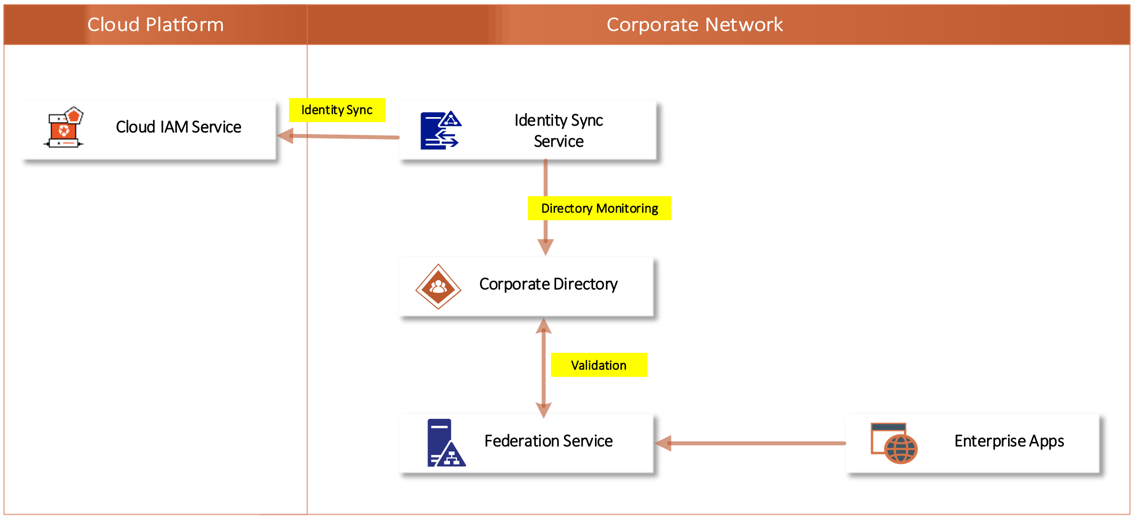


Figure 7: Hybrid Cloud Identity Architecture model

Table 1: Hybrid IAM Architecture components

Component	Description
On-Premise (Corporate) Directory	Directory service that enables authentication to access enterprise resources (e.g., Active Directory). Typically contains directory objects (accounts) that represent a human (user account) or non-human identity (service account).
On-Premise Federation Service	Identity service that implements common access management capabilities (authentication and authorization) for enterprise applications. Typically supports identity standards like SAML or OpenID Connect to enable access to internal or external resources.
Identity Sync Service	Infrastructure service that monitors directory objects in the enterprise directory for changes and synchronizes changes to a mapped cloud directory object. Sync direction can be one-way or two-way but is typically implemented in an Enterprise to Cloud direction to minimize risk and complexity. Standards such as SCIM can be used for this data transfer.
Cloud IAM Service	Platform service in a public cloud that implements core IAM capabilities (Authentication, Federation, Access Management) and can be leveraged to access on-premise resources as well.

Important considerations for Hybrid IAM:

- User Provisioning: User objects can be configured to synchronize when added to either the cloud or the on-premises environment.   The best practice is to restrict user provisioning to one environment and sync account and profile data to the other environment (typically from enterprise to the cloud).

- Profile Data: Manually maintaining identities in more than one environment can add unnecessary complexity and risk to your security posture. Cloud identity objects may not need the entire set of user profile data available for an on-premises user; the IAM practitioner should take care (e.g., understand the business requirements for authentication) when deciding how much user profile data should be stored on a cloud user object. A principle of "least privilege" should be applied to avoid data spillage.
- Single Sign-On: Cloud IAM environments can enable SSO to on-premises applications or services. For SSO to be successful, the user object must have been provisioned and enabled for sign-in. It is critical to understand the authentication scenarios available from the cloud IAM platform (e.g., pass-through authentication or federation) and ensure that there is a fit with the enterprise requirements.

As enterprises place increasing importance on "time to value", a hybrid IAM architecture will be critical to support infrastructure expansion beyond the enterprise perimeter and leverage cloud-enabled benefits (e.g., agility, scalability, reliability). The IAM professional will find use-cases where IDaaS solutions offer rapid deployment and appealing software update methods, when compared as an alternative to on-premises solutions. However, hybrid scenarios may require both types of deployments, cloud and on-premise, to working together. In some cases, the cloud identity service will be the 'source of truth' for identity data within the organization. Such an IDaaS approach can reduce the overhead of managing on-premise infrastructure for an enterprise, an activity that can be costly and inflexible.

Applying an Architectural Approach

An architectural approach can be taken to an IAM project regardless of whether it is in the collection and management of identity information or access management, using identity information for access control to protected resources.

Identity Governance and Administration

Identity Governance and Administration (IGA) covers the identity management side of IAM, e.g., the 'admin-time' events that establish user entitlements, as opposed to 'real-time' events that occur when users request access to protected resources. IGA combines administration and governance over the collection, use, and disposal of identity information. It requires a governance facility that enables managers to certify the entitlements that their staff have been granted. In addition, IGA typically includes monitoring and reporting functions for identity services that, in turn, support corporate requirements.

IGA systems support:

- Administering accounts and credentials
- Identity and account provisioning
- Managing entitlements
- Segregation of duties
- Role management
- Analytics and reporting

IGA systems provide additional functionality beyond standard IAM systems. In particular, they help organizations meet compliance requirements and enable them to audit access for compliance reporting. They also automate workflows for tasks such as access approvals and provisioning/deprovisioning.

Identity Lifecycle

The business rules that tie these elements together are generally referred to as the identity lifecycle.[iii] In the identity lifecycle, an identity is created that defines who or what (human or non-human) needs access to a protected resource. Every stage of the identity lifecycle sees the activities of the identity managed to ensure business rules are enforced according to the identity and security rules of the enterprise.

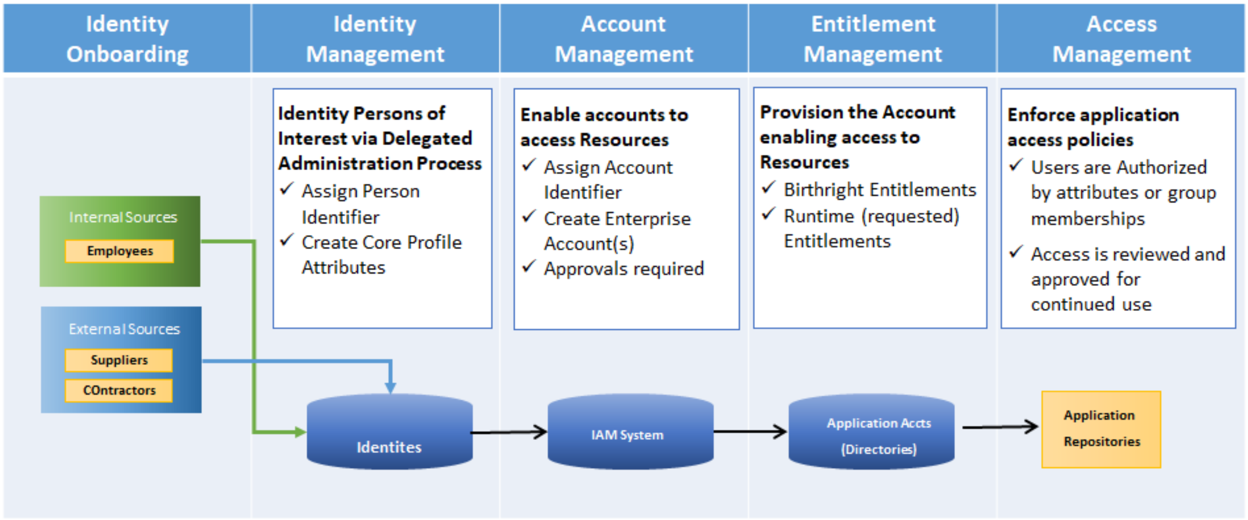

Identity Onboarding
Identity Management
Account Management
Entitlement Management
Access Management
Identity Persons of Interest via Delegated Administration Process
✓ Assign Person Identifier
✓ Create Core Profile Attributes
Enable accounts to access Resources
✓ Assign Account Identifier
✓ Create Enterprise Account(s)
✓ Approvals required
Provision the Account enabling access to Resources
✓ Birthright Entitlements
✓ Runtime (requested) Entitlements
Enforce application access policies
✓ Users are Authorized by attributes or group memberships
✓ Access is reviewed and approved for continued use
Internal Sources
Employees
External Sources
Suppliers
COntractors
Identites
IAM System
Application Accts (Directories)
Application Repositories

Figure 8: Identity Lifecycle Categories

IGA System Components

Identity governance and administration tools help facilitate identity lifecycle management. IGA systems generally include the following components for identity administration:

- **Password management**: using tools like password vaults or, more often, SSO, IGAs ensure users don't have to remember many different passwords to access applications.

- **Integration connectors**: used to integrate with directories and other systems that contain information about users and the applications and systems they have access to, as well as their authorization in those systems.
- **Access request approval workflows**: support the automation of a user's request for access to applications and systems and ensures all access is properly authorized.
- **Automated de-provisioning**: supports the removal of a user's entitlement to access an application when the user is no longer authorized to access a system.
- **Attestation reporting**: used to periodically verify user entitlements in various applications (such as add, edit, view, or delete data) and is usually sent to a user's manager.
- **Recertification of user entitlements**: often a response to an attestation report, recertification of user entitlements involves recording a manager's approval of their staff's system access. If access is no longer required, this shifts to automatic de-provisioning.
- **Segregation of duties**: rules that prevent risky sets of access from being granted to a person. For example, if a person has the ability to both view a corporate bank account and transfer funds to outside accounts, this might enable a user to transfer money to a personal account.
- **Access reviews**: reviews include tools that streamline the review and verification (or revocation) of a user's access to different apps and resources. Some IGA tools also provide discovery features that help identify entitlements that have been granted.
- **Role-based management**: also known as Role-based Access Control (RBAC), this includes defining and managing access through user roles.
- **Analytics and reporting**: include tools that log activities, generate reports (including for compliance), and provide analytics to identify issues and optimizations.

IGA Solution Architecture

An example of how an IGA solution could support an authentication service is shown in Figure 9 (access management shown for context):

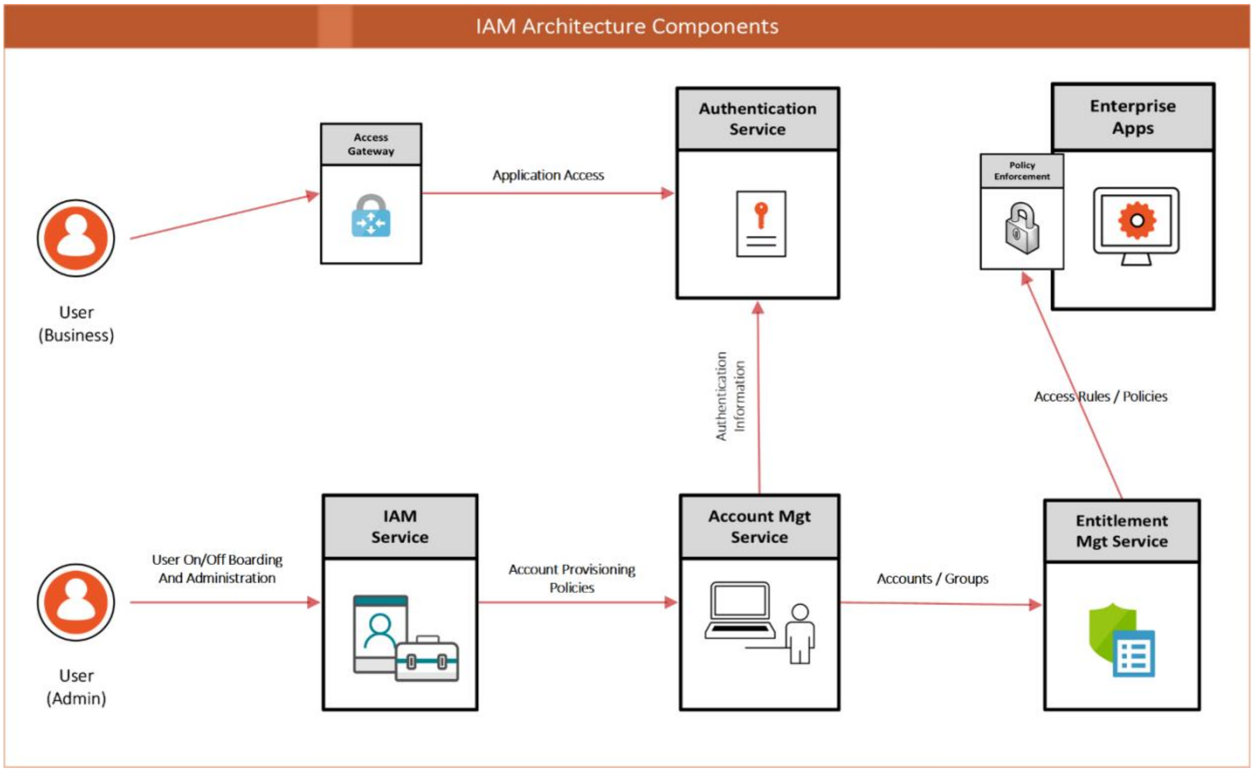


Figure 9: IAM Architecture Components

This architecture supports the following IAM Processes:

Table 2: IAM Processes

Process	Description
Identity Provisioning	Creates identity records based on initiation from trusted identity sources (e.g., the HR System)
Account Provisioning	Creates accounts in Enterprise Directories based on birthright provisioning rules. Also supports the creation of application accounts based on request / approval workflows.
Entitlement Management	Supports the workflow and administration requirements of enabling user-to-group/role mappings that enable access management rule creation.

Access Management

Access Management is the ‘real-time’ component of IAM. It encompasses the processes that are critical in protecting corporate resources and securing the digital business. Whether it is giving access to customers to enable e-commerce or securing resources for partners to conduct business securely, the Access Management architecture will control the planning, design, and development of the enabling technology.

Access Management Overview

An access management architecture will have components that enable only those accounts that are authorized to perform an action on a protected enterprise resource.

The key functions supported in an Access Management Architecture are:

- User Authentication (staff, contractors, business partners)
- Access Policy Management
- Access Policy Decision making and enforcement
- Authorization Control (Coarse / Fine-Grained)
- Adaptive Access controls
- Single Sign-On (SSO)
- Authenticated Session Management
- Security Token Services
- Access Event Logging
- User Behavior Analytics
- Access Management Solution Architecture

The two most common Access Management services supported in most scenarios are:

- Authentication – logging into a computer system - typically role-based
- Authorization – accessing computer system functionality – typically attribute-based
- Policy-based authorization is increasingly being deployed. It provides access control to corporate resources in accordance with centrally managed corporate policy rather than entitlements established on a system-by-system basis.

An example of a fine-grained authorization environment is shown in Figure 10. The components of the solution combine to control access to corporate resources based on the policies in the Decision Point.

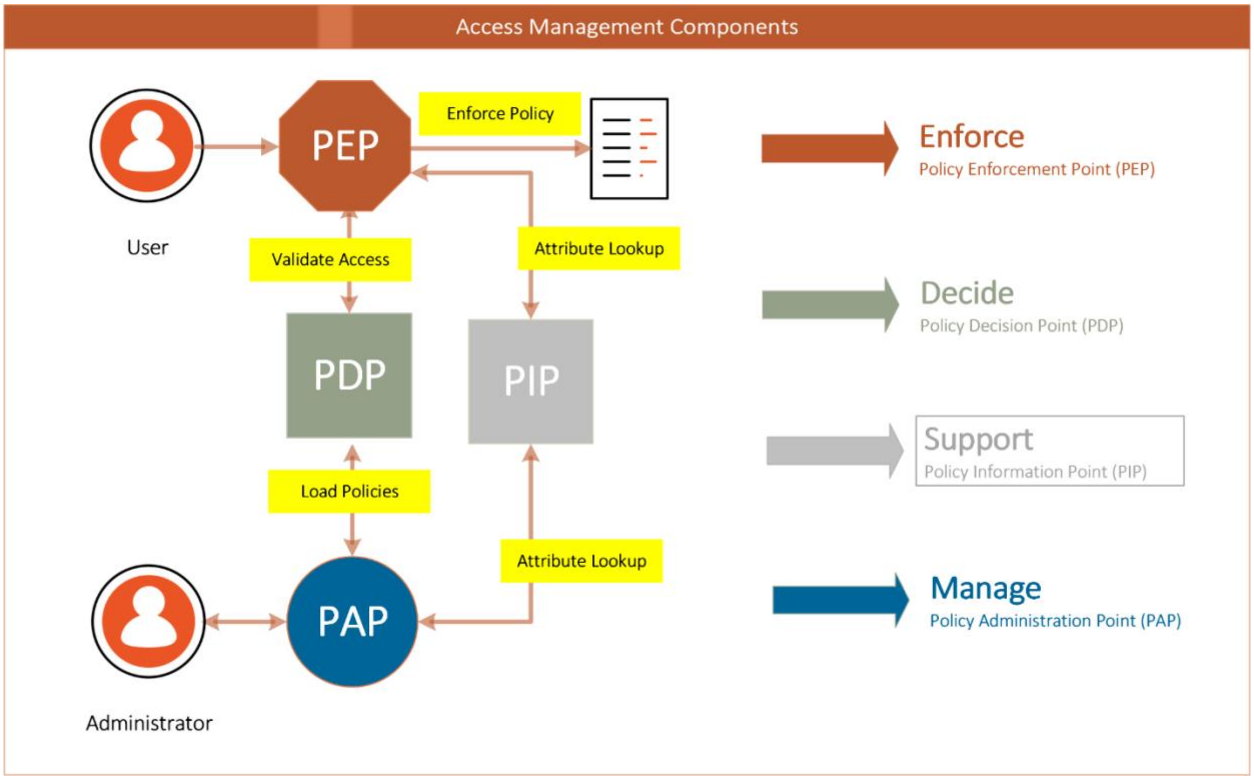

Access Management Components
User
PEP
Enforce Policy
Validate Access
Attribute Lookup
PDP
PIP
Load Policies
Attribute Lookup
PAP
Administrator
Enforce
Policy Enforcement Point (PEP)
Decide
Policy Decision Point (PDP)
Support
Policy Information Point (PIP)
Manage
Policy Administration Point (PAP)

Figure 10: Typical Components of an Authorization Service

The architecture of an authorization service will typically contain the key elements that are involved in the flow from an actor (person or system) on a device (mobile or desktop) that accesses an application or service (typically over the internet) that resides within an enterprise boundary (behind network firewalls).

Table 3: Policy Control Points

Policy Control Point	Definition
Policy Administration Point (PAP)	responsible for creating policy statements that tie the user to a role or group and defines the type of access to a resource
Policy Enforcement Point (PEP)	responsible for protecting the resource; intercepts traffic to the resource, and validates access with the PDP
Policy Decision Point (PDP)	determines access to a resource, uses policy to determine if a subject (user) has access to a resource, usually via an attribute value or role or group membership.
Policy Information Point (PIP)	typically a user or attribute store that provide information about managed users (e.g., Active Directory or LDAP directory)

Access Management Patterns

A well-crafted IAM architecture is able to both improve user experience and increase security by combining the flow between architecture components in a connected, orchestrated framework. Historically, organizations have seen security and ease of use as tradeoffs, but with the new identity technologies available today, it is possible to have both. When combining these key components in a deployment blueprint (solution configuration), an architecture pattern evolves to support most, if not all, access management needs across the organization.

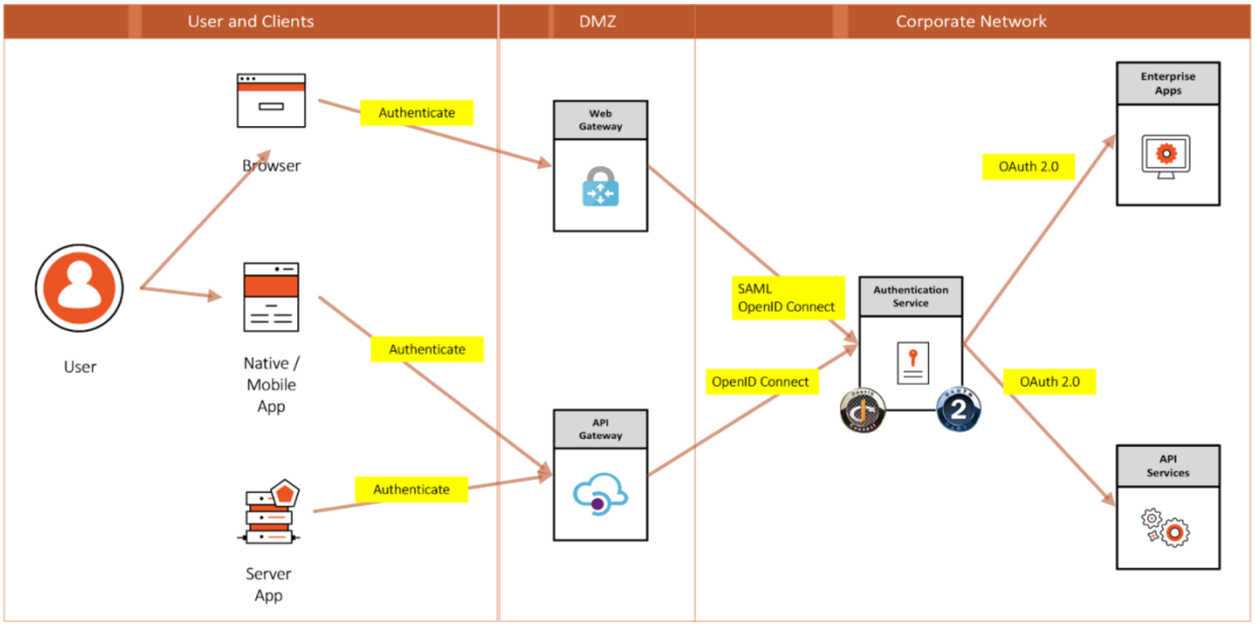


Figure 11: Access Management Patterns

Table 4: Access Management Pattern descriptions

Pattern	Description
Browser to Web Application	A user needs to sign in to a web application that is secured by an Authentication Service
Native App (also Single Page App) to Web API	A native application needs to authenticate a user to access resources from a web API that is secured by an Authentication Service
Server App to Web API	A server application with no web user interface needs to get resources from a web API secured by an Authentication Service

Identity Standards

No IAM solution architecture is complete without addressing the applicable standards. Because IAM touches virtually all corporate systems, interfaces need to adhere to standards in order to minimize the amount of customization that would otherwise be required. An IAM Architecture should support a “pluggable” approach that facilitates interconnection and ties together key security enablers that are built on industry standards. There are several industry organizations (standards bodies) like IETF, OASIS, Kantara Initiative, and the OpenID Foundation.

The key standards that support modern identity and access management today are OIDC, OAuth2, and SAML.[iv]

OpenID Connect (OIDC) 1.0	OAuth 2.0	SAML 2.0
OpenID Connect	OAUTH 2	SAML

Figure 12: Logos for OIDC, OAuth2 , SAML

Conclusion

IAM practitioners should adopt the enterprise architecture approach used within the organization in which they are working. In the absence of a corporate approach to architecture, IAM practitioners should develop an architectural approach that ensures their IAM projects consider all the business systems that might be affected, the types of applications to be supported, and the infrastructure on which IAM solutions are to be deployed.

An IAM project that takes such an approach will have a significantly better chance of being completed within schedule and budget constraints. It will also be much more likely to satisfy users.

Authors

Andrew Cameron

Andrew Cameron is the Enterprise Architect for Identity and Access Management at General Motors. His responsibilities include Defining the Strategy and Implementation Roadmaps of their IAM technology platform and ensuring the architectural quality of the many initiatives driving the GM digital business.

Graham Williamson

Graham Williamson is an IAM consultant working with commercial and government organizations for over 20 years with expertise in identity management and access control, enterprise architecture and service-oriented architecture, electronic commerce, and public key infrastructure, as well as ICT strategy development and project management. Graham has undertaken major projects for commercial organizations such as Cathay Pacific in Hong Kong and Sensis in Melbourne, academic institutions in Australia such as Monash University and Griffith University, and government agencies such as Queensland Government CIO's office and the Northern Territory Government in Australia and the Ministry of Home Affairs in Singapore.

Change Log

Date	Change
2020-06-17	V1 published
2021-09-30	Additional information added regarding hybrid cloud infrastructures; removed specific mention of RACF; minor editorial updates

[i] Readers may find the IDPro BoK article "Introduction to Project Management for IAM Projects" of interest. See Williamson, Graham, and Corey Scholefield, "Introduction to Project Management for IAM Projects," IDPro Body of Knowledge, vol 1, issue 1, March 2020, https://bok.idpro.org/article/id/25/.

[ii] "SCIM: System for Cross-domain Identity Management" http://www.simplecloud.info/

[iii] Cameron, Andrew and Olaf Grew, "An Overview of the Digital Identity Lifecycle," IDPro Body of Knowledge, 30 October 2021, https://bok.idpro.org/article/id/31/.

[iv] OpenID Connect, website, OpenID Foundation, https://openid.net/connect/; OAuth2, website, https://oauth.net/2/ ; "Security Assertion Markup Language (SAML) V2.0 Technical Overview," OASIS, http://docs.oasis-open.org/security/saml/Post2.0/sstc-saml-tech-overview-2.0.html

IAM Reference Architecture (v2)

By George B. Dobbs

To comment on this article, please visit our GitHub repository and submit an issue.

Table of Contents

Abstract

This article provides a reference model to organize the presentation of technical details associated with various implementations of identity and access management (IAM) architectural concepts. The model is conceptual, as are the set of abstract components which it provides.

Additional articles will be made available in the IDPro Body of Knowledge that offer more specific technical use-cases based on the abstract concepts in this document.

Introduction

It has been said that all models are wrong, but some are useful.[i] This model attempts to find a level of generality that is broadly useful. Too general, and the model becomes untethered to reality and definitely not useful. Too specific, and the model will only work in some cases.

This Identity and Access Management (IAM) Reference Architecture leans more towards technical implementation and touches on some of the process, legal, and capability dimensions. This breadth of coverage is intended to give the reader a set of concepts that can be applied when thinking about IAM.

The principle behind this model assumes that the management of identities and access can (mostly) be separated from their use. This concept can apply to distributed systems as well as self-contained systems. So, when you see IAM working together with, say, an application, it may mean that these are separate physical systems. Alternatively, it could mean these parts are separate pieces of software running on a single system.

The main goal of this article is to allow consistent discussion of more specific use-cases by offering a common set of terms and concepts to be used across all IAM architectures.

While the model incorporates guidance from various standards and best practice documents, the primary structure for the model started with the ISO/IEC framing.[ii] The Unified Modeling Language (UML) detail was removed for simplicity, and the IAM model has been extended so that authorization, governance, and risk-control can be included.

Some of the ISO/IEC names have been changed to reflect more common usage. In some cases, the ISO names have been used in a more expansive way than their original definition.

In an attempt to adopt the most useful terminology, the model has been reviewed in conjunction with the FICAM,[iii] Internet2,[iv] NIST SP-800-63 definitions,[v] NIST Zero Trust frameworks,[vi] and with the Identity Stack as presented at Identiverse 2019.[vii]

The model can be used to support varying levels of system complexity. For example:

- in a Distributed System environment, where the architecture may have a web-hosted application the Relying Party (RP) that depends on a cloud identity service, the Identity Provider (IDP). The RP, in this case, could be a customer-facing application or a workforce-facing application;
- in a Single System model, where a computer's file system (the RP) provides access control based on the user information acquired at login (the IDP). In this case, both the file system and IAM function are encapsulated in an operating system.

Terminology

The terms are defined below. Those with a ✓ mark are the abstract components that comprise the model.

Two of the terms, IDM and Access Management, are used for a conceptual grouping of components. This is to aid understanding.

Item	Definition
Access Control	Various methods to limit access to data, systems, services, resources, locations by a user, a device or thing, or a service.
✓Access Governance (also known as Identity Governance and Administration (IGA))	Access Governance provides oversight and control over access rights implemented in multiple local or shared authorization systems. These rights may be controlled in a variety of ways, starting with the existence and validity of the digital identity. Other controls include various mechanisms such as policies, the mapping of roles, permissions, and identities. The abbreviation used is for Identity Governance and Administration and is commonly used in the commercial sector. This roughly corresponds to the Access Certification section of the first-class component Governance Systems in the FICAM model. IGA is not specifically addressed in the ISO/IEC model.
✓Access Management	The process and techniques used to control access to resources. This capability works together with identity management and the Relying Party to achieve this goal. The model shows access management as a conceptual grouping consisting of the Access

Governance function and the shared authorization component. However, access management impacts local authorization as well (through the governance function).

Assertion

A formal message or token that conveys information about a principal, typically including a level of assurance about an authentication event and sometimes additional attribute information. Sometimes this is called a Security Token.

Assurance Level

A category describing the strength of the identity proofing process and/or the authentication process. See NIST SP.800-63-3 for further information.

✓Attribute Provider

Sometimes the authority for attributes is distinguished from the authority for identities. In this case, the term Attribute Provider is sometimes used. It is a subset or type of an Identity Information Authority.

✓Audit Repository

A component that stores records about all sorts of events that may be useful later to determine if operations are according to policy, support forensic investigations, and allow for pattern analysis. Typically, this is highly controlled to prevent tampering. Audit Repository is the ISO name for this concept and is localized to the IDM. In this model, the term is generalized to indicate a service that supports event records from any part of the ecosystem.

✓Authentication (AuthN)

The act of determining that to a level of assurance, the principal/subject is authentic.

AuthN Assertion

A security token whereby the IDP provides identity and authentication information securely to the RP.

Authorization (AuthZ)

Authorization is how a decision is made at run-time to allow access to a resource. We break this down into two types: shared and local. The FICAM framework includes this as a subcomponent of the Access Management System. AuthZ is not included in the ISO or Internet2 models.

✓Shared AuthZ

Shared authorization is provided by a facility outside of the RP. It is shown here as part of the access management grouping.

✓Local AuthZ

Local authorization is handled by the RP.

Credential

A credential allows for authentication of an entity by binding an identity to an authenticator.

✓Credential Service Provider (CSP)

Following the guidance included in NIST 800-63-3, we include both the enrollment function and credential services together under the name Credential Services Provider.

Credential Services

Credential Services issue or register the subscriber authenticators, deliver the credential for use, and subsequently manage the credentials. We include PKI information for IAM architectures that must include system components that need certificates and private keys. This roughly corresponds to the FICAM component called Credential Management Systems.

Enforcement

The mechanism that ensures an individual cannot perform an action or access a system when prohibited by policy.

Enrollment

Also known as Registration. Enrollment is concerned with the proofing and lifecycle aspects of the principal (or subject). The entity that performs enrollment has sometimes been known as a Registration Authority, but we (following NIST SP.800-63-3) will use the term Credential Service Provider.

Entitlement

The artifact that allows access to a resource by a principal. This artifact is also known as a privilege, access right, permission, or an authorization. An entitlement can be implemented in a variety of ways.

✓Identity Information Authority (IIA)

This represents one or more data sources used by the IDM as the basis for the master set of principal/subject identity records. Each IIA may supply a subset of records and a subset of attributes. Sometimes the IIA is distinguished from the Identity Information Provider or IIP. We use IIA to include the service that actually provides the information as well as the root authority. This corresponds to Identity Information Source in ISO/IEC 24760-2 and Identity Sources in Internet2.

✓Identity Management (IDM)

A set of policies, procedures, technology, and other resources for maintaining identity information. The IDM contains information about principals/subjects, including credentials. It also includes other data such as metadata to enable interoperability with other components. The IDM is shown with a dotted line to indicate that it is a conceptual grouping of components, not a full-fledged system in itself.

Identity Provider (IDP)

Identity Provider or IDP is a common term. We treat this as a subset of Identity Management. It consists of the service interfaces: AuthN/Assertion, Service Provisioning Agent, Session Management, Discovery Services, and Metadata Management.

✓Identity Register

This is the datastore that contains the enrolled entities and their attributes, including credentials. See the IDM section for elaboration. The terms Directory, Identity Repository, and Attribute Store are sometimes used as synonyms.

✓Metadata Management

The processes and techniques that allow the collection, use, and eventual deletion of control data used by the IDM to recognize and trust the Relying Party. This corresponds to Relying Party data in the Internet2 model.

✓Relying Party (RP)

A component, system, or application that uses the IDP to identify its users. The RP has its own resources and logic. Note that the term ‘relying service’ is used in the ISO/IEC standards to encompass all types of components that use identity services, including systems, sub-systems, and applications, independent of the domain or operator. We will use the more common Relying Party (or RP). An RP roughly corresponds to the Agency Endpoint in the FICAM model or to Identity Consumers in the Internet2 model.

✓Risk Context (RCTX)

Risk Context consists of additional facts that can be brought to bear to improve the overall security of the ecosystem. Internal or external events and facts can be applied to enable, limit, or terminate access. This is similar to the section Monitors and Sensors under FICAM’s Governance Systems and to many of the inputs of the Policy Decision Point in the NIST Special Publication 800-207, a paper on Zero Trust.

Session

A period of time after an authentication event when an RP grants access to resources for the principal/subject. The duration of the session and the mechanism for enforcement vary by implementation.

✓Session Management

A coordinating function provided by an IDP to control sessions of subscribing RPs.

Trust Framework

This component represents the legal, organizational, and technical apparatus that enables trust between the IDM and the RPs.

✓Trust Root

A technical structure that provides the IDP and RP the ability to recognize each other with a high degree of certainty. This is similar to the concept of Trust Anchor (NIST SP.800-63-3), but we allow for a structure that relies on a mutually agreed-upon third party. A trust root derives from the operation of a Trust Framework.

Basic Structure of the Model

The most basic function of the identity system is to provide secure storage of the information about identities and a way for Relying Parties (RPs) to use that data to control access to resources. The following diagram shows the core components of an identity management system (IDM) that supports multiple RPs.

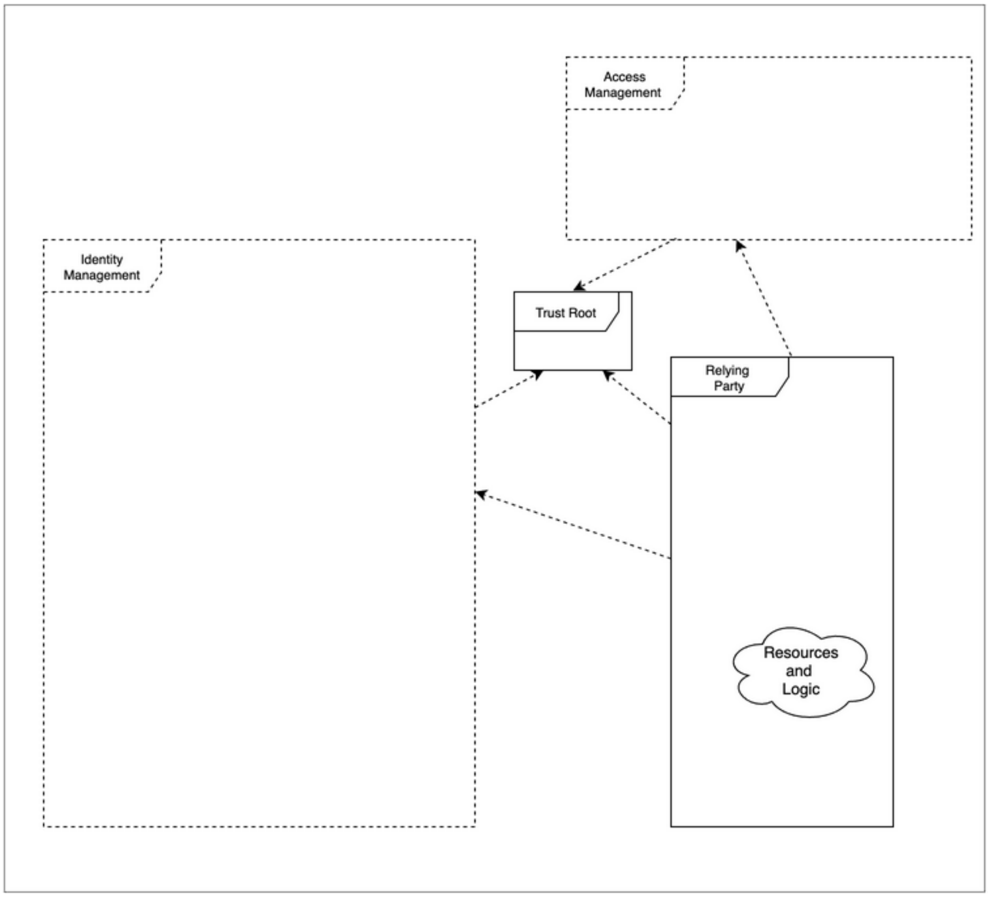


Figure 1: Basic Component Dependencies the IDM supports multiple relying parties. The core components of the IDM are shown. The dotted arrowed lines show dependencies

Identity Management

Identity Management (IDM) is a set of policies, procedures, technology, and other resources for maintaining identity information. In this model, it contains information about principals/subjects, including credentials. It also includes other data such as metadata to enable interoperability with other components. The IDM is shown with a dotted line to indicate that it is a conceptual grouping of components, not a full-fledged system in itself.

Relying Party

The Relying Party (RP) is a component, system, or application that uses the IDM to identify its users. The RP has its own resources and logic. It comes in many forms, all of which use identity services, including systems, sub-systems, and applications, independent of the domain or operator.

Trust Framework

This component represents the legal, organizational, and technical apparatus that enables trust between the IDM and the RPs.

When the IDM and the RP are not in the same organization, the Trust Framework takes on a salient aspect, resulting in multilateral or bilateral agreements. In simple cases, this may be a contract between two parties. In other cases, there may be a multilateral agreement. We will use the term federation loosely to cover both cases. These frameworks are described further in Laws Governing Identity Systems (v2).[viii]

These agreements, rules, and policies govern how the federation members operate and interact.[ix] The parties of a federation establish mutual agreement upon an acceptable identity to be used between the parties in a federated relationship (for instance, the level of assurance used) in order to operate well. In addition, the definition and values of attributes of federated identities should be agreed upon. The parties should agree on the security/access policies of federated users between the parties in a federated relationship. For instance, whether there are duties to notify others in the event of security failures.

When IDM and the RP are in the same organization, the agreements may be more tacit.

When the IDM and RP are both built into a single system framework that allows for mutual trust may be completely opaque to the system operator, although the system developer may be aware of the framework or at least its implications since he or she will need to implement mechanisms that support the trust.

Trust Root

A trust root is a technical structure that provides the IDP and RP the ability to recognize each other with a high degree of certainty. This is similar to the concept of Trust Anchor (NIST SP.800-63-3), but we allow for a structure that relies on a mutually agreed-upon third party. A trust root derives from the operation of a Trust Framework. There is a need for a trust root so that the systems can operate without human involvement in every transaction. This may be done through a Public Key Infrastructure (PKI), where the parties agree to trust a common certificate authority that signs the certificates of all parties in the federation. This may be done through a set of several independent certificates that the parties agree to trust.

Provisioning

Provisioning is a term that encompasses the processes and methods that create, modify, and, eventually, delete the identity and profile information used by IT infrastructure and business applications. By these methods, records are created or updated in the identity register and removed from it. Often, provisioning needs to extend to applications to support authorization decisions. This is sometimes known as “downstream provisioning”. The term “Onboarding” is sometimes used to refer to the sum of the initial provisioning activities in both the identity and access aspects.

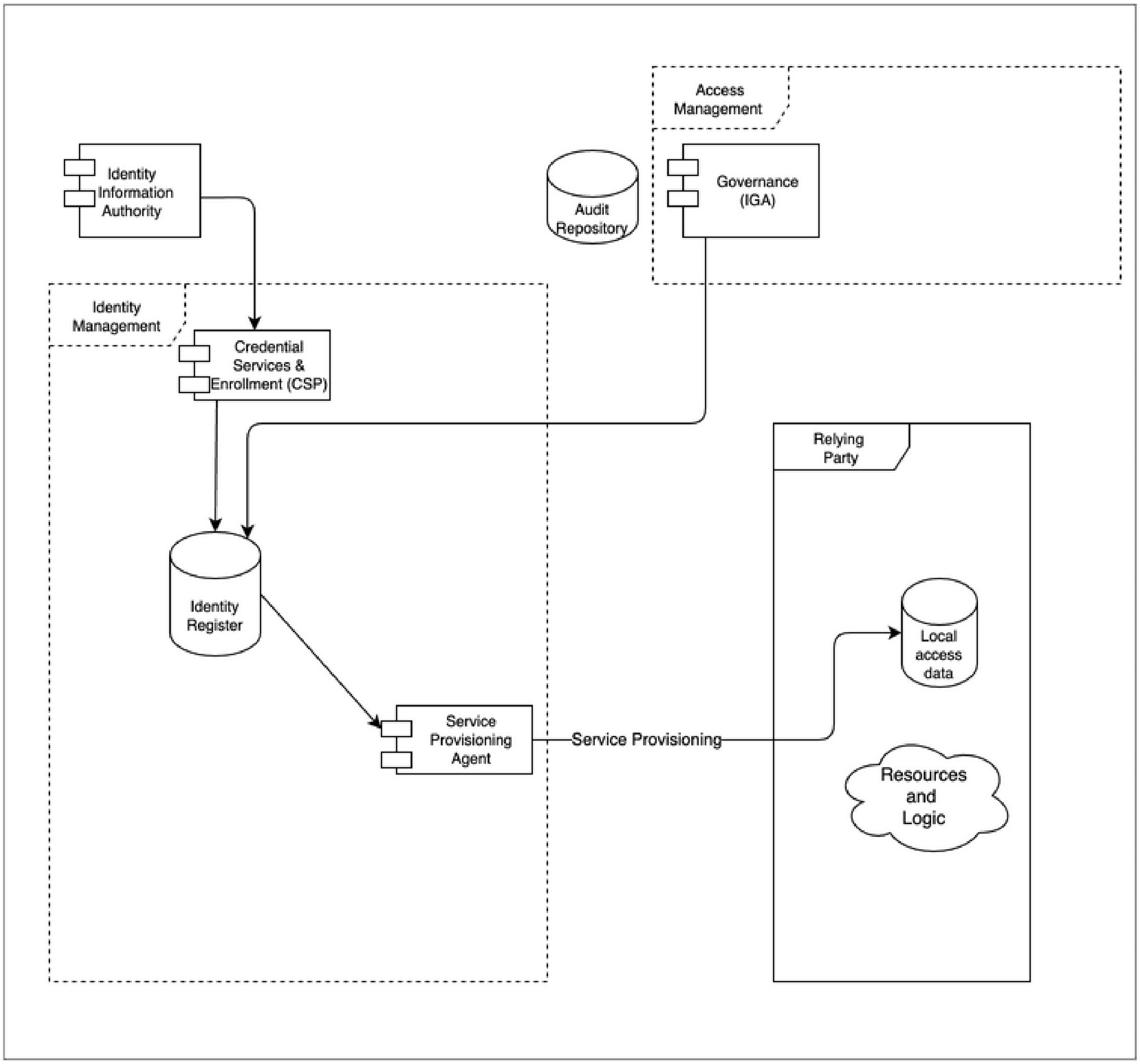


Figure 2: Provisioning: The Identity register receives updates from one or more external sources and administrative actions, passing the information on as needed.

Identity Information Authorities

While it is possible to have an IDM populated without attaching to an external data service, this is typically not the case. Usually, employee or customer data needs to be imported. This can be referred to as upstream provisioning.

Note that the authoritative sources for identity attributes transcend the HR system and may include email, phone, training certification system, etc. In some cases, a company may have more than one HR system.

Governance

The act of provisioning may include certain logic, best modeled as governance. In some cases, the IGA system takes on all the provisioning duties (see also the section on Access Governance below).

Credential Services & Enrollment

This function includes steps needed to originate and activate an identity. It is also concerned with ongoing maintenance such as password reset and key rotation. This function includes administrative activities and self-serve activities.

Enrollment

Also sometimes known as Registration. It involves such activities as proofing, verification or vetting, and recording sponsorship if needed. It also is responsible for the secure delivery of credentials. Enrollment ends when a user formally receives ownership of their digital identity and assumes control and ownership of their account's credentials.

Credential Services

Credential services include the creation of passwords, cryptographic keys, and other authenticators. It associates or "binds" these to an identity record. It is also concerned with ongoing maintenance such as password reset and key rotation and revocation of credentials as needed.

Identity Register

This is the datastore that contains the enrolled entities and their attributes, including credentials. In this model, we use the singular, as if it were one singular database. In practice, designs may store some attributes separately from identities. We also use this term to include the storage related to credentials, although all or some of the credentials may be stored in their own physical repository.

Identity Registers, by their nature, have high availability requirements. Often at the physical level, they contain multiple instances which are synchronized. The Identity Register could be implemented in several ways. Common methods include the use of general-purpose databases, optimized stores such as directories, either physical or virtual.

Importing data does not necessarily mean making a physical copy of data, although it often does. The notion also supports the idea of virtualization - where the import of identity information is done at run-time.

Service Provisioning Agent

Also noted is the function of propagating selected information further into the ecosystem. This typically occurs when an RP needs additional information about the users, e.g., for access control or personalization. The RP makes a copy of the identity data for future use in the application processes. A complete solution will support the full data lifecycle, including creation, update, and eventual deletion of the identity data stored locally.

Just in Time Provisioning

So far, the discussion of the provisioning function has been focused on “admin-time”. However, there are some cases where provisioning occurs at run time.

Not shown here, but sometimes implemented, are provisioning actions that occur on a just-in-time basis. This can happen when additional identity information is passed to an RP in real-time to support a specific application requirement, possibly including identity attributes (See Authentication and Sessions). A similar case involves the RP querying the IDM to acquire attributes (see Authorization later in this document)

Audit Repository

The audit repository is shown to indicate the accumulation of historical event data. To avoid clutter, we assume audit information is written but call that out with arrows in the diagram.

Authentication and Sessions

Authentication

Authentication is the process by which a subject’s credentials are used to verify their identity. The IDP checks and verifies credentials that are presented to it. There are multiple scenarios. Typically, the RP asks the Identity provider to gather the credentials from the user and receives an assessment from the IDP regarding the level of certainty that the user is authentic. Often the assessment (and more information about the user) is delivered to the RP via a security token, which is protected by cryptography. There are several varieties of security tokens. The diagram uses bidirectional arrows to show that use cases exist that require ongoing exchange of information as describe in the section in this document called “Sessions.”

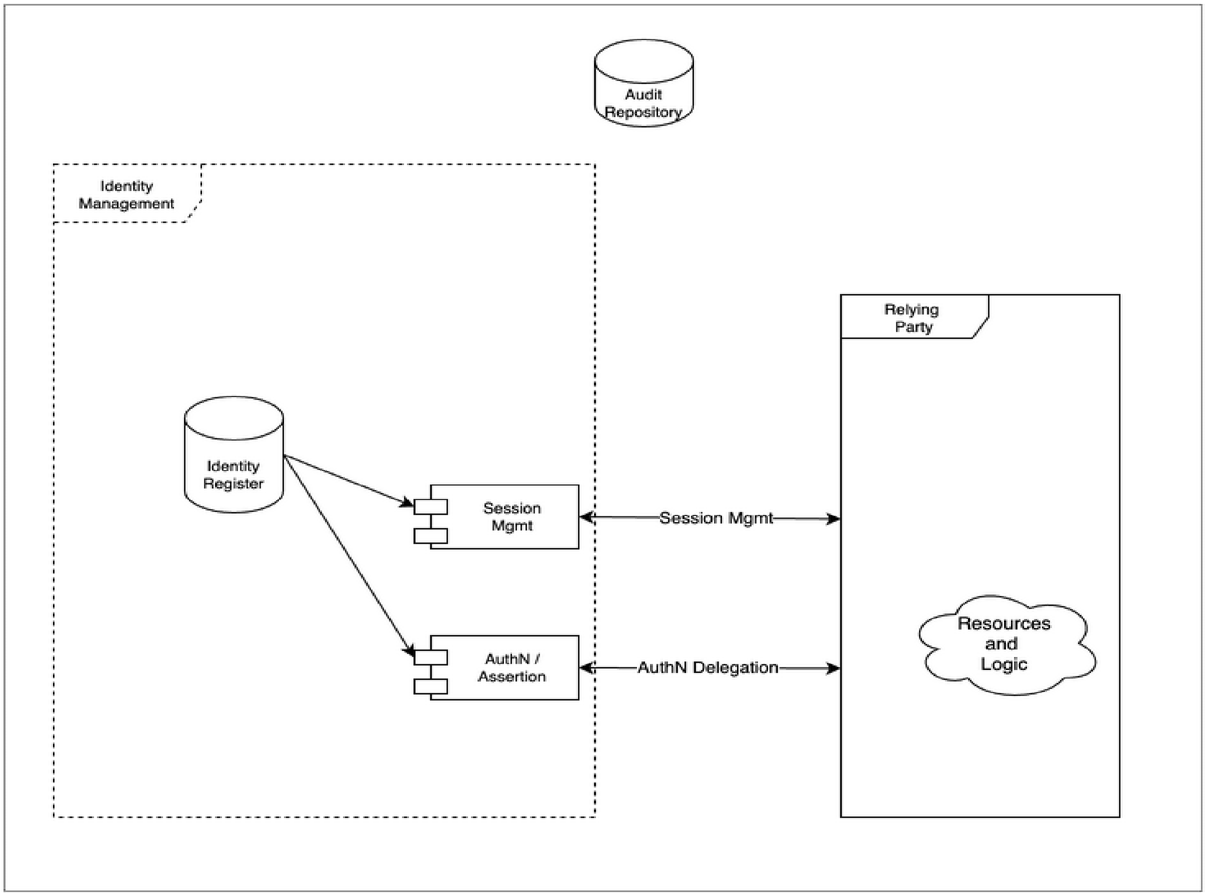


Figure 3: Authentication and Sessions: The Identity Register supports authentication scenarios. The IDP may monitor or participate in the full session lifecycle with the relying parties.

Sessions

A common pattern is to associate the authentication event with the start of a session. The session is mostly the concern of the RP. However, it is sometimes desirable to keep the sessions of several relying parties in synch. For instance, logging out of one session will terminate concurrent sessions. To do this, often the IDP will act to orchestrate sessions termination. In high-security environments, session management must support termination based on real-time identity data, such as when a user's entitlements have been modified.

The existence of a centralized point of view about sessions can be leveraged to support good security practices. For example, if the identity attributes of a user with an active session change and these new values then contravene an access control policy, the session should terminate. If session management becomes aware of a terminated account, it should end any active session that the user has. This could also occur in advanced scenarios that include facts presented by external risk monitors. See Risk Context below.

Sessions also support another important concept: step-up authentication. A session can keep track of the level of assurance of a particular authentication, so when a user requests access to a transaction or application requiring a higher level of identity assurance, the IDP can be prepared to determine the course of action, such as improving the certainty that the user is the right person by asking the user provide additional evidence. For example, maybe the password is good enough to review some information, but to withdraw money, the additional factor of a one-time password from a phone app is required. The detection of the assurance gap and subsequent action will logically be done at the RP, but to avoid a poor user experience in multiple RP scenarios, the step-up needs to be recorded in the session.

Authorization

Authorization models are many and diverse. The diagram illustrates two approaches for authorization: local and shared. As noted below, both approaches are subject to Access Governance.

Both approaches typically use subject attributes to help determine access, although some systems rely on direct enumerations mapping users to resources known as access control lists.

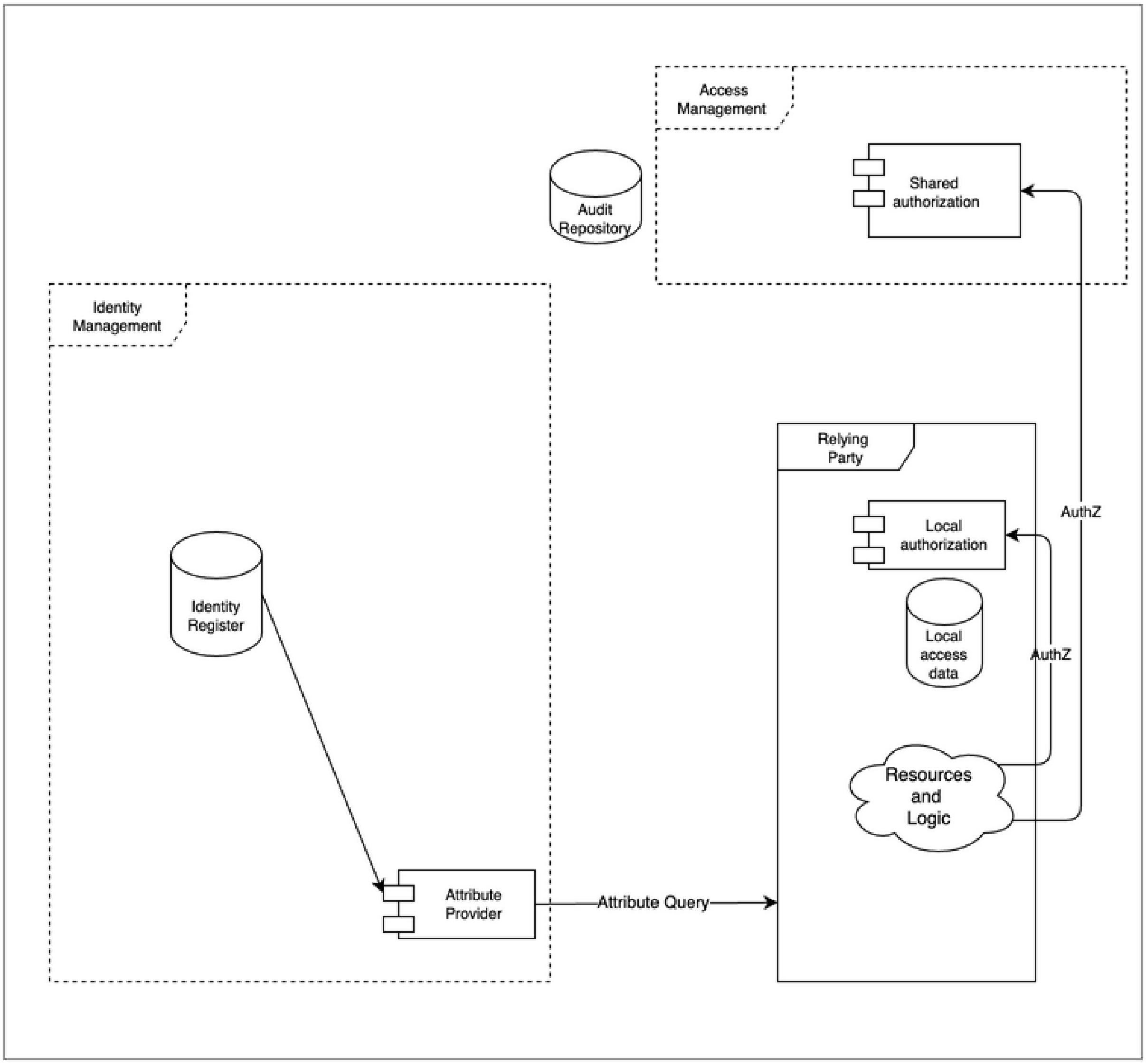


Figure 4: Authorization models: Some RPs perform authorization tasks internally. Sometimes authorization is a shared resource for many RPs.

Local Authorization

Many Relying parties perform authorization tasks internally. Often the fine-grained access control required by a protected resource makes this appealing. For instance, a financial management system may maintain a user's entitlements to specific functionality within the application. In this scenario, the application makes the authorization decision and implements (enforces) the result.

The controlling values may have been provisioned into the local access data store by the Provisioning process described above. Or the values can be acquired at run-time from the IDM as shown by the attribute query, which may provide the user's role or other attributes during the sign-on, perhaps as a value in the security token.

Shared Authorization

Sometimes authorization is a shared resource for many Relying parties. This design can improve the consistency of authorization decisions and supports organizations wishing to include advanced access decisions strategies such as those required by a "Zero Trust" access control approach. Shared authorization systems typically have a consistent approach to policy, such as a standardized policy language. In this scenario, the RP asks the shared authorization function to make the decision but implements (enforces) that itself.

Authorization Mechanisms

In either approach, the access rights may be established, maintained, and revoked in a variety of ways, starting with the existence and validity of the digital identity. Other controls include various mechanisms such as policies, roles, permissions, and identities. Some controls rely on user attributes, including group memberships or roles stored in an Identity Register. Some controls may depend on the properties of the accessed resource or the context of the request, such as time, device, or location.

Each mechanism relies on a particular logical data structure to implement the access control; that data structure becomes the focus of implementers. For instance, in role-based access control, there is some art involved in "Role Management" (defining and managing a useful set of roles) since too many roles become difficult to manage, whereas too few leads to users with access to things they don't need. Similarly, in the case of policy-based access control, the set of policies (the Policy Rules) needs to be designed, stored, and managed.

Access Governance

Access Governance, also known as Identity Governance and Administration (IGA), provides control over access rights implemented in multiple local or shared authorization systems. This function is often broken into the administration of these rights and the oversight needed to ensure that these rights are in good order over time.

In enterprise systems, Access Governance focuses on managing staff (employee/contractor) entitlements. The concept can also apply to other scenarios, such as when business-to-business delegated administrative rights are required or to in business-to-customer scenarios where authorized third parties such as attorneys are required.

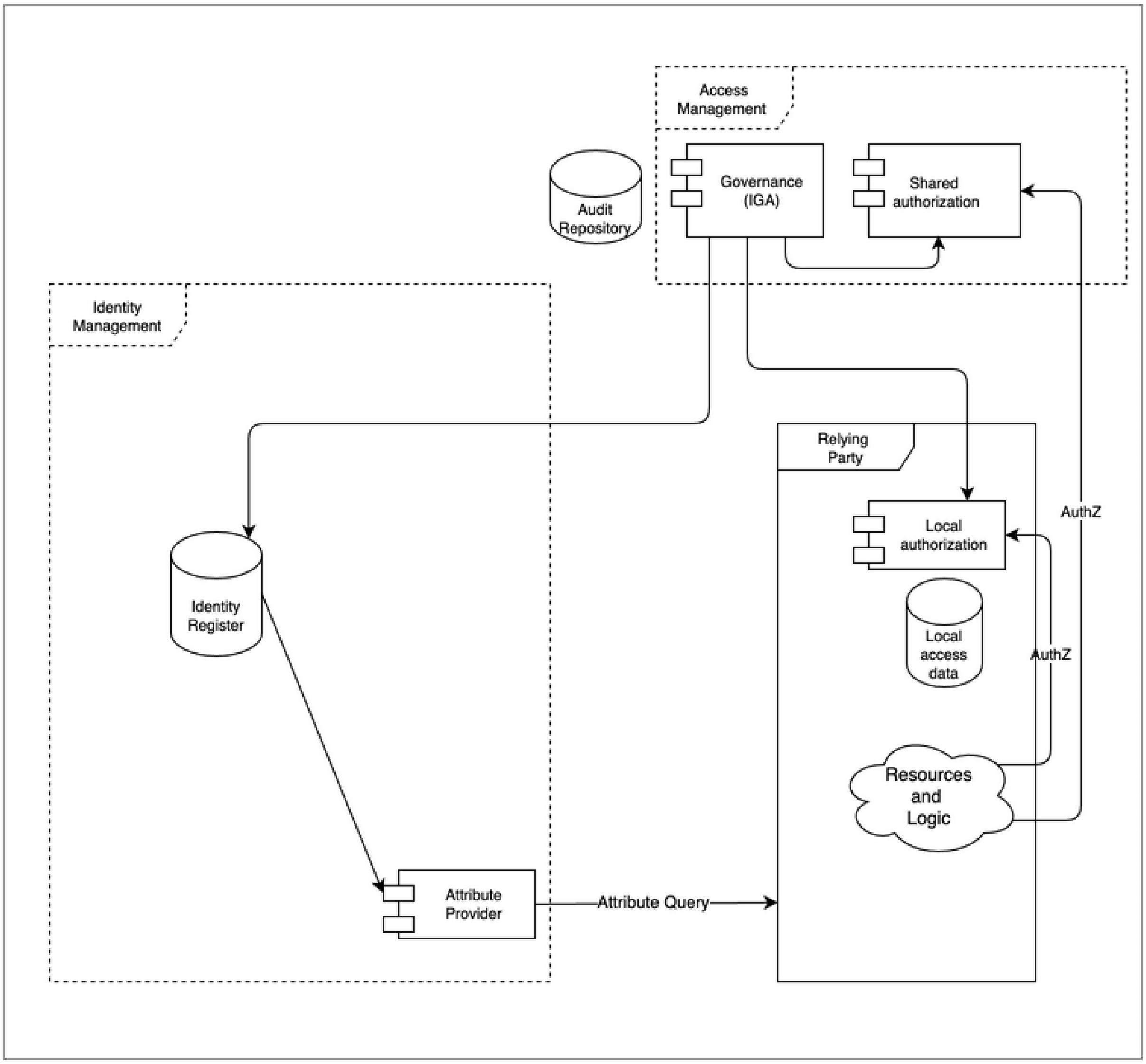


Figure 5: Access Governance provides oversight and control over access rights implemented in many Local authorization systems and, sometimes, in Shared authorization systems.

Control

The controls may also include methods such as procedures and workflows to ensure proper review. Typically, a request for access to resources is passed to one or more approvers and an audit trail is created.

Often deployed to prevent internal fraud is the “segregation of duties” control. The control defines groups of access rights that cannot be held by the same person. This control is best implemented in a location that has visibility to all the implicated access rights, i.e., the IGA system.

Oversight

Typically, governance activities review and potentially modify the data in one or more of the authorization components in order to effect a change in entitlements. Often organizations will have a formal process to review existing entitlements and may require a responsible party to certify or attest that the entitlements are in good order. Additional tools to ensure that IAM policies are effective at enforcing their stated controls include internal and external audits as well as analytic reports.

Risk Context

Risk Context (often abbreviated as RCTX) information can be valuable to improve the security of the relying service. Risk can be judged based on information in the request, information about the history of the user, or assertions/evidence from third parties.

The linkage from the Audit Repository illustrates that the Risk Context may consume the local historical data about events.

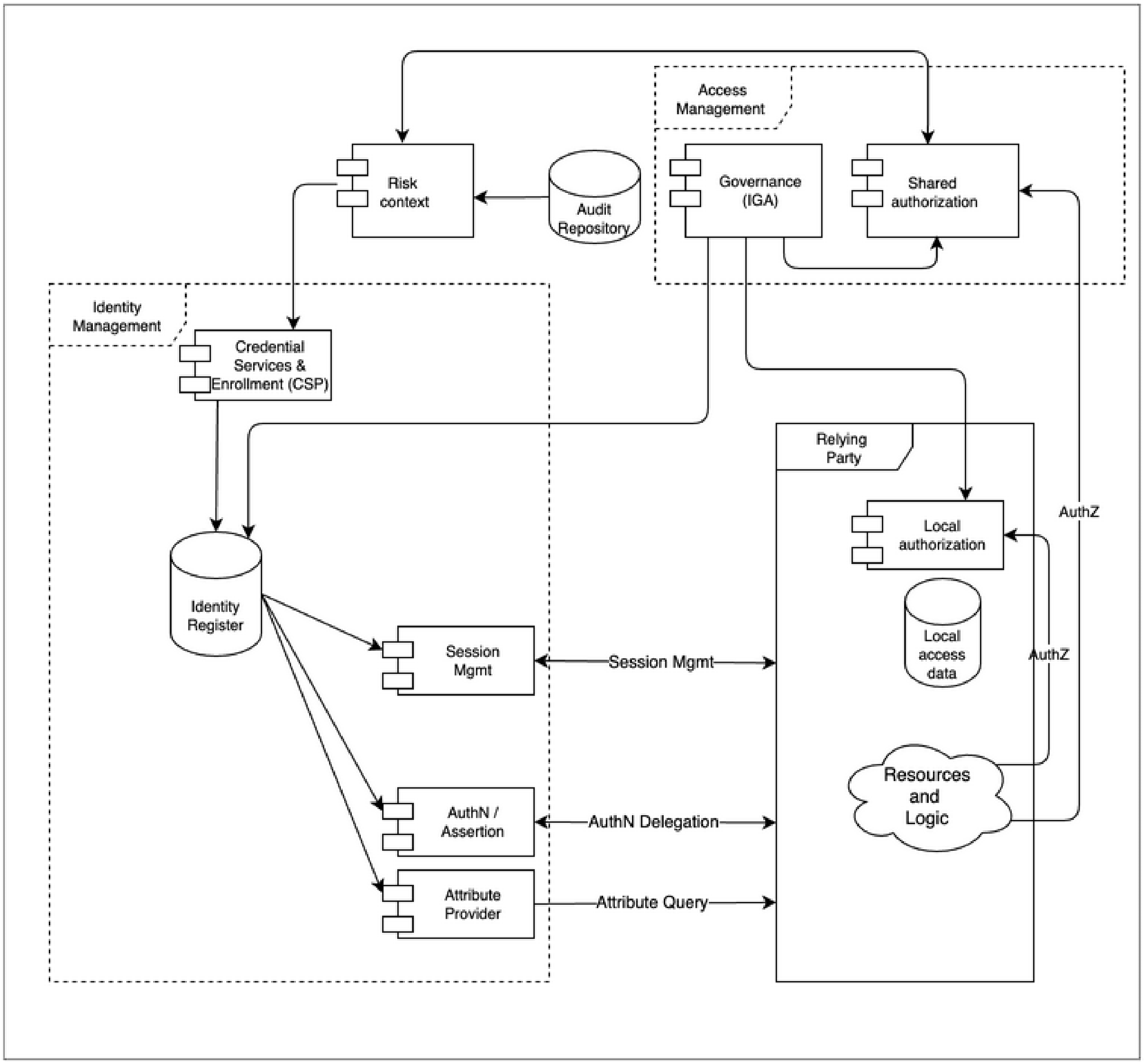


Figure 6: Risk Context: It is possible to use risk information in authentication decisions. For instance, if a stolen password is found on the dark web, don't allow login.

External events may be visible to the IDM operator through consortia or vendor packages. In some mutual-support scenarios, it may be possible for the IDM operator to also publish events for the benefit of others, supporting other operators' risk management requirements.

Events need to be delivered into the IDM so that they can selectively be used to modify the behavior of the authentication function. For example, armed with additional event data, the authentication function may request a step-up authentication or even plainly deny access.

In some severe scenarios, attaching the events to the session management function may be desirable so that current sessions can be reviewed and terminated if needed. The

OpenID Shared Signals and Events working group is developing standard ways to deliver these signals. [x]

As shown in the diagram, shared authorization systems may consume risk data as well. For example, an authorization might be denied if the subject's recent activity history is outside of normal bounds, possibly indicating a compromised credential. Logically this could happen with local authorization as well, but this is not shown.

Example: Information in the request

Boundary control

An authentication or authorization decision may be influenced by specific criteria, such as whether a request is coming from a known or unknown network. A more sophisticated version of this attempts to prohibit access from, say, certain countries.

Examples: Historical usage

Usage pattern match

Determine if this request is outside the normal usage patterns for a given individual. The reference to historical usage patterns allows for pattern detection and can help establish a metric for risk for a user, a specific transaction, or in general. Such activity can be called risk profiling.

Land speed violation

Amending the user's request and history with location information makes it possible to identify a likely compromised account because the user can't be in two places at once.

Such examples depend on signals from the local environment, but it is also possible to obtain signals from further afield.

Example: Third party

it is possible to determine commonly used passwords based on postings on the "dark web." Bad actors acquire these in the hope that users will use the same password at other sites. A countermeasure is for the IDM operator to require additional certainty if one of those passwords were presented.

Metadata and Discovery

Metadata refers to control data that allows the IDM and the Relying Parties to interoperate.

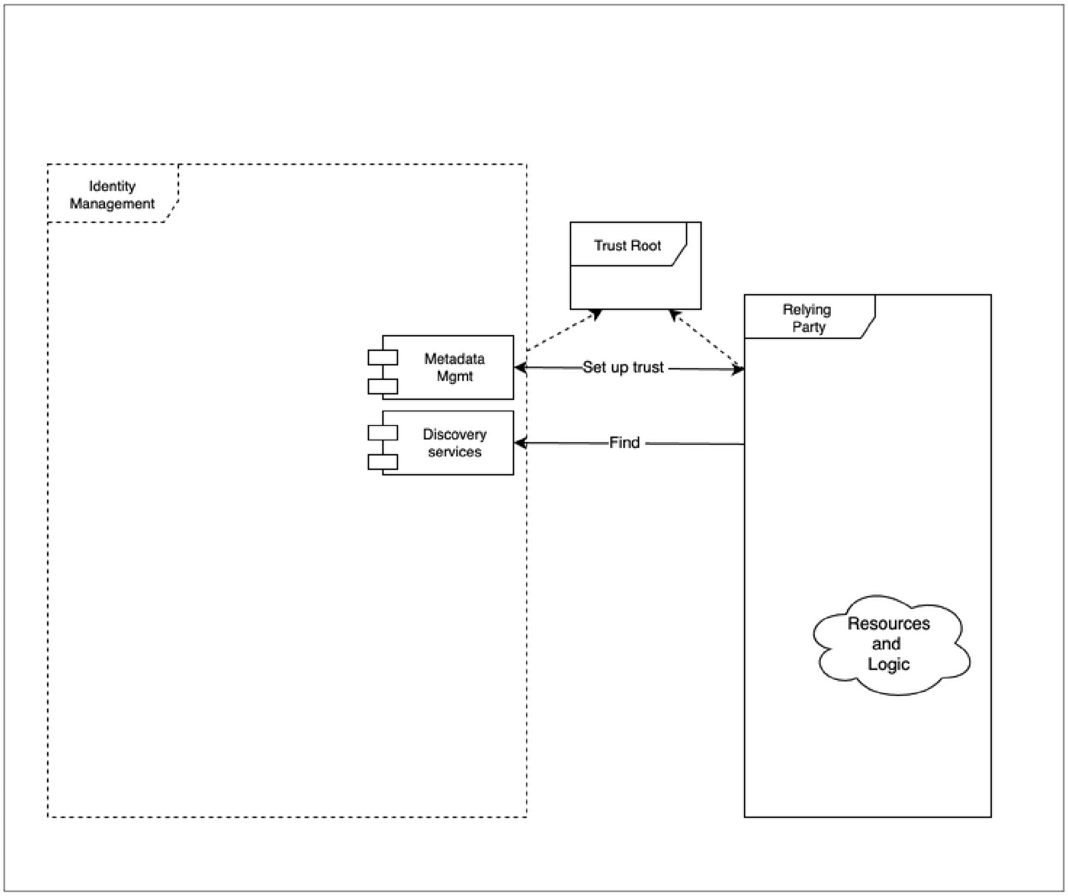


Figure 7: Metadata and discovery these two functions are involved with mutual recognition of the IDM and Relying Service.

One example is the registration of public-key certificates to enable mutual authentication. In some scenarios, this information is shared between the parties manually. At run-time for distributed systems, the technical root of trust is needed to validate the security channel (PKI)

Another example points out that configuration information is another form of metadata. OpenID Connect has a list of required, recommended, and optional values that describe a particular implementation aimed at providing a degree of automation during setup.

The metadata may include information that limits the types of interactions and scope of the data that is exchanged. It can also contain security information to allow the counterparties to authenticate each other. For instance, public key components such as certificates with a common certificate authority may be used.

Discovery refers to protocols that facilitate automation. For instance, OpenID Connect defines a method for RPs to locate an end-point where a user's identity can be verified.[xi] The concept is more supported by other methods such as SAML.[xii] A Discovery service can advise where specific data can be accessed and which end-points are maintained to allow an RP to use the identity service.

Author Bio

George Dobbs manages architects at a major insurance company. He is also the chairman of the IDPro Body of Knowledge Committee. One of his interests is modernizing the use of Identity and Access Management techniques used by the firm. He is particularly interested in the area of customer-facing applications, including approaches to fraud prevention in call center and digital contexts. Related to this, he is interested in the evolution of distributed session management – notably distributed session termination. He is a founding member of IDPro and represented his firm in the Identity Ecosystem Steering Group (IDESG). Prior to his current position, he led the development of customer-facing identity for websites at three other insurers. He has led a local identity and access management user group since 2004. Prior to that, he was the chairman of the Network Applications Consortium.

Acknowledgments

The author would like to express gratitude to Ian Glazer, Graham Williamson, and Corey Scholefield for the detailed reviews of early drafts; Jon Lehtinen and Steve Hutchinson for some of the definitions from their unpublished Introduction to Identity Part 3 document; and Bertrand Carlier for his thorough and thoughtful review.

Change Log

Date	Change
2021-09-30	V1 published
2022-12-15	V2 published; minor editorial changes; some clarification in the text re: Credential Services and in Authentication

References

[i] Wikipedia contributors, "All models are wrong," *Wikipedia, The Free Encyclopedia,* https://en.wikipedia.org/w/index.php?title=All_models_are_wrong&oldid=1111346950 (accessed November 28, 2022).

[ii] ISO/IEC 24760-1 Second edition “IT Security and Privacy — A framework for identity management — Part 1: Terminology and concepts,” https://www.iso.org/standard/77582.html and ISO/IEC 24760-2, 2015 “Information technology — Security techniques — A framework for identity management — Part 2: Reference architecture and requirements,” https://www.iso.org/standard/57915.html (accessed 28 November 2022).

[iii] “FICAM Playbooks – FICAM Architecture – System Component Examples,” Identity Assurance and Trusted Access Division in the GSA Office of Government-wide Policy, https://playbooks.idmanagement.gov/arch/components/ (accessed 28 November 2022).

[iv] Hazelton, Keith “The TAP Reference Architecture (RA)” https://spaces.at.internet2.edu/pages/viewpage.action?pageId=98306902 (accessed 28 November 2022).

[v] Grassi, Paul A., Michael E. Garcia, James L. Fenton, “NIST Special Publication 800-63-3 – Digital Identity Guidelines,” National Institute of Standards and Technology, U.S. Department of Commerce, June 2017, https://doi.org/10.6028/NIST.SP.800-63-3.

[vi] Rose, Scott, Oliver Borchert, Stu Mitchell, Sean Connelly, “NIST Special Publication 800-207 – Zero Trust Architecture,” National Institute of Standards and Technology, U.S. Department of Commerce, August 2020, https://doi.org/10.6028/NIST.SP.800-207.

[vii] Hutchinson, Steve, “Introduction to Identity Part 2 - June 25,” Identiverse 2019, recording starting minute 27:39, https://www.youtube.com/watch?v=zxKRUXmTLJs&list=PLpKq7xRiIHaTDwAqpIU1UYpKZY03tfTMf&index=8.

[viii] Smedinghoff T. J., (2021) “Laws Governing Identity Systems (v2),” *IDPro Body of Knowledge* 1(5). https://bok.idpro.org/article/id/8/.

[ix] Temoshak, David, Christine Abruzzi, “NISTIR 8149 - Developing Trust Frameworks to Support Identity Federations,” National Institute of Standards and Technology, U.S. Department of Commerce, January 2018, https://doi.org/10.6028/NIST.IR.8149.

[x] “Shared Signals and Events WG” https://openid.net/wg/sse/ (Accessed 28 November 2022).

[xi] Sakimura, N., J. Bradley, M. Jones, E., Jay, “OpenID Connect Discovery 1.0 incorporating errata set 1,” OpenID Foundation, 8 November 2014, https://openid.net/specs/openid-connect-discovery-1_0.html (accessed 28 November 2022).

[xii] Widdowson, Rod, Scott Cantor, “Identity Provider Discovery Service Protocol and Profile,” OASIS Committee Specification 01, 27 March 2008, https://docs.oasis-open.org/security/saml/Post2.0/sstc-saml-idp-discovery.pdf (accessed 28 November 2022).

Delegated Authentication Using a SAML Web Browser SSO Profile (v2)

By George B. Dobbs

To comment on this article, please visit our GitHub repository and submit an issue.

Table of Contents

Abstract

This article builds on a generic IAM reference architecture to describe a common use case of an authentication service using the Security Assertion Markup Language (SAML). We show how a service (the relying party) uses the authentication capability of an identity provider during a web-based single sign-on action.

Introduction

This article is one of a set that illustrates several abstract components defined in the IDPro Body of Knowledge article, "IAM Reference Architecture."[i] This particular article focuses on a specific method of web-based single sign-on via the common use case of a service (the relying party) that uses the authentication capability of an identity provider (IDP) via the Security Assertion Markup Language (SAML) standard. This method allows the RP to delegate the authentication function to the IDP.

This widespread use case relies on trust between the IDP and the relying party (RP). There are various ways of doing this; this article assumes that the trust is based on the use of public-key cryptography, which involves exchanging certificates.

The SAML specification defines three different kinds of assertion statements; this article is only about the authentication assertion.

The SAML specification supports the mapping of identities between different names, known as federated identity. This article is restricted to a single domain, such as an organization providing access for its employees to web-based services provided by third-party vendors. In other words, a single domain of administration allows for the user identifiers to be shared.

Even a cursory review of the OASIS SAML standards documents will reveal an extremely rich and flexible structure.[ii] This article represents a very thin slice of its possibilities focusing on the run-time aspect of authentication using the web (HTTPS) messaging protocol. Technically, we are discussing what OASIS calls the Web Browser SSO profile, using the POST binding.[iii]

This synopsis stresses the importance of the Trust Root. The messaging between the IDP and RP passes through the User Agent. The User Agent must be considered untrusted, as a corrupted agent could potentially modify the messages. To protect against this modification, the messages are protected by a digital signature, which must be validated. It is the common certificate authority that acts as the Trust Root to support these signatures.

The topic of signatures becomes quite deep quickly and is not covered in detail here. The SAML specification relies on the W3C Recommendation XML Signature Syntax and Processing, which may be of interest.[iv]

Terminology

Please see also the terminology in the IDPro Body of Knowledge article, "IAM Reference Architecture."

Item	Definition
User Agent	A user agent is any software that retrieves, renders, and facilitates end-user interaction with Web content. [v]

Use Case

Summary

The web user works through a user agent to access resources at an RP. The access request results in a redirection of the user to an IDP as part of an authentication action. This result of the authentication is an authentication assertion that is consumed by the RP and used to establish a security context for the web user. In effect, the RP has delegated the authentication to the IDP.

Architecture Types

Different architecture types are defined in the "Introduction to IAM Architecture" article in the IDPro Body of Knowledge. The SAML-based authentication use case applies specifically to the architecture of Cloud Environments.[vi]

Actors

The user is the only actor. The user acts through the User Agent (the browser). The other participants in the use-case are systems "actors", which we show as components.

Components

The following components are defined in the article, "IAM Reference Architecture."[vii]

- Audit Repository
- AuthN / Assertion (part of IDP)
- Identity Register
- Relying Party (RP)
- Trust Root

Please note that the SAML documents refer to the relying party as the service provider.

Assumptions

The user wishes to access a protected resource and has requested access via a web browser, the User Agent.

The RP has a single IDP. The RP may support several IDPs, so a method to determine which one to use would be needed in that case.

Preconditions

There must be an established trust between the RP and the IDP before SAML can be used for authentication. “The primary mechanism is for the relying party and asserting party to have a pre-existing trust relationship which typically relies on a Public Key Infrastructure (PKI). While SAML does not mandate using a PKI, it is recommended.”[viii]

The IDP and the RP use the same user identifiers. The SAML specification establishes ways to map these, but we don’t discuss this subject here.

Postconditions

The user is logged into the RP’s site.

Basic Course of Events

The following shows the “happy path”, without errors. See also the sequence diagram below. See Alternative Paths for some variations.

1. The user selects the login function on the RP’s site. This selection may be automatic when the user attempts to access the protected resource.
2. RP determines that the user is not logged in.
3. The RP prepares an Authentication Request message, which the RP may sign. It is delivered to the user agent as a form targeted at the IDP, which is known since there is a single configured IDP. The user agent (automatically via a client-side script) sends the request to the IDP.
4. The IDP ensures the signing certificate from the RP is still valid by checking for revocation.
5. The IDP validates the request and interprets its contents. The signature and some field values (such as Issuer, AuthnContextClassRef, etc.) are checked.
6. The IDP interacts with the user agent to gather the user’s identifier and credentials. For example, this could ask for a username and password, but it could be something else.
7. The IDP uses its Identity Register to validate the credentials.
8. The IDP prepares a Response message, which the IDP signs.

9. The response is then sent back to the user agent with instructions to use an HTTP POST to forward it to the RP. (The target RP URL is typically known to the IDP through initial configuration).
10. The RP ensures the signing certificate from the IDP is still valid by checking for revocation.
11. The RP validates the response and interprets its contents. The signature **must** be checked. The RP checks it against the already active assertions to prevent replay and makes other checks. The RP then determines whether the authentication was successful.
12. Not shown in the chart are the audit records being written. The various components should write these.

Alternative Paths

Step 3 may be replaced with an HTTP redirect. This formulation is an allowed composition of the POST binding and the Redirect binding.[ix]

There are also alternatives to the POST method in step 3.[x]

In SAML terms, this is the service provider-initiated variant. There is also an IDP-initiated alternative.

The messages may be encrypted. For instance, in step 8, the IDP may encrypt, and in step 10, the RP would need to decrypt the response.

Not recommended: some implementations have ignored request signing and signature verification, possibly due to historical performance issues.

Exception Paths

Failure to authenticate at the IDP does not return an assertion.

Failure to validate the signature indicates that the assertion should not be honored.

Various error conditions, such as the validity period expiring, are described in the standard.

Sequence Diagram

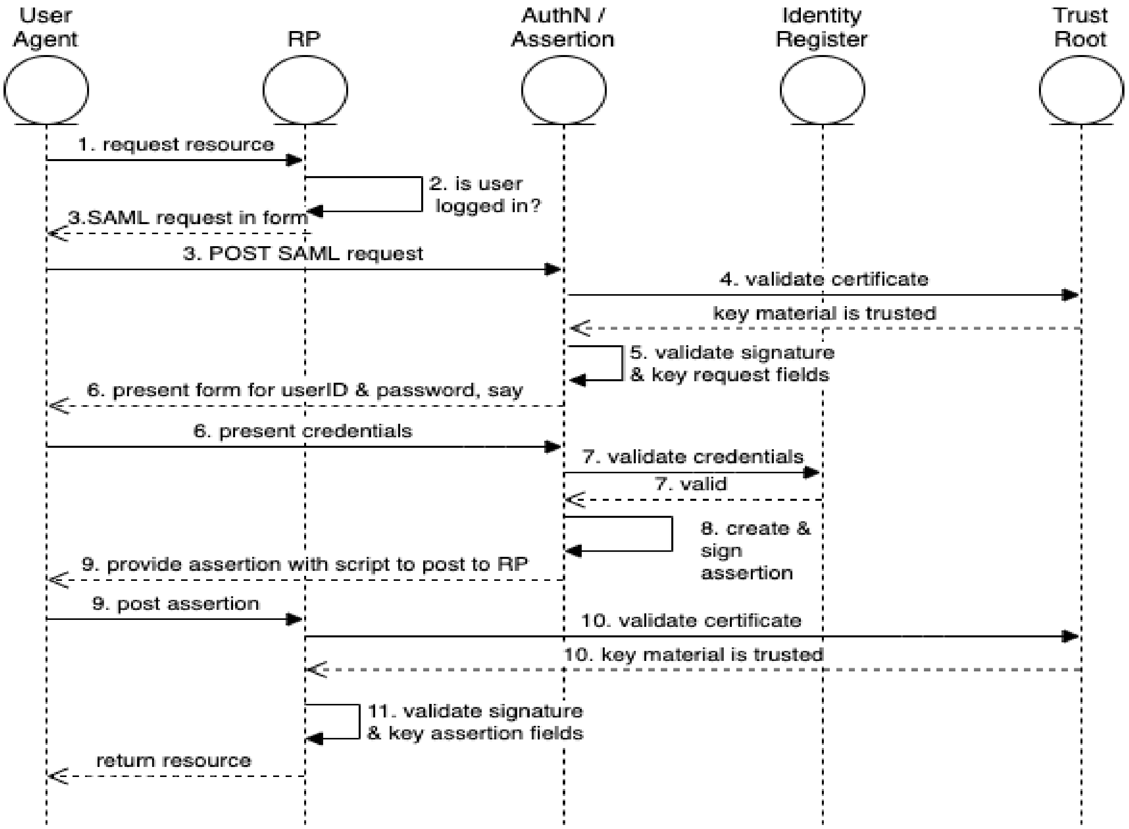


Figure 1: The "happy path" of the Web Browser SSO Profile

Author Bio

George Dobbs manages architects at a major insurance company. He is also the chairman of the IDPro Body of Knowledge Committee. One of his interests is modernizing the use of Identity and Access Management techniques used by the firm. He is particularly interested in the area of customer-facing applications, including approaches to fraud prevention in call center and digital contexts. Related to this, he is interested in the evolution of distributed session management – notably distributed session termination. He is a founding member of IDPro and represented his firm in the Identity Ecosystem Steering Group (IDESG). Prior to his current position, he led the development of customer-facing identity for websites at three other insurers. He has led a local identity and access management user group since 2004. Prior to that, he was the chairman of the Network Applications Consortium.

Acknowledgments

The author would like to express gratitude to Chris Olsen and Bernard Carlier for their review and suggestions for improving this text.

Change Log

Date	Change
2021-09-30	V1 published
2022-12-15	V2 published; title changed, intro and use case summary clarified; Alternate Path includes a “not recommended”

References

[i] Dobbs, George, “IAM Reference Architecture,” IDPro Body of Knowledge, 30 September 2021, https://bok.idpro.org/article/id/76/.

[ii] “OASIS SAML Wiki – Front Page,” wiki page, OASIS, https://wiki.oasis-open.org/security/FrontPage#SAML_V2.0_Standard (accessed 28 November 2022).

[iii] Hughes, John, and Scott Cantor, Jeff Hodges, Frederick Hirsch, Prateek Mishra, Rob Philpott, Eve Maler, eds. “Profiles for the OASIS Security Assertion Markup Language (SAML) V2.0,” OASIS, 15 March 2005, https://docs.oasis-open.org/security/saml/v2.0/saml-profiles-2.0-os.pdf (accessed 28 November 2022).

[iv] Bartel, Mark, and John Boyer, Barb Fox, Brian LaMacchia, Ed Simon, “XML Signature Syntax and Processing Version 1.1,” Section: Core Validation, World Wide Web Consortium, 11 April 2013, https://www.w3.org/TR/xmldsig-core/#sec-CoreValidation, (accessed 28 November 2022).

[v] “User Agent Accessibility Guidelines (UAAG) 2.0,” W3C Working Group Note, https://www.w3.org/TR/UAAG20/#glossary (accessed 28 November 2022).

[vi] Cameron, Andrew, and Graham Williamson, “Introduction to IAM Architecture,” IDPro Body of Knowledge, 17 June 2020, https://bok.idpro.org/article/id/38/ (accessed 28 November 2022).

[vii] “IAM Reference Architecture,” https://bok.idpro.org/article/id/76/.

[viii] Ragouzis, Nick, and John Hughes, Rob Philpott, Eve Maler, Paul Madsen, Tom Scavo, “Security Assertion Markup Language (SAML) V2.0 Technical Overview - Committee Draft 02,” OASIS, 25 March 2008, https://www.oasis-open.org/committees/download.php/27819/sstc-saml-tech-overview-2.0-cd-02.pdf (accessed 28 November 2022)

[ix] Cantor, Scott, and Frederick Hirsch, John Kemp, Rob Philpott, Eve Maler, eds. “Bindings for the OASIS Security Assertion Markup Language (SAML) V2.0,” OASIS, 15 March 2005, https://docs.oasis-open.org/security/saml/v2.0/saml-bindings-2.0-os.pdf (accessed 28 November 2022).

[x] Ibid.

Federation Simplified (v2)

Patrick Lunney, Product Owner - Single Sign-On & Multi Factor Authentication
Capital One

To comment on this article, please visit our GitHub repository and submit an issue.

Table of Contents

Abstract

This article describes the fundamentals of enterprise identity federations, focusing on SAML and OpenID Connect (a protocol built on OAuth2.0). It will also contain common scenarios where federations are used and high-level terminology. Academic identity federations are out of scope but are mentioned briefly for comparison.

Introduction

This article describes identity federation in the context of single sign-on in enterprises and outlines some use cases for enterprise federation integrations. Enterprises have various ways to manage federation connections: the connections may be full service within the enterprise, self-service with controls in place for governance, or manual integrations. Each integration model has its strengths and weaknesses, which will be discussed in turn below.

Terminology

Term	Definition
Identity Federation	An identity federation is a group of computing or network providers that agree to operate using standard protocols and trust agreements. In a **Single Sign-On (SSO)** scenario, identity federation occurs when an **Identity Provider (IdP)** and **Service Provider (SP)** agree to communicate via a specific, standard protocol. The enterprise user will log into the application using their credentials from the enterprise rather than creating new, specific credentials within the application. By using one set of credentials, users need to manage only one credential, credential issues (such as password resets) can be managed in one location, and applications can rely on the appropriate enterprise systems (such as the HR system) to be the source of truth for a user's status and affiliation. Identity federations can take several forms. In academia, **multilateral federations**, where a trusted third party manages the metadata of multiple IdPs and SPs, are fairly common.[i] This article focuses, however, on the enterprise use case where **bilateral federation** arrangements, where the agreements are one-to-one between an IdP and an SP, are the most common form of identity federation in use today.

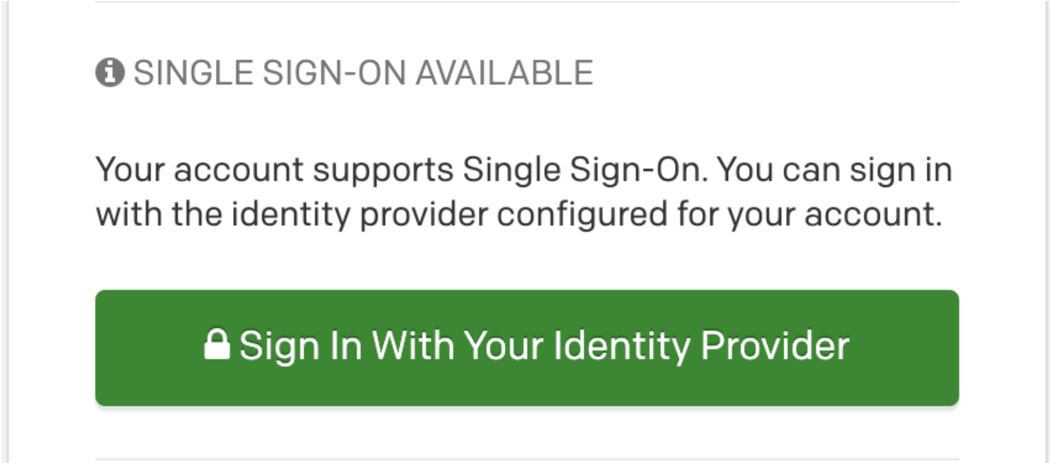
Bilateral Federation	A bilateral federation is one that consists of only two entities: one **Identity Provider (IdP)** and one **Service Provider (SP)**. This is the most common model for an enterprise identity federation.
Identity Provider (IdP)	An Identity Provider (IdP) performs a service that sends information about a user to an application. This information is typically held in a user store, so an identity provider will often take that information and transform it to be able to be passed to the service providers, AKA apps. The OASIS organization, which is responsible for the SAML specifications, defines an IdP as "A kind of SP that creates, maintains, and manages identity information for principals and provides principal authentication to other SPs within a federation, such as with web browser profiles." [ii]
Multilateral Federation	A federation that consists of multiple entities that have agreed to a specific trust framework. There are several forms of multilateral federations, including hub-and-spoke and mesh. Multilateral federations are the most common model for academic identity federations.
OAuth 2.0	OAuth 2.0 is an open-source protocol that allows Resource Owners such as applications to share data with clients by facilitating communication with an Authorization Server.[iii] That data takes the form of credentials given to applications to obtain information/data from other applications. The Authorization Server is usually the Identity Provider (IdP). The Authorization Server (AS) may provide authorization directly or indirectly. For example, the AS may supply attributes or profile data of the Resource Owner or provide access to data that can later be used for authorization purposes, such as entitlements from an Identity Management or Governance Solution.
OpenID Connect	OpenID Connect is a simple identity layer on top of the OAuth 2.0 protocol. It enables Clients to verify the identity of the End-User based on the authentication performed by an Authorization Server, as well as to obtain basic profile information about the End-User in an interoperable and REST-like manner.
Security Assertion Markup Language (SAML)	SAML is an XML-based communication protocol between SPs and IdPs.[iv] Usually, the enterprise hosts the IdP, whereas applications (including cloud services) are the SPs.
Service Provider (SP)	Defined by the OASIS organization, which is responsible for the SAML specification, as "A role donned by a system entity where the system entity provides services to principals or other system entities." This usually takes the form of an application that offers services requiring authentication and authorization to a user.

Single Sign-On	Single Sign-On is a service that enables SPs to verify the identities of **End Users** by facilitating communication with IdPs. SSO acts as a bridge to decouple SPs and IdPs. This can happen via numerous protocols such as agent-based integrations, direct LDAP integration, SAML, and OpenID Connect, to name a few.

Exploring Identity Federation in the Enterprise

There are several common scenarios where an identity practitioner is likely to encounter identity federation in an enterprise context. This section explores the most common protocols, OpenID Connect, and SAML.

Use Case 1: SAML


SINGLE SIGN-ON AVAILABLE

Your account supports Single Sign-On. You can sign in with the identity provider configured for your account.

Figure 1 - Example of a Single Sign-On User Interface

SAML is most often found in SaaS (Software as a Service) applications. An application is purchased or created by an enterprise to do "something" and employees need to log into the application. The application will need to exchange information with the enterprise to form this federation. Usually, an IdP (the enterprise) and an SP (the app) will exchange metadata, allowing them to set up the connections in the SSO system. Metadata exchange can be done manually, but that often takes time and can cause headaches for IdPs and SPs.

See Appendix Item 1 for an example of a metadata file from an IdP. In that example, the IdP operator will give this metadata to the SP operator. The SP can then input this information manually (or import it, depending on their SSO platform) into their SSO system to allow the enterprise's users to sign in to the application using their SSO accounts. The IdP operator will need to do the same, either by importing an SP metadata file or manually updating the configuration of the IdP.

An IdP metadata file contains the enterprise's entityID, the various URLs used in SAML, and the attributes that will be passed in the SAML assertion (the data that is passed to the app). An entity ID is a unique name for a SAML entity, both an IdP and an SP. No two IdPs or SPs can share the same entityID.

Think of a SAML assertion as a voucher or ticket. The IdP gives the user a voucher to the user to get into the SP, and the SP is validating the voucher using certificate validation. After the voucher is validated, the SP will look at the attributes to see what the user can do. For example, in Appendix Item 2, you can see a user's username and email address were passed to this SP.

For more information on the details of SAML assertions and components, see the SAML specification and associated supporting documents.[v]

One last piece of information regarding enterprise SAML federations: there are two different types of URLs for applications. Sometimes it is the SP’s URL, for example, ‘https://myhrapp.com/enterprise’. This is known as an SP-initiated request. Other times, the IdP will initiate the request. For example, ‘https://authn.enterprise.com/idp/SAML20=myhrapp’. In both cases, the user will be logging into the same app tenant for the enterprise. Some applications only support IdP-initiated login requests. Some applications only support SP-initiated requests.

Here is a diagram flow of a standard SAML authentication:

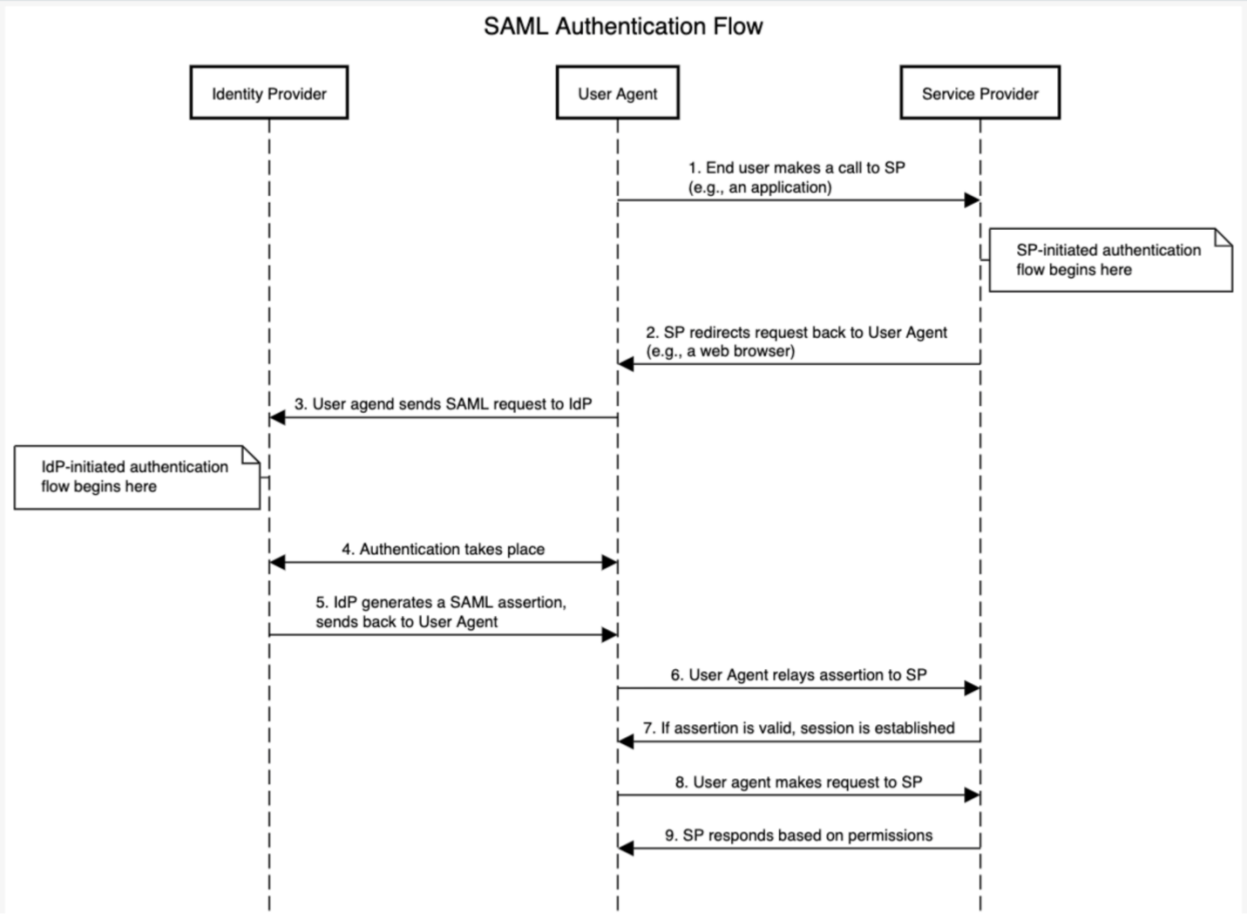

SAML Authentication Flow
Identity Provider
User Agent
Service Provider
1. End user makes a call to SP (e.g., an application)
SP-initiated authentication flow begins here
2. SP redirects request back to User Agent (e.g., a web browser)
3. User agend sends SAML request to IdP
IdP-initiated authentication flow begins here
4. Authentication takes place
5. IdP generates a SAML assertion, sends back to User Agent
6. User Agent relays assertion to SP
7. If assertion is valid, session is established
8. User agent makes request to SP
9. SP responds based on permissions

Figure 2 - SAML Authentication Flow

It should be noted that the *authentication* of the user is completed at step 5; the IdP has validated the user's credentials and is now passing the SAML assertion back to the browser. Federation is completed at step 7; the browser forwards the assertion to the application so that the application can know the user has been authenticated and create a session for that user. In steps 8 and 9, that is where *authorization* takes place. Based on the information provided by the IdP, the application will allow or deny the user access to certain parts of the application.

Use Case 2: OpenID Connect

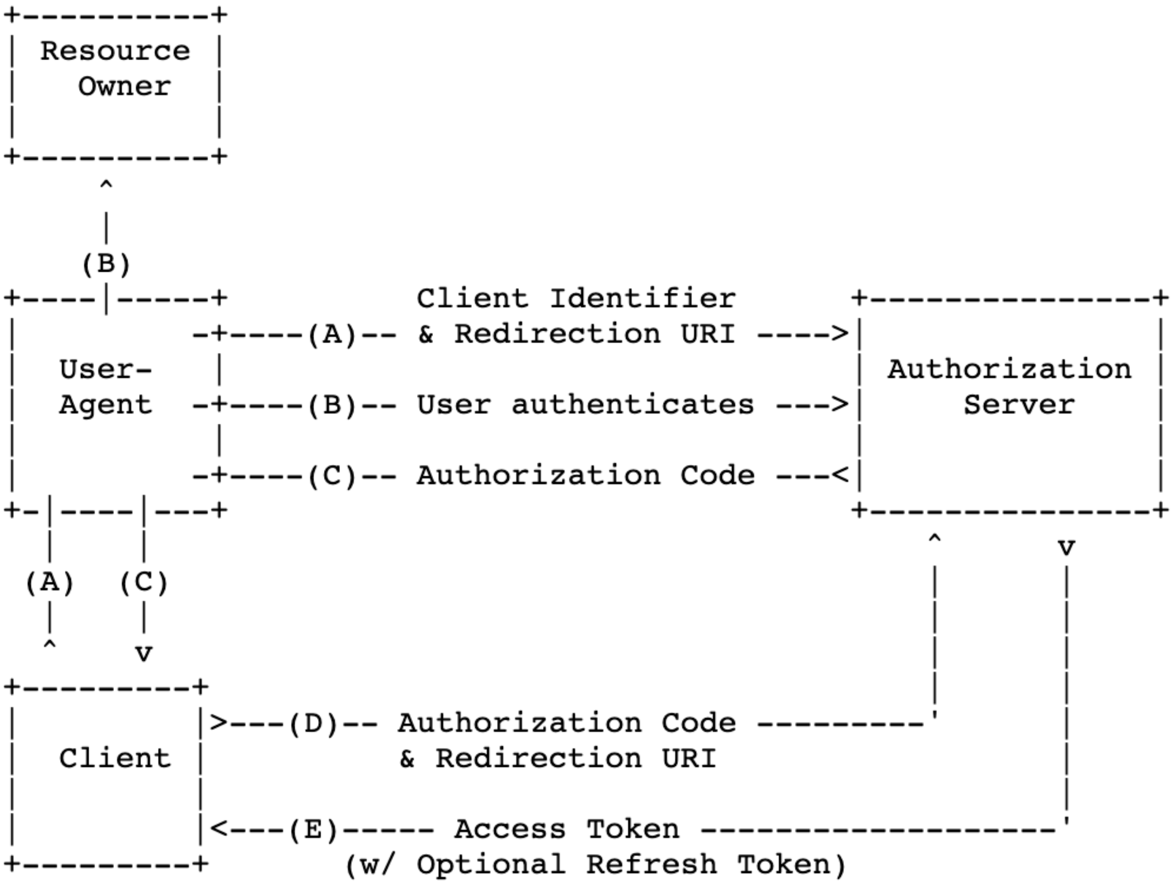
Another common type of identity federation is internal to the enterprise and increasingly found in SaaS offerings. Previously, enterprises would use "agents," which they would install on web servers hosting applications. The agents would communicate with something called a policy server to determine what a user could do, if anything at all. That agent/policy server technology is old and not used as much in enterprises anymore.

Instead, a popular protocol that is increasingly being used is OpenID Connect. OpenID Connect is newer than SAML and based on the OAuth2.0 protocol; most in-house enterprise apps are based on APIs and microservices, which is why OIDC is favored.[vi] It should be noted that some SaaS applications do support OpenID Connect.

OpenID Connect uses the authorization_code grant type of OAuth2.0. It is important to note that OpenID Connect is meant to share user attributes, so it will be the only part of OAuth2.0 in this document. There are many other grant_types in OAuth2.0 which authenticate users or clients in different ways but are not part of user *authentication* and *authorization* and are outside the scope of this document.

Authorization_Code Flow

The authorization_code grant type is explained in the OAuth2.0 spec.[vii] OpenID Connect 1.0 is based on this flow. An important consideration to note involves the scopes in OpenID Connect: they must contain openid (and most often include profile). Here is a diagram of that authorization_code flow:[viii]



```
+----------+
| Resource |
|   Owner  |
|          |
+----------+
     ^
     |
    (B)
+----|-----+          Client Identifier      +---------------+
|         -+----(A)-- & Redirection URI ---->|               |
|  User-   |                                 | Authorization |
|  Agent  -+----(B)-- User authenticates --->|     Server    |
|          |                                 |               |
|         -+----(C)-- Authorization Code ---<|               |
+-|----|---+                                 +---------------+
  |    |                                         ^      v
 (A)  (C)                                        |      |
  |    |                                         |      |
  ^    v                                         |      |
+---------+                                      |      |
|         |>---(D)-- Authorization Code ---------'      |
|  Client |          & Redirection URI                  |
|         |                                             |
|         |<---(E)----- Access Token -------------------'
+---------+       (w/ Optional Refresh Token)
```

Figure 3 - OAuth 2.0 authorization_grant Flow

In this diagram, we can see that the user will first go to a browser (user agent) and initiate a request against the authorization server. The authorization server will then prompt the user to enter their credentials (B). After collecting the credentials, the browser will send that information to the authorization server, which then will respond with a code to the browser (C). The backend of the application (Client, C) will take that code and exchange it for an access token (D, E). In OpenID Connect, there is an optional step F in which the client may request additional information about the user (attributes) by making an API request against a 'userinfo' endpoint. With this API request, the AS will return the user's information allowing the client to *authorize* the user.

To see the API calls, please see Appendix Item 3.

Challenges in Enterprise Federations

When to Use SAML versus OpenID Connect

The short answer to this question is: it depends. Sometimes there are limitations as to what SPs can do, as well as IdPs. There are pros and cons to both integrations, so it really is just a matter of choice (or limitation) between the IdPs and SPs.[ix]

The IDPro Body of Knowledge article "Introduction to Identity - Part 2: Access Management" by Pamela Dingle offers an interesting view of the evolution of authentication and access control tools.[x] In particular, the section 'Mobile & API Innovation Gave Us OAuth & Delegated Authorization Frameworks' offers some interesting insights into the evolution that led to the development of OAuth despite the existence of SAML.

Attributes - Data and Formatting

Applications require different names for attributes. Sometimes an attribute must be called firstname, where other applications may need firstName, or perhaps even givenName. This can cause issues, as the application might not be able to pick up the attribute in the SAML Assertion / userinfo endpoint it needs to authorize the user. This is where the IdP and SP need to collaborate to determine how the attributes should be sent. In some enterprises, the attribute names do not change; the enterprise forces the application to adopt its formatting of the attribute. Other times, the application forces the IdP to change the attributes. There is also something called attribute mapping which can take place. Most SAML and OpenID Connect plugins allow this to take place in attribute mapping files, like Shibboleth.[xi] The IdP will send attributes, and upon receiving them, the SP can transform them into the correct format.

Assertion Sizing

Quite a bit of information can be passed to SPs, and the assertion can become so large that it will break the SP. This is somewhat common when applications authorize users via Active Directory or LDAP groups (also known as SID bloat, essentially a large data blob of information about the user), and the IdP sends an array of all Active Directory groups. The SAML assertion will contain so much information that the SP will not be able to parse it out, and the user will not be able to get into the application. Resolving this issue often requires custom integrations, where there needs to be a special configuration within the IdP to manage assertions for that single application. Additionally, assertion sizes can be limited based on web servers, browsers, and even proxies. This problem can be alleviated via identity governance processes that limit the number of Active Directory groups and removes memberships no longer required.

Cross-Origin Resource Sharing (CORS)

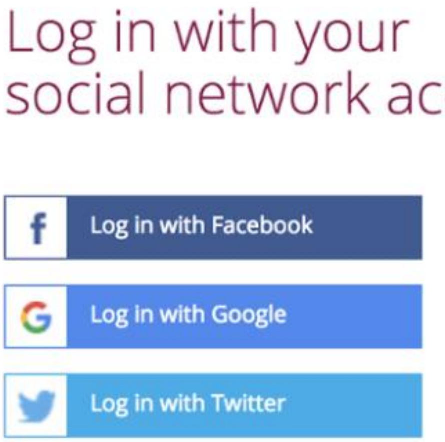
Cross-Origin Resource Sharing, commonly known as CORS, causes issues in many enterprises. CORS is a standard that allows a server to relax the same-origin policy.[xii] Usually, an API call from one application cannot be returned to a separate application. For example, if I make a request to application1.com/api, I would expect the request to come back to me and not be sent to application2.com/api. These are two different domains and application1.com could potentially be sending malicious data to application2.com.

CORS is used to explicitly allow some cross-origin requests while rejecting others. For example, if a site offers an embeddable service, it may be necessary to relax certain restrictions. If I attempt to load application1.com, and that application requires resources from application2.com, my browser will make that request through application1.com into application2.com, thus making it a cross-domain API call. CORS allows the request to pass through and retrieve information so I can visit the application.

Setting up such a CORS configuration is a challenge. It is also potentially not secure. What most IdPs can do is relax their policies to allow sharing between top-level domains, for example, *.enterprise.com or *.partner.com. This way, there will be no restrictions on the origin of requests.[xiii]

Conclusion

This document is a high-level review of application federations in the enterprise. The most common protocols used are SAML and OpenID Connect. Both are widely used today in the enterprise world as well as the consumer world as well. When you see this screen:



Log in with your
social network accounts

Figure 4 - Sample Social Login Screen

you are actually selecting the IdP you'd like to sign into the SP with. You also have the ability (in most cases) to sign up in the app directly. One thing to note, when you do sign in to an application using an Identity Provider such as social media sites, you are passing information about yourself, the same way your enterprise passes information about you to SPs in the enterprise. On social networks, it is important to understand the terms and conditions of what can be done with this data. In enterprise applications, this is usually done by legal teams to ensure there will be no data exfiltration.

With more and more applications becoming SaaS applications, enterprises are creating more and more federations. With that, there will continue to be innovations in the single sign-on community to make them safer, such as adding multifactor authentication into the flow.

Author Bio

My name is Patrick Lunney. I have managed/owned identity providers in two fortune 50 companies over the past eight years. In that time, I've worked with 100s of SaaS applications as well as in-house applications to ensure federations are set up securely and properly. Currently, I am the product owner for Capital One's internal workforce Single Sign-On and Multi Factor Authentication products. All applications in our enterprise must use either OpenID Connect or SAML for SSO, with very few exceptions. I have held this role since July of 2019.

Change Log

Date	Change
2022-06-03	V2 published; Changed title, updated OIDC definition, added detail re: SP-initiated flows
2021-04-19	V1 published

Appendix:

Item 1: SAML Request

```
<md:EntityDescriptor xmlns:md="urn:oasis:names:tc:SAML:2.0:metadata" ID="mzWO1kVu-
dAmFIdmN.08s9bOaCH" cacheDuration="PT1440M" entityID="IdProvider">
  <md:IdPSSODescriptor protocolSupportEnumeration="urn:oasis:names:tc:SAML:2.0:protocol"
  WantAuthnRequestsSigned="false">
    <md:KeyDescriptor use="signing">
      <ds:KeyInfo xmlns:ds="http://www.w3.org/2000/09/xmldsig#">
        <ds:X509Data>
          <ds:X509Certificate>
          </ds:X509Certificate>
        </ds:X509Data>
      </ds:KeyInfo>
    </md:KeyDescriptor>
    <md:NameIDFormat>urn:oasis:names:tc:SAML:1.1:nameid-format:unspecified</md:NameIDFormat>
    <md:SingleSignOnService Binding="urn:oasis:names:tc:SAML:2.0:bindings:HTTP-POST"
    Location="https://authn.enterprise.com/idp/SSO.saml2"/>
    <md:SingleSignOnService Binding="urn:oasis:names:tc:SAML:2.0:bindings:HTTP-Redirect"
    Location="https://authn.enterprise.com/idp/SSO.saml2"/>
    <saml:Attribute xmlns:saml="urn:oasis:names:tc:SAML:2.0:assertion" Name="firstname"
    NameFormat="urn:oasis:names:tc:SAML:2.0:attrname-format:unspecified"/>
    <saml:Attribute xmlns:saml="urn:oasis:names:tc:SAML:2.0:assertion" Name="groups"
    NameFormat="urn:oasis:names:tc:SAML:2.0:attrname-format:unspecified"/>
    <saml:Attribute xmlns:saml="urn:oasis:names:tc:SAML:2.0:assertion" Name="lastname"
    NameFormat="urn:oasis:names:tc:SAML:2.0:attrname-format:unspecified"/>
    <saml:Attribute xmlns:saml="urn:oasis:names:tc:SAML:2.0:assertion" Name="userid"
    NameFormat="urn:oasis:names:tc:SAML:2.0:attrname-format:unspecified"/>
    <saml:Attribute xmlns:saml="urn:oasis:names:tc:SAML:2.0:assertion" Name="email"
    NameFormat="urn:oasis:names:tc:SAML:2.0:attrname-format:unspecified"/>
  </md:IdPSSODescriptor>
  <md:ContactPerson contactType="administrative"/>
</md:EntityDescriptor>
```

Item 2: SAML Response

```
<samlp:Response Destination="https://serviceprovider.com/acs"
  ID="HpiyLr_zVMK.jxdUHXxRvjJ8Fwy" IssueInstant="2020-11-24T01:53:06.809Z" Version="2.0"
  xmlns:samlp="urn:oasis:names:tc:SAML:2.0:protocol">
  <saml:Issuer xmlns:saml="urn:oasis:names:tc:SAML:2.0:assertion">IDprovider</saml:Issuer>
  <ds:Signature xmlns:ds="http://www.w3.org/2000/09/xmldsig#">
    <ds:SignedInfo>
      <ds:CanonicalizationMethod Algorithm="http://www.w3.org/2001/10/xml-exc-c14n#"/>
      <ds:SignatureMethod Algorithm="http://www.w3.org/2001/04/xmldsig-more#rsa-sha256"/>
      <ds:Reference URI="#HpiyLr_zVMK.jxdUHXxRvjJ8Fwy">
        <ds:Transforms>
          <ds:Transform Algorithm="http://www.w3.org/2000/09/xmldsig#enveloped-signature"/>
          <ds:Transform Algorithm="http://www.w3.org/2001/10/xml-exc-c14n#"/>
        </ds:Transforms>
        <ds:DigestMethod Algorithm="http://www.w3.org/2001/04/xmlenc#sha256"/>
        <ds:DigestValue>PwJICHFA1QIlML2p5MyJaRib5TDY4TWj5J7IEAjn1Yo=</ds:DigestValue>
      </ds:Reference>
    </ds:SignedInfo>
    <ds:SignatureValue> Signature
    </ds:SignatureValue>
    <ds:KeyInfo>
      <ds:X509Data>
        <ds:X509Certificate>
</ds:X509Certificate>
</ds:X509Data>
      <ds:KeyValue>
        <ds:RSAKeyValue>
                 <ds:Modulus>
                 </ds:Modulus>
          <ds:Exponent>AQAB</ds:Exponent>
        </ds:RSAKeyValue>
      </ds:KeyValue>
    </ds:KeyInfo>
</ds:Signature>
<samlp:Status><samlp:StatusCode Value="urn:oasis:names:tc:SAML:2.0:status:Success"/>
</samlp:Status>
<saml:Assertion ID="bJUFiJZEXV0rDgdTh9HnF2CbrIq" IssueInstant="2020-11-24T01:53:07.104Z"
    Version="2.0" xmlns:saml="urn:oasis:names:tc:SAML:2.0:assertion">
    <saml:Issuer>IDprovider</saml:Issuer>
    <saml:Subject>
      <saml:NameID Format="urn:oasis:names:tc:SAML:2.0:nameid-format:entity">ztl593</saml:NameID>
      <saml:SubjectConfirmation
Method="urn:oasis:names:tc:SAML:2.0:cm:bearer"><saml:SubjectConfirmationData NotOnOrAfter="2020-11-
24T01:58:07.104Z"
        Recipient="https://serviceprovider.com/acs"/></saml:SubjectConfirmation>
    </saml:Subject>
    <saml:Conditions NotBefore="2020-11-24T01:48:07.104Z" NotOnOrAfter="2020-11-24T01:58:07.104Z">
      <saml:AudienceRestriction>
        <saml:Audience>http://www.serviceprovider.com/</saml:Audience>
      </saml:AudienceRestriction>
    </saml:Conditions>
    <saml:AuthnStatement AuthnInstant="2020-11-24T01:53:07.103Z"
      SessionIndex="bJUFiJZEXV0rDgdTh9HnF2CbrIq">
      <saml:AuthnContext>

<saml:AuthnContextClassRef>urn:oasis:names:tc:SAML:2.0:ac:classes:Telephony</saml:AuthnContextClassRef>
      </saml:AuthnContext>
    </saml:AuthnStatement>
    <saml:AttributeStatement>
      <saml:Attribute Name="mail" NameFormat="urn:oasis:names:tc:SAML:2.0:attrname-format:basic">
        <saml:AttributeValue xmlns:xs="http://www.w3.org/2001/XMLSchema"
          xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:type="xs:string">Patrick.Lunney@idprovider.com</saml:AttributeValue>
      </saml:Attribute>
    </saml:AttributeStatement>
  </saml:Assertion>
</samlp:Response>
```

Item 3: OpenID Connect

To begin the process the user agent will first make a GET request against the authorization server, passing along information about the application the user wishes to go to.

```
curl --request GET \
--header ‘content-type: application/x-www-form-urlencoded' \
 --url
“${sso_prefix}/authorization?response_type=code&redirect_uri=${redirect_uri}&scope="op
enid profile”&client_id=${client_id}
```

What will return from this request is the login page (assuming there is no session), and a user will enter their credentials so the authorization server can authenticate the user. Afterward, an authorization_code is sent to the application in the browser. The application backend must take that authorization_code and exchange it for an access token.

To exchange the authorization_code for the access token:

```
curl --request POST \
    --url “https://${sso_prefix}/token” \
    --header 'content-type: application/x-www-form-urlencoded' \
    --header 'Authorization: Basic base64(urlencode("${client_id}:${client_secret}))' \
    --data “code=${code}” \
    --data “grant_type=authorization_code” \
    --data “redirect_uri=${redirect_uri}” \
    --data 'scope=openid profile'
```

After this exchange, the application can then make a backend API call to the authorization server to obtain additional information about the user for further authorization.

```
curl --request GET \
--header ‘content-type: application/x-www-form-urlencoded' \
--header 'Authorization: Bearer ${token}
 --url “${sso_prefix}/userinfo
```

This will give applications information like this:

```
{
  "sub"        : "83692",
  "name"       : "Alice Adams",
  "email"      : "alice@example.com",
  "department" : "Engineering",
  "birthdate"  : "1975-12-31"
}
```

[i] “Multilateral federation,” InCommon Federation wiki, last updated 17 February 2020, https://spaces.at.internet2.edu/display/federation/Multilateral+federation.

[ii] Hodges, Jeff, Rob Philpott, Eve Maler, eds. “Glossary for the OASIS Security Assertion Markup Language (SAML) V2.0,” OASIS Standard, 15 March 2005, https://docs.oasis-open.org/security/saml/v2.0/saml-glossary-2.0-os.pdf.

[iii] Hardt, D., Ed., "The OAuth 2.0 Authorization Framework", RFC 6749, DOI 10.17487/RFC6749, October 2012, <https://www.rfc-editor.org/info/rfc6749>.

[iv] Ragouzis, Nick, John Hughes, Rob Philpott, Eve Maler, Paul Madsen, Tom Scavo, eds. “Security Assertion Markup Language (SAML) V2.0 Technical Overview,” OASIS Committee Draft, 25 March 2008, https://docs.oasis-open.org/security/saml/Post2.0/sstc-saml-tech-overview-2.0.pdf.

[v] OASIS Standards landing page, https://www.oasis-open.org/standards/.

[vi] Hardt, D., Ed., "The OAuth 2.0 Authorization Framework", RFC 6749, DOI 10.17487/RFC6749, October 2012, <https://www.rfc-editor.org/info/rfc6749>.

[vii] Ibid, see Section 4.1.

[viii] Hardt, D., Ed., "The OAuth 2.0 Authorization Framework", RFC 6749, Section 4.1, DOI 10.17487/RFC6749, October 2012, <https://www.rfc-editor.org/info/rfc6749>.

[ix] For further discussion on the pros and cons between SAML and OAuth, see https://www.okta.com/identity-101/whats-the-difference-between-OAuth-openid-connect-and-saml/ or https://auth0.com/intro-to-iam/saml-vs-openid-connect-oidc/

[x] Dingle, Pamela, “Introduction to Identity – Part 2: Access Management,” IDPro Body of Knowledge, 17 June 2020, https://bok.idpro.org/article/id/45/.

[xi] Shibboleth Consortium, https://www.shibboleth.net/.

[xii] “Same-origin Policy,” MDM Web Docs, https://developer.mozilla.org/en-US/docs/Web/Security/Same-origin_policy.

[xiii] For additional information, see https://developer.mozilla.org/en-US/docs/Web/HTTP/CORS and https://web.dev/cross-origin-resource-sharing/.

Designing MFA for Humans

By Nishant Kaushik

Table of Contents

Abstract

This article describes how to deploy a thoughtful, consumer-friendly multi-factor authentication (MFA) program that will allow the IAM practitioner to successfully deliver on both the security and usability needs of their authentication systems. The approach is based on a framework of six pillars: determining the viability of different forms of MFA, allowing a multimodal rollout of MFA options, encouraging adoption, supporting MFA across all services and access channels, designing support processes, and creating a trusted environment where MFA can offer additional security to both the consumer and the company.

Introduction - Designing MFA For Humans

If every year is The Year of PKI, then when exactly was The Year of Two-Factor Authentication? Was it 2012, when the epic hacking of Mat Honan highlighted just how vulnerable all of our digital lives are? [i],[ii],[iii],[iv] Was it 2014, when the even higher profile iCloud leaks of celebrity photos pushed various consumer services to rush offering two-factor authentication an option available to users?[v] Or did it really arrive in 2018, at least for financial institutions, when PSD2 delivered a regulation with some real teeth?[vi],[vii]

Terminology

- Adaptive Authentication - Adaptive authentication aims to determine and enforce the authentication level required at any time during a user session - when the session is commenced, during the session when access requirements force a re-evaluation, or when the session token expires. The factors to be used in achieving that authentication level are determined dynamically based on the access control policy governing the resources being accessed, and a variety of environmental conditions and risk factors in effect at that time for that user.
- Account Takeover - Account takeover is a form of identity theft and fraud, where a malicious third party successfully gains access to a user's account credentials.
- Continuous Authentication - Continuous authentication is a mechanism that uses a variety of signals and measurements to determine during a user session if there is any change in the confidence that it is still the same user that authenticated at the beginning of the session, and trigger an authentication action if there is a drop in confidence.
- PSD2 - PSD2 (the Revised Payment Services Directive, Directive (EU) 2015/2366) is an EU Directive, administered by the European Commission (Directorate General Internal Market) to regulate payment services and payment service providers throughout the European Union (EU) and European Economic Area (EEA). It contains many requirements specifically related to Strong Client Authentication.
- Social Engineering - Social engineering is a method of manipulating people so they give up confidential information, such as passwords or bank information, or grant access to their computer to secretly install malicious software.
- Step-Up Authentication - A method to increase the level of assurance (or confidence) the system has regarding a user's authentication by issuing one or more additional authentication challenges, usually using factors different from the one(s) used to establish the initial authenticated session. The need for increasing the level of assurance is typically driven by the risk associated with the sensitive resource the user is attempting to access.
- Threat Modeling - Threat modeling is an analysis technique used to help identify threats, attacks, vulnerabilities, and countermeasures that could impact an application or process.
- Two-Factor Authentication (2FA) - A specific case of Multi-Factor Authentication (see: IDPro's Consolidated Terminology) where two factors must be checked to validate a user's identity.

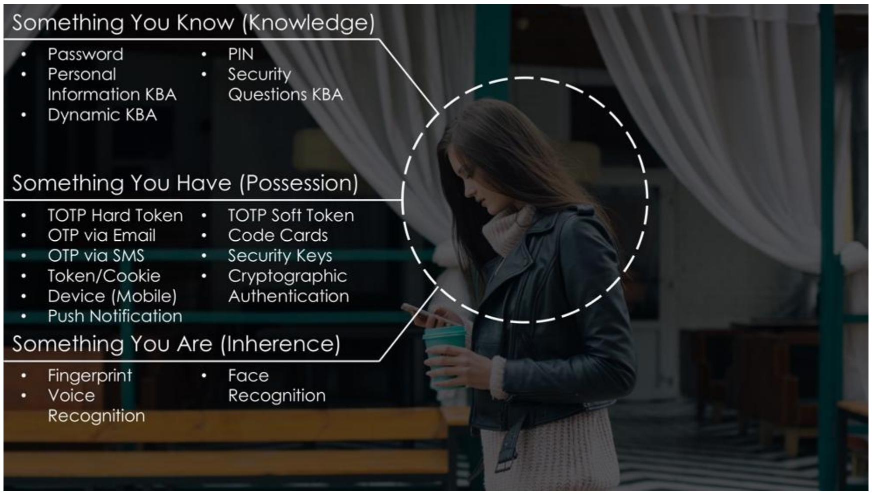
The Struggle is Real

Two-factor authentication (2FA) is not new. IAM practitioners are certainly familiar with it through their professional lives (remember keychains full of hardware tokens?), but organizations still struggle with rolling out 2FA to customers. Why?

Figure 1: Remember carrying these?

The simple reason is that while employees are a captive audience that will submit to whatever painful, inconvenient mechanism they are forced to adopt (ok, except for mobile device management on their personal phones), customers are a different story. The customer experience matters, and if it is not done well, people are either not going to enable it (when it is optional), will work their way around it, or decide not to engage at all.

For any organization starting down the path of implementing 2FA, it can be confusing and challenging. They find an extensive list of factors spread across the "something you ___" categories, but little guidance on how to put a good 2FA scheme in place. It's like getting all the parts in a model kit, but without the instruction manual.



Something You Know (Knowledge)

- Password
- Personal Information KBA
- Dynamic KBA
- PIN
- Security Questions KBA

Something You Have (Possession)

- TOTP Hard Token
- OTP via Email
- OTP via SMS
- Token/Cookie
- Device (Mobile)
- Push Notification
- TOTP Soft Token
- Code Cards
- Security Keys
- Cryptographic Authentication

Something You Are (Inherence)

- Fingerprint
- Voice Recognition
- Face Recognition

Figure 2: A vast menu to choose from

Most organizations simply end up taking the approach of picking an additional factor that they can simply tack on to the end of their password authentication step, and then call it a day. Unfortunately, that simplified approach falls far short of successfully addressing the problem, resulting in continued breach vectors, brittle infrastructure, and unsatisfied customers.

Thinking in Factors

Multi-factor authentication (MFA) aside, the goal of any authentication framework is to validate that the returning person (or thing) is the same one that the system saw last time, *to the required level of assurance*. That last part is what makes authentication difficult to implement well. Measuring the assurance of authentication is subjective, and cannot be normalized across organizations, industries, or end-user communities since a critical element in evaluating the assurance of authentication is trying to determine how easy or likely it is for an adversarial party to get around it. Determining this correctly requires doing threat modeling and risk analysis (more on that later), and then translating this into how to authenticate in different contexts.

This requirement for assurance is where factors of authentication become relevant. Factors make the abstract concrete by giving the authentication framework something tangible to evaluate, invoke, and measure. An important evolution that has happened is the realization that not all factors are created equal. This realization has expanded the kind of factors that can be used, while also creating the understanding that the same level of authentication assurance can be achieved using different sets of factors that are not numerically the same (i.e., one set of 2 factors can achieve the same level of assurance as a different set of 4 factors). The requirements around assurance are helping drive the discussion away from 2FA and towards **MFA**.

One important consideration that overshadows all of this is the nature of the factors and their impact on the user experience. When authentication factors translate into explicit challenges (or "active" factors) that an end-user has to engage with (as opposed to "passive" factors that work silently in the background), then the impact on usability will drive organizations to try and reduce the number of factors used in the authentication process. One way that they can compensate is to invoke additional (active) factors when the risk associated with the access request is elevated (often called **step-up authentication** or **adaptive authentication**). An even more refined approach to evaluating authentication assurance and risk is **continuous authentication**, which recognizes that the assurance level degrades over the life of the user session given how identity or access information may change during the session, and that passive factors can be used to constantly measure any changes or degradation to that assurance level, and determine if step-up authentication is required to bump the assurance level back up.

In all of these approaches, the factors of authentication are the control vector that allows the authentication framework to measure, achieve, and maintain the assurance level of the authenticated session as required by the business.

A Framework for Designing Your MFA Schema

MFA has become an imperative across industries, user bases and threat models, and the challenges and practices described below apply equally to both small and large organizations. It lays out a basic framework to build an MFA program that should prove useful to product teams, employees, and clients. This framework is built on six pillars that address the challenge of balancing security, usability and privacy.

1. Viability

The first pillar of that framework is **Viability**. When going through the long list of factors possible, implementors and decision makers must assess which of those factors is viable for their MFA scheme. Assessing viability has multiple considerations:

- User Acceptance: Think of the people that make up your user base, and what factors they'd be willing to accept and use.
- Cost: Think about the cost of the factor, and whether that is a cost that the business will bear, or the customer will bear. Hardware tokens are great, but expensive. Is the business buying it for their customers, or are they expecting the customer to buy it themselves?
- Threat Model: Consider the threat model associated with the factor. A USB device can be a very secure authentication factor, where the user has to plug the key into a port on their desktop in order to authenticate. But research studies have shown that people will often leave them plugged into their desktop even when they leave the office, virtually negating its assurance as a possession factor. Discussing this in detail is out of scope for this article, but do note that there is a need to introduce threat modeling as a core discipline in identity management.
- Effectiveness: Consider the effectiveness of the factor. For example: security questions, a widely deployed form of MFA, are universally acknowledged as being ineffective in this age of public social media profiles and social engineering threats.
- Regulatory Compliance: In many cases, regulatory compliance can enter the equation, since regulators are increasingly rendering opinions on which factors are acceptable for different industries.

2. Multimodal

The second pillar of the framework is **Multimodal**. When implementing 2FA, the goal is to have each user employ at least two factors when authenticating. However, that does not mean that the business should only support two factors. Not all factors work for all users, and when a business is trying to increase the number of customers turning on MFA, they must offer options (i.e., be multimodal) that work with their vast and diverse user base. The idea that they can find two factors that work for everyone leads them to a least common denominator approach, and that's how so many industries have ended up with SMS OTP as the de facto "standard" in MFA, and a weakening of the security model.[viii]

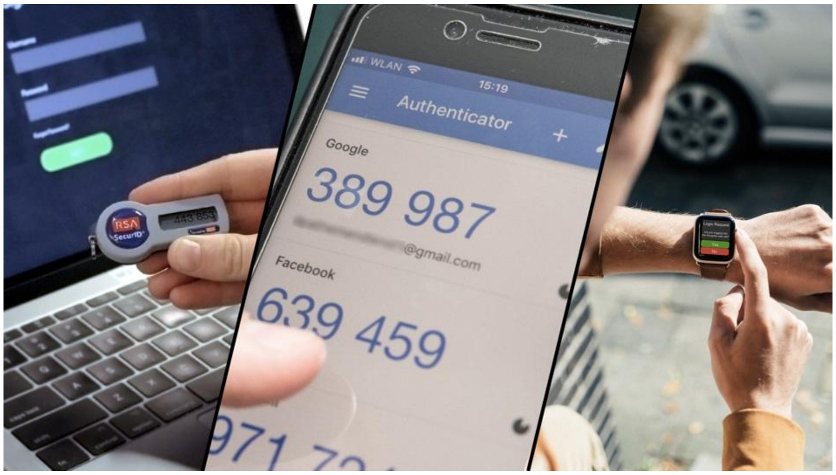


Figure 3: Different strokes for different folks

Offering choice allows a business to address the varying capabilities, preferences, and circumstances of their end-users, and avoid a "one size fits all" approach that alienates customers and often weakens security.

Being multimodal will necessarily require the authentication platform become adaptive, not just to risks, but also to user (cap)abilities. This ability to adapt will require the authentication service/platform/provider to create intelligent user flows – a concept commonly being referred to as orchestration.[ix]

3. Adoption

The third pillar is the one that is the most misunderstood - **Adoption**. The reality is that unlike enterprise environments where the business can mandate MFA, the customer environment requires a business to convince their end-users to start using MFA. While acceptance for this pattern is growing, in general this is easier said than done.[x]

Organizations need to make UX research a core element of their IAM program, especially as they design their MFA scheme. It is a critical and foundational element to creating the right set of messaging, training, and incentive components that the business will have to incorporate into their rollout plan to drive adoption.[xi]

4. Omnichannel

An overlooked pillar when designing MFA is **Omnichannel**. Businesses have often failed to recognize that MFA should not apply just to their web or mobile channels; they must be deployed across all their customer-facing channels. This ties back to the pillar on multimodality. Businesses are engaging with customers and partners across many channels – web, mobile, call center, in-person, chat, smart home assistants, and more - and each channel usually brings a completely different way of authenticating the end-user.

This inconsistency frustrates end-users, creates a headache for customer-facing staff and IT staff, and delights bad actors. Attackers look for the weakest link across those channels, and go after that one, exploiting not only the weakness of the channel but also the frustration that customers and employees feel. The result is rampant account takeover attacks and fraud. Watch this video of a classic social engineering attack that exploits weaknesses in the customer service authentication process to take over an account.

Businesses today have a pressing imperative to transition away from an inconsistent hodge-podge of varying authentication models and bring some consistency and equality of security levels across their various channels.

5. Processes

The fifth pillar of the framework is the one that most organizations do not pay enough attention to: **Processes**. Enabling and maintaining MFA for individual customers involves many different processes, each of which needs to be properly designed:

- **Enrollment**: If the enrollment process is flawed, the assurance of your MFA is suspect from the very beginning. Many organizations will allow users to set up their second factor after they have authenticated solely using their first, and that is a massive vulnerability point in the overall security scheme.
- **Backup / Alternate**: No authentication factor is immune from loss or destruction, so the business has to think about ways to not only allow, but proactively encourage, customers to set up additional authenticators as backups. And those backups must have the same strength as the primary; otherwise, this creates a backdoor for attackers.
- **Escape Paths**: Not all authentication factors are always available for use, and the alternate mechanism may not be available. Consider what happens to push notification-based authentication for someone working in a part of the building, or on a plane, where they get no signal. It is not out of the question that they left their FIDO security key that they use as a backup safely locked in their office drawer. Locking them out under those circumstances can prove to be hugely problematic and result in workarounds that open up exploitation vectors. Escape paths may not be appropriate in all circumstances; consider carefully whether they should be designed into your system.
- **Recovery**: Consider how the business will support an end-user that has lost their authentication factor(s), so that they are not faced with the dire consequence of being permanently locked out (think of all the horror stories of bitcoin wallets irrecoverably locked up because their owner lost the hardware token containing their private key). Recovery paths must also be designed properly to avoid having them turn into backdoors for bad actors. *Never* use an authentication factor as the verification factor for also doing recovery (e.g., every service that uses SMS OTP as a second factor of authentication, and also as a way of resetting a forgotten password). This effectively creates a backdoor that turns a 2FA scheme into a one-factor authentication scheme.
- **Deprovisioning**: Of course, the business must to consider how to invalidate a factor that is no longer available to the customer, or is no longer acceptable to the business because of vulnerabilities or issues discovered in it (whether it be at an individual level or system wide).

Importantly, escape paths and recovery flows need to be treated as *exceptions* with higher risks associated with them. That implies increasing the risk evaluation and security of those flows, which often means adding friction. It is important to remember that in these circumstances, customers will frequently be understanding of the increased scrutiny in those paths (provided the business offers adequate explanations). One of the techniques emerging for these exception flows is the use of identity verification tools (e.g., document-based identity proofing) in these scenarios.

6. Trusted Environment

The sixth and final pillar of the framework is establishing a **Trusted Environment** within which to execute MFA. It will not matter how good or strong a business's factors of authentication are if the environment within which those factors are being accepted, stored, transmitted, and evaluated is compromised, allowing them to be stolen, manipulated, or replayed. Keyloggers that capture secrets, malware apps that intercept SMS codes or steal keys, malicious WiFi, reverse proxies, and rogue cell towers that capture and replay credentials or tokens – threats like these reduce the effectiveness of MFA and degrade organizational trust in those factors. All multi-factor authentication projects must be part of a larger security program that enforces defense-in-depth (or, to use the industry term du jour, **zero-trust security**) to not only leverage the factors of authentication, but also look at the health of the devices and hardware being used and the networks being relied upon, as well as other signals of risk, in order to build trust in (hopefully) the simple act of authenticating your customer.

Conclusion

This framework offers guidance for rolling out a strong and usable MFA service for their users. The considerations of factor viability, multimodal support, adoption rates, omnichannel applicability, and the infrastructure that guarantees a trusted environment, applies to any organization in any sector, and should be considered at every stage -- designing, building, and rollout – of any MFA program.

May all your authentications be strong, and all your customers be happy, engaged, and protected.

[This article is adapted from my talks at EIC, Identiverse, and Identity Week. You can watch the Identiverse talk here.]

Author Bio

Nishant Kaushik is the CTO of Uniken, the first security platform that tightly integrates identity, authentication and channel security. He brings over 15 years of experience in the identity management industry architecting and delivering market leading products, with stints at Thor Technologies, Oracle, SCUID and CA Technologies. His current role allows him to focus on his latest passion of solving the user experience problem in delivering exceptional security by leveraging identity. Nishant is a recognized thought leader and notorious photoshopper of the identirati, regularly speaking at conferences and provoking discussion through his blog (blog.talkingidentity.com) and on twitter (@NishantK).

Nishant Kaushik
CTO, Uniken Inc.

[i] Berinato, Scott, "Only Mostly Dead," CSO Online, 23 May 2002, https://www.csoonline.com/article/2113027/only-mostly-dead.html.

[ii] Stiennon, Richard, "The new Entrust: is 2011 the year of PKI?" Forbes, 9 May 2011, https://www.forbes.com/sites/richardstiennon/2011/05/09/the-new-entrust-is-2011-the-year-of-pki/#2bdb8f5a171e.

[iii] Hils, Adam, "2015 Network Security Predictions: 8 Things That Won't Happen," blog, Gartner, 29 December 2014, https://blogs.gartner.com/adam-hils/2015-8-network-security-trends-that-wont-gain-t-raction/.

[iv] Honan, Matt, "How Apple and Amazon Security Flas Led to My Epic Hacking," Wired, 6 August 2012, https://www.wired.com/2012/08/apple-amazon-mat-honan-hacking/.

[v] Wikipedia contributors, "ICloud leaks of celebrity photos," *Wikipedia, The Free Encyclopedia*, https://en.wikipedia.org/w/index.php?title=ICloud_leaks_of_celebrity_photos&oldid=974755004 (accessed September 21, 2020).

[vi] *Directive 2015/2366/EU of the European Parliament and of the Council of 25 November 2015 on payment services in the internal market, amending Directives 2002/65/EC, 2009/110/EC and 2013/36/EU and Regulation (EU) No 1093/2010, and repealing Directive 2007/64/EC*, European Commission, 12 January 2016, https://ec.europa.eu/info/law/payment-services-psd-2-directive-eu-2015-2366/law-details_en.

[vii] Constantin, Lucian, "What is PSD2? And how will it impact the payments processing industry," CSO Online, 13 September 2019, https://www.csoonline.com/article/3390538/what-is-psd2-and-how-it-will-impact-the-payments-processing-industry.html.

[viii] Suau, Roxanne, "SMS OTP Authentication: Not As Safe As You May Think," blog, Pradeo, 17 February 2020, https://blog.pradeo.com/sms-otp-authentication-not-safe.

[ix] Goldberg, Joel, "Workflow Orchestration: An Introduction," DevOps Blog, BMC, 15 October 2019, https://www.bmc.com/blogs/workflow-orchestration/.

[x] Camp, L. Jean, Sanchari Das, "Studies of 2FA, Why Johnny Can't Use 2FA and How We Can Change That," RSA Conference2019, video session, 12 February 2020, https://www.youtube.com/watch?v=UH9yWvvp4k8.

[xi] Das, Sianchari, Andrew Dingman, L. Jean Camp, "Why Johnny Doesn't Use Two Factor: A Two-Phase Usability Study of the FIDO U2F Security Key,"in *2018 International Conference on Financial Cryptography and Data Security (FC),* 2018, https://fc18.ifca.ai/preproceedings/111.pdf.

Multi-factor Authentication

By Dean H. Saxe (Amazon) and Khaled Zaky (Amazon)

Table of Contents

Abstract

Multi-factor authentication (MFA) is critical in securing account access and guarding against account takeover. In this article, we explain the core concepts that define MFA, explore the characteristics of different MFA types, and discuss the various threats mitigated by using MFA.

Introduction

This article describes multi-factor authentication (MFA), a key component in securing account access and guarding against account takeover. Organizations and individuals typically have multiple types of MFA and several strategies for implementing its use. Not all MFA offers the same level of security, and some types of MFA are generally not recommended.

Terminology

Many of these terms have been sourced from "Terminology in the IDPro Body of Knowledge." [i]

Term	Definition
Authentication	Authentication is the process of proving that the user with a digital identity who is requesting access is the rightful owner of that identity. Depending on the use case, an 'identity' may represent a human or a non-human entity; may be either individual or organizational; and may be verified in the real world to varying degrees, including not at all.
Authorization	Determining a user's rights to access functionality with a computer application and the level at which that access should be granted. In most cases, an 'authority' defines and grants access, but in some cases, access is granted because of inherent rights (like patient access to their own medical data). Authorization is evaluating what access or rights an identity should have in an environment.
Identity and Access Management	Identity and Access Management (IAM) is the discipline used to ensure the correct access is defined for the correct users to the correct resources for the correct reasons.
Identity Provider	An Identity Provider (IdP) performs a service that sends information about a user to an application. This information is typically held in a user store, so an identity provider will often take that information and transform it to be able to be passed to the service providers, AKA apps. The OASIS organization, responsible for the SAML specifications, defines an IdP as "A kind of SP that creates, maintains, and manages identity information for principals and provides principal authentication to other SPs within a federation, such as with web browser profiles."
Multi-Factor Authentication (MFA)	An approach whereby a user's identity is validated to the trust level required according to a security policy for a resource being accessed using more than one factor (something you know (e.g., password), something you have (e.g., smartphone), something you are (e.g., fingerprint).
MFA Prompt Bombing	Also known as MFA fatigue, MFA prompt bombing is a cyber-attack technique that describes when an attacker bombards a user with mobile-based push notifications, which sometimes leads to the user to

	approve the request out of annoyance which might lead to an account takeover.

What is Multi-factor Authentication?

MFA is an authentication mechanism that requires a user logging into an application or an online account to present two or more factors to sign in and complete their authentication flow. Traditionally this would have been just a username and a password combination or another form of single-factor authentication, such as fingerprint biometrics. Adding multiple factors reduces the likelihood of bad actors gaining unauthorized access in case any of the factors are compromised. For example, single factors, such as passwords (which are subject to reuse and compromise), are one of the most common ways malicious actors can gain unauthorized access to your accounts, data, and online assets. Adding additional factors reduces the risk of account compromise and raises authentication assurance. Check out the NIST 800-63-B, which provides recommendations on types of authentication processes, authenticator types, and various assurance levels.[ii]

There are three types of MFA factors:

- The knowledge factor is something you know. This factor could be something like a password or a PIN code.
- The possession factor is something you have. This factor could be something like a USB key, a smartphone, or an access card.
- The inherence factor is something you are. This factor could be a biometric, like facial recognition, fingerprint, or voice recognition.

What is the Difference between MFA and 2FA?

Two-factor authentication, or 2FA, is an identity and access management authentication method that requires exactly two factors of identification to gain access. It is worth mentioning that 2FA is sometimes referred to as two-step verification or 2SV in some online services. 2FA is usually used interchangeably with MFA. However, in the case of MFA, more than two factors can be required, such as a combination of password + one-time password (OTP) + on a device with mobile device management (MDM). Therefore, 2FA is a subset of MFA.

History of Multi-factor Authentication

How did the industry come to embrace MFA? Although the original ideas and patents are up for debate, we can say that the concept of MFA was first commonly used with automated teller machines (ATMs, cash machines). First introduced in Europe in 1967, ATMs required a physical card containing information encoded on the magnetic stripe as the possession factor (something I have) and a PIN (something I know) to conduct bank transactions.[iii]

> *The breakthrough in security was the idea that a public number (PAN) was to be combined with a private identification number (PIN). The PAN was printed in punched holes on the card and of course, could be forged. It would be secured through the use of a PIN that would correspond to the PAN through a complex coding system. The key was that such system should be of sufficient strength to prevent anyone getting to the PIN from the PAN. Chubb tested the system by printing off 1001 cards and attempting to break this system. They failed and Goodfellow's system became the basis of the security system in the 'Chubb MD2' cash dispenser. Goodfellow's patent was filed on May 2, 1966 (GB1197183).*[iv]

In 1987, RSA introduced the first hardware key fob, enabling the use of one-time passwords (OTPs) as an authentication factor. These hardware key fobs are still in use today and sold by numerous vendors using both Time-based One Time Passwords (TOTP, see RFC6238)[v] and HMAC-Based One Time Passwords (HOTP, see RFC4226).[vi]

By the early 2000s, MFA solutions began to see a broad rollout in enterprise, government, and consumer use cases. In 2004 the United States Homeland Security Program Directive 12 (HSPD-12) was signed by President George W. Bush.

> *"US policy is to enhance security, increase Government efficiency, reduce identity fraud, and protect personal privacy by establishing a mandatory, Government-wide standard for secure and reliable forms of identification issued by the Federal Government to its employees and contractors (including contractor employees). This directive mandates a federal standard for secure and reliable forms of identification."* [vii]

In response to HSPD-12, the US Federal Government, through the National Institute of Standards and Technology (NIST), released FIPS-201-1, specifying the requirements for Personal Identity Verification (PIV) for US Federal Government employees and contractors.[viii] NIST Special Publication 800-73-1[xi], released in March 2006, "specifies the

PIV data model, Application Programming Interface (API), and card interface requirements necessary [...] for interoperability across deployments or agencies. Interoperability is defined as the use of PIV identity credentials such that client-application programs, compliant card applications, and compliant integrated circuit cards (ICC) can be used interchangeably by all information processing systems across Federal agencies." [ix]

In December 2004, the US Federal Deposit Insurance Corporation (FDIC) released the paper "Putting an End to Account-Hijacking Identity Theft," which concluded with the recommendation for "upgrading existing password-based single-factor customer authentication systems to two-factor authentication."[x] Shortly after that, in 2005, the Federal Financial Institutions Examination Council released guidance for the US banking industry entitled "Authentication in an Internet Banking Environment," which stated, "The agencies consider single-factor authentication, as the only control mechanism, to be inadequate for high-risk transactions involving access to customer information or the movement of funds to other parties. Financial institutions offering Internet-based products and services to their customers should use effective methods to authenticate the identity of customers using those products and services."[xi] The requirements were not compulsory. In 2011, the RAND Corporation noted:

> *The financial sector is potentially the most varied in its implementation practices. Despite regulations (more like "guidelines") that require financial institutions to protect certain data to a certain minimum level and indicate that MFA meets these criteria, organizations in this sector make network access decisions internally.*[xii]

These changes did not arrive without debates about their value and consumer concerns about using MFA.[xiii] First deployed in Nigeria in 2005 by Neticash, SMS OTP was broadly adopted in the 2010s. At the same time, consumer use of OTPs became more common with readily available authenticator apps, such as Google Authenticator, becoming available for various smartphone devices.

The FIDO Alliance was founded in 2012 to develop a password-less authentication protocol and later, an open, second-factor protocol.[xiv] The World Wide Web Consortium (W3C) released the first WebAuthn specification in conjunction with FIDO Alliance Client to Authenticator Protocol (CTAP) in March 2019, enabling FIDO2 as a phishing-resistant authentication protocol across platforms, browsers, and devices.[xv] With the public release of passkeys by the FIDO Alliance, W3C, and commercial partners in 2022, the tools for strong, highly phishing-resistant authentication already exist in many consumer and enterprise devices such as laptops, tablets, and phones.[xvi]

Bruce Schneier wrote in 2005, "Two-factor authentication isn't our savior. It won't defend against phishing. It's not going to prevent identity theft. It's not going to secure online accounts from fraudulent transactions. It solves the security problems we had ten years ago, not the security problems we have today."[xvii] Schneier's blog post was prescient. Even after the broad rollout of MFA mechanisms starting in 2005, we are still fighting against phishing, fraud, and identity theft in the 2020s. As the industry has adapted to these ills, malicious actors have also adapted their mechanisms. As we close the front door with better technology, what new paths will actors take to achieve their nefarious goals?

Why Choose Multi-factor Authentication?

The key benefit of adopting MFA is that it improves individuals' and enterprises' security posture and delivers a higher level of assurance to guard against unauthorized account access. With MFA enforced, users are required to authenticate by presenting multiple factors, for example, a username, password, and fingerprint from their device. These additional factors reduce the risk of unauthorized access when one of the authentication factors is compromised, such as a leaked password through a third-party data breach or a phishing attack. You can think of every factor added as an additional lock as an access security layer to prevent unauthorized users from breaking in.

The Problem with Single-Factor Authentication

Single-factor authentication is when access is provided when a user presents one factor. This presentation could be in the form of a password, access card, or fingerprint biometric. The most common single-factor authentication mechanism is the password. Password-less mechanisms, such as passkeys, designed to replace passwords as an authentication factor, are expected to see broad consumer rollout after their introduction in 2022.

However, passwords are still the most widely used mechanism to authenticate to various online services. Passwords are vulnerable to various attack techniques commonly used by attackers to gain access to online accounts. Here are some examples of those techniques:

- **Identity Theft:** This is when an attacker illegally acquires personal information such as date of birth, credit card details, or even answers to security questions that could be used for password guessing or resets.
- **Phishing:** This is when an attacker falsely presents themselves as a trusted party through fraudulent emails, websites, or pop-ups, hoping that they collect someone’s personal information, such as username/password.
- **Brute force:** This involves an attacker guessing username and password combinations in hopes that they would eventually gain unauthorized access to an account
- **Credential Stuffing:** This is when an attacker uses a list of known compromised passwords to take over someone’s account.
- **Key-logging:** This requires an attacker to compromise the end-point like a public computer where they would have installed a key-logger to monitor and record

actual keystrokes for personal information such as login information and credit cards.

- **Man in the middle:** An attacker could use URLs that closely resemble the intended website. This deceptive URL is then used to direct the user to a reverse proxy server used by the attacker to intercept the communication between the user and the intended website in order to steal sensitive data, such as a user's password.

The introduction of passkeys presents an interesting dilemma: Are passkeys an MFA mechanism when they are syncable across cloud services? What about when they are resident on a single device? If passkeys are primarily designed as a single authentication factor to replace passwords, will we see passkeys deployed with additional factors? With varying security models depending on where the passkeys are generated and how they are stored, synced, and shared, we believe it is likely that some passkey implementations will require additional authentication factors. For additional passkey considerations, see the FIDO section below.

Multi-factor Authentication Mechanisms

Grid Cards & Grid-Based Mechanism

Possession Factor: Card
Knowledge Factor: Password
Inherence Factor: None
Phishing Resistance: None

A form of a challenge-response protocol, a unique grid with named columns and rows is printed on card stock, a plastic card (e.g., student ID), etc.[xviii] At each set of coordinates is a cell containing an alphanumeric value. Upon first factor authentication with a password, the user is presented with a dynamic challenge requiring entry of the values at multiple coordinates on the grid as a second factor.

Credential Calculators Hardware Token

Possession Factor: Credential Calculator
Knowledge Factor: Password
Inherence Factor: None
Phishing Resistance: None

In another form of challenge-response protocol, users authenticate to a service with a password and receive a numeric challenge. This challenge is entered into the device using a keyboard, and the response is calculated. The user enters the output into the service to complete the authentication process.

One-Time Passwords - HOTP

Possession Factor: HOTP Generator
Knowledge Factor: PIN or Password
Inherence Factor: None
Phishing Resistance: None

Described by RFC 4226 as “An HMAC-based One Time Password Algorithm,” HOTP is a commonly used second factor. Successive HOTP values are generated through the application of the HMAC-SHA1 algorithm, whose inputs are a static seed value, unique per device and shared with the server, and the counter, a numeric value that increments on each iteration. The output is truncated to a set of human-readable numbers, often 4-8 bytes in length.

Generally found on hardware devices with small display screens showing a set of numbers after pressing a button, the HOTP output is entered into a form field by the user to complete authentication. Of note with HOTP generation is that the codes are generated dynamically in response to a user action, such as a button press. This can lead to devices becoming out of sync with the server state when multiple HOTPs are generated by the client and unused. Desynchronization must be addressed through a re-synchronization process that is undefined by RFC.

One-Time Passwords - TOTP

Possession Factor: TOTP Generator
Knowledge Factor: PIN or Password
Inherence Factor: None

Similar to a HOTP, a TOTP is defined by RFC 6238 as a time-based one-time password algorithm. The RFC describes TOTPs as a “variant of the HOTP algorithm [that] specifies the calculation of a one-time password value, based on a representation of the counter as a time factor.” Since the successive values are not generated in response to a user action, desynchronization is less of an issue with TOTPs vs. HOTPs, assuming the services are not subject to excessive clock-skew.

Similar to HOTP, TOTP is often implemented in hardware devices with a small display screen that is constantly refreshed over time, displaying 4 to 8 digits. Additionally, TOTPs are often implemented by software such as password managers, Authy, etc., as a convenient mechanism for users with smartphones to carry multiple TOTP generators for different services on a device they already possess. In this case, the user will scan a QR code with their TOTP software application, instantiating the TOTP in the software. The user then enters the current TOTP into the relying party's service to validate the TOTP has been instantiated correctly. These TOTPs may exist on multiple devices, either through a cloud-based sync or re-scanning the QR code on multiple devices as a backup of the TOTP generator.

TOTP, like HOTP, was developed by the Initiative for Open Authentication, an industry group that developed the open specifications, which later became IETF RFCs. The standards developed by OATH enabled the creation of an ecosystem of hardware devices and software implementations, eliminating the need for context-specific second factors.

One-Time Passwords - SMS (Short Messaging Service)

Possession Factor: Access to SMS on a mobile device
Knowledge Factor: PIN or Password
Inherence Factor: None

SMS OTP allows a user to authenticate using a one-time password sent over to the user's mobile number using SMS. The user configures their phone number with a relying party to receive OTPs during authentication. As noted above, NIST-800-63rev3 identifies SMS OTP as a "restricted" authenticator.

> *"The use of a RESTRICTED authenticator requires that the implementing organization assess, understand, and accept the risks associated with that RESTRICTED authenticator and acknowledge that risk will likely increase over time. It is the responsibility of the organization to determine the level of acceptable risk for their system(s) and associated data and to define any methods for mitigating excessive risks. If at any time the organization determines that the risk to any party is unacceptable, then that authenticator SHALL NOT be used.*[xix]

> *Verifiers SHOULD consider risk indicators such as device swap, SIM change, number porting, or other abnormal behavior before using the PSTN to deliver an out-of-band authentication secret."*[xx]

One-Time Passwords - Email

Possession Factor: Email address (no physical possession)
Knowledge Factor: PIN or Password
Inherence Factor: None

Email OTP allows a user to user to authenticate using a one-time password sent over to a registered email address registered to the user's account. The user must provide the OTP value during the authentication ceremony. The security of email OTP is dependent upon the security of the user's email service.

One-Time Passwords – Magic Links

Possession Factor: Indeterminate
Knowledge Factor: Indeterminate
Inherence Factor: Indeterminate

Magic links provide a fast and easy sign-in user experience. Users are authenticated by providing their email address only; they are then sent an email with a link for the user to click and complete their sign-in. This link is an embedded token that can only be used once. This provides a password-less login experience, which has many user experience advantages. However, it is worth mentioning that magic links are only as secure as a user's email address. For example, if someone gets access to a user's inbox, they can now access the magic links as they get sent to the user, which might lead to an authorized access event. Therefore, we classify the possession, knowledge, and inherence factors are indeterminate – the security is dependent upon the authentication credentials to the email service and any devices which have persistent access to the same.

FIDO U2F / FIDO2

Possession Factor: *Devices such as a phone, tablet, laptop, or a FIDO hardware security key*
Knowledge Factor: PIN code (optional, may be used in place of an inherence factor)
Inherence Factor: *fingerprint, iris, or faceprint (optional, may be used in place of a PIN code)*

The FIDO protocols (U2F/CTAP1, CTAP2.x) and WebAuthn use asymmetric cryptography to authenticate users on external hardware devices (e.g., security keys) and platform authenticators built into laptops, tablets, and phones. Authentication credentials are scoped to origins controlled by the relying party; relying parties cannot discover credentials for unrelated origins to protect privacy.[xxi] The credentials may be bound to a single device, as with hardware keys and some platform authenticators, or synchronized across a cloud fabric, ensuring availability across the user's devices. FIDO credentials are considered to be highly phishing resistant.

Some FIDO credentials are attestable. At registration, the authenticator emits a signed attestation statement identifying the provenance of the authenticator. Relying parties can validate the signature on the attestation and collect additional authenticator metadata through the FIDO Metadata Service (MDS).[xxii] This data may include information about the authenticator's certification level and conformance to standards such as FIPS140-1.[xxiii]

Implementers should note that not all FIDO credentials are created equally. FIDO credentials may be created and managed entirely in software, within TPMs, Secure Enclaves, or other hardware embedded in general-purpose computers, phones, and tablets, or on hardware security keys. While all of these credentials use the same cryptographic primitives and protocols, relying parties should have an understanding of the differences between FIDO authentication mechanisms to help them make effective choices when implementing FIDO solutions.

- Passkeys are discoverable credentials that reside on the system that created them.
- Passkeys may be used as a highly phishing-resistant, single-factor credential, replacing passwords.
- The number of passkeys that can be configured on a single hardware security key is limited by the properties of the hardware and credentials.
- Passkeys created on hardware security keys do not leave the device.
- Passkeys may be synchronized across a fabric provided by platforms (Apple, Google, Microsoft) or password managers (1Password, Dashlane). Synchronization fabrics are provider-specific. Synchronized keys are sometimes called “multi-device credentials”. Non-synchronized keys are “single-device credentials”.
- Passkeys cannot be synchronized across providers.
- Synchronized credentials create an alternative credential recovery pathway. Credential recovery mechanisms are provider-specific.
- Passkeys may be shared by exporting them to nearby contacts through the AirDrop protocol on Apple platform devices.
- Passkeys, like all FIDO credentials, may not carry an attestation during registration. Relying parties may request attestation during credential registration. Authenticators and browsers may restrict whether an attestation is returned.
- In the event that a credential does not meet the relying party’s requirements, the RP must reject credential registration after the credential is created on the authenticator.
- Relying parties cannot be assured of the origin or security properties of unattested credentials. High-assurance use cases should require and validate all attestations.

The breadth of the FIDO2/WebAuthn ecosystem is too broad for this article. Look for a future BoK article on the FIDO protocols to address these protocols in more depth.

Push-Based Authentication

Possession Factor: *Access to the mobile device where the push notification is sent*
Knowledge Factor: PIN or Password (optional)
Inherence Factor: *Biometric on the device (optional)*

Push-based authentication is primarily a mobile-based experience. At authentication time, the service sends a push notification to the user's registered device(s) or applications. The user receives the notification and may approve or decline the request. As with most technologies, this has been abused by malicious actors who use social engineering or prompt bombing attacks to obtain the user's help to complete the authentication process.[xxiv] These attacks can be mitigated by providing additional context data to the user, such as the location of the authentication session or device identity, or requiring the user to copy a number from the push notification to the device attempting authentication.[xxv]

Smart Cards

Possession Factor: Smart Card
Knowledge Factor: PIN (optional, may use inherence factors)
Inherence Factor: Fingerprint (optional, may use PIN)

Smart Cards are physical devices of varying sizes (e.g., nano-SIM, SIM, credit card form factors) used to store a credential, often in the form of a cryptographic certificate, which can be unlocked by the user presenting a PIN or inherence factor to facilitate authentication. The card may be presented by insertion into a physical reader or via a contactless protocol, such as NFC.

Smart cards exist in a wide variety of formats with different use cases depending on the industry in which they are used. A common deployment is the use of a Common Access Card (CAC) by the US Federal Government. After identity proofing, the federal government issues a CAC to an individual as both a physical identity document used to access government property, as well as a multi-factor authenticator. Upon inserting the CAC into a reader, the user enters a PIN to unlock the device. Once unlocked, the CAC authenticates the user against a directory service via the public key certificate embedded in the hardware.

Threat Mitigation by MFA Mechanism

The NIST Special Publication 800-63B is a recommended read as it provides an informative section on the various threat and security considerations and how to mitigate them. In this section, we highlight a subset of threats against MFA mechanisms and whether the mechanism is susceptible to the threat (❌), partially mitigates the threat (∼), or completely mitigates the threat (✅).

The threats considered below are:

- Credential duplication - Can the credential be duplicated and used in a manner undetectable to the owner? For example, a grid card could be photographed and used illicitly if the password was known, but the attack is not scalable.
- Eavesdropping / Man in the Middle - Active or passive eavesdropping of communications can compromise flows that depend on secrets, either by sniffing the secret off the wire as they are being delivered to the recipient (e.g., attacks on mobile SMS networks or SIM swapping), or by replaying secrets obtained through phishing.
- Replay - Some MFA mechanisms are designed for one-time use. Implementations may fail to enforce one-time use of these secrets, allowing sniffed secrets to be replayed.
- Social Engineering - Manipulating a target through psychological means such as authority, intimidation, urgency, and other mechanisms to force a victim to take actions that may not be in their own best interests. In the realm of MFA, this may be seen through attacks such as prompt bombing.
- Phishing - A form of social engineering where the victim is enticed into entering their credentials into a fraudulent site designed to look like a legitimate service. Phishers will collect credentials, including passwords and second factors, and use them immediately to authenticate to the legitimate site to further their schemes. In 2020, phishing was the most frequent crime reported to the FBI Internet Crime Complaint Center (IC3), representing almost one-third of all complaints (241,343 of 791,790).[xxvi]

Threats (--->)	**Credential Duplication**	**Eavesdropping / Man in the Middle / Replay**	**Phishing**	**Social Engineering**
Mechanisms (down)				
Grid Cards & Grid-Based Mechanism	❌	~	❌	❌
Credential Calculators Hardware Token	❌	~	❌	❌
One-Time Passwords - HOTP	~	~	❌	❌
One-Time Passwords - TOTP	❌	~	❌	❌

One-Time Passwords - SMS	N/A	~	❌	❌
One-Time Passwords - Email	N/A	~	❌	❌
FIDO U2F / FIDO2	~	✅	✅	✅
Push-Based Authentication	N/A	✅	✅	~
Smart Cards	✅	✅	✅	✅

Conclusion

Using MFA is now considered an essential security best practice. It protects against many cyber threats, and the user experience has significantly improved since the early days of heavy hardware tokens. There is more to learn when it comes to deploying MFA in an environment; we suggest further exploring this space by reading Nishant Kaushik's "Designing MFA for Humans".[xxvii]

References

[i] "Terminology in the IDPro Body of Knowledge," IDPro Body of Knowledge, updated 30 September 2021, https://bok.idpro.org/article/id/41/.

[ii] Grassi, Paul A., James L. Fenton, Elaine M. Newton, Ray A. Perlner, Andrew R. Regenscheid, William E. Burr, and Justin P. Richer, "NIST Special Publication 800-63B: Digital Identity Guidelines: Authentication and Lifecycle Management," National Institute of Standards and Technology, U.S. Department of Commerce, updated 2 March 2022, https://doi.org/10.6028/NIST.SP.800-63b.

[iii] Bátiz-Lazo, Bernardo, "A Brief History of the ATM: How automation changed retail banking, an Object Lesson," The Atlantic, 26 March 2015, https://www.theatlantic.com/technology/archive/2015/03/a-brief-history-of-the-atm/388547/ (accessed 14 December 2022).

[iv] Batiz-Lazo, Bernardo and Reid, Robert J. K., "Evidence from the Patent Record on the Development of Cash Dispensing Technology," MPRA: Munich Personal RePEc Archive, University of Leicester, University of Leicester and University of Glasgow, 30 June 2008, https://mpra.ub.uni-muenchen.de/9461/1/MPRA_paper_9461.pdf (accessed 14 December 2022).

[v] M'Raihi, D., Machani, S., Pei, M., and J. Rydell, "TOTP: Time-Based One-Time Password Algorithm", RFC 6238, DOI 10.17487/RFC6238, May 2011, <https://www.rfc-editor.org/info/rfc6238>.

[vi] M'Raihi, D., Bellare, M., Hoornaert, F., Naccache, D., and O. Ranen, "HOTP: An HMAC-Based One-Time Password Algorithm", RFC 4226, DOI 10.17487/RFC4226, December 2005, <https://www.rfc-editor.org/info/rfc4226>.

[vii] U.S. Department of Homeland Security,"Homeland Security Presidential Directive 12: Policy for a Common Identification Standard for Federal Employees and Contractors," last updated 27 January

2022, https://www.dhs.gov/homeland-security-presidential-directive-12 (accessed 14 December 2022).

[viii] National Institute of Standards and Technology, "Personal Identity Verification (PIV) of Federal Employees and Contractors," Federal Information Processing Standard (FIPS) 201-1, March 2006, https://csrc.nist.gov/CSRC/media/Publications/fips/201/1/archive/2006-06-23/documents/FIPS-201-1-chng1.pdf. Please note version is for historical reference only; the current version of this publication is FIPS 201-3, published January 2022 and available at https://doi.org/10.6028/NIST.FIPS.201-3.

[ix] Dray, James, Scott Guthery, and Teresa Schwarzhoff, "NIST Special Publication 800-73-1: Interfaces for Personal Identity Verification," Computer Security Resource Center, National Institute of Standards and Technology, U.S. Department of Commerce, March 2006, https://csrc.nist.gov/publications/detail/sp/800-73/1/archive/2006-03-15. Please note version is for historical reference only; the current version of this publication is NIST 800-73-4, published May 2015 and available at https://doi.org/10.6028/NIST.SP.800-73-4.

[x] U.S. Department of Justice, Office of Justice Programs, "Putting an End to Account-Hijacking Identity Theft," NCJ Number: 210758, December 2004, https://www.ojp.gov/ncjrs/virtual-library/abstracts/putting-end-account-hijacking-identity-theft (accessed 14 December 2022).

[xi] Federal Financial Institutions Examination Council, "Authentication in an Internet Banking Environment," 2005, https://www.ffiec.gov/pdf/authentication_guidance.pdf (accessed 14 December 2022)

[xii] Libicki, Martin C., Edward Balkovich, Brian A. Jackson, Rena Rudavsky, Katharine Watkins Webb, "Influences on the Adoption of Multifactor Authentication," technical report, RAND Homeland Security and Defense Center, 2011, https://www.rand.org/content/dam/rand/pubs/technical_reports/2011/RAND_TR937.pdf (accessed 14 December 2022).

[xiii] "Banks to Use 2-factor Authentication by End of 2006," Slashdot forum discussion, 2005, https://it.slashdot.org/comments.pl?sid=165833&cid=13832042 (accessed 14 December 2022).

[xiv] FIDO Alliance, "History of FIDO Alliance," n.d., https://fidoalliance.org/overview/history/ (accessed 14 December 2022).

[xv] "Web Authentication: An API for accessing Public Key Credentials Level 1," W3C Recommendation, 4 March 2019, and "Client to Authenticator Protocol (CTAP)," FIDO Alliance, 21 June 2022, https://fidoalliance.org/specifications/download/.

[xvi] Passkey.dev website, W3C WebAuthn Community Adoption Group and the FIDO Alliance, https://passkeys.dev/ (accessed 14 December 2022).

[xvii] Schneier, Bruce, "The Failure of Two-Factor Authentication," Schneier on Security blog, March 2005, https://www.schneier.com/blog/archives/2005/03/the_failure_of.html (accessed 14 December 2022).

[xviii] Williams, Andy, "Grid-based two-factor authentication comes to campus cards," SecureIDNews, 25 September 2006, https://www.secureidnews.com/news-item/grid-based-two-factor-authentication-comes-to-campus-cards/# (accessed 14 December 2022).

[xix] NIST 800-63B Section 5.2.10 Restricted Authenticators.

[xx] NIST 800-63B Section 5.1.3.3 Authentication using the Public Switched Telephone Network.

[xxi] "Web Authentication: An API for accessing Public Key Credentials Level 2," W3C Recommendation, 8 April 2021, section 3. Dependencies, https://www.w3.org/TR/webauthn-2/#scope.

[xxii] FIDO Alliance Metadata Service, website, https://fidoalliance.org/metadata/ (accessed 14 December 2022).

[xxiii] FIDO Alliance Certified Authenticator Levels, website, https://fidoalliance.org/certification/authenticator-certification-levels/ (accessed 14 December 2022).
[xxiv] Goodin, Dan, "A Sinister Way to Beat Multifactor Authentication Is on the Rise," Ars Technica, 30 March 2022, https://www.wired.com/story/multifactor-authentication-prompt-bombing-on-the-rise/.
[xxv] Cybersecurity & Infrastructure Security Agency, "Implementing Number Matching in MFA Applications," October 2022, https://www.cisa.gov/sites/default/files/publications/fact-sheet-implement-number-matching-in-mfa-applications-508c.pdf.
[xxvi] "Internet Crime Report 2020," Internet Crime Compliant Center, Federal Bureau of Investigation, 2021, https://www.ic3.gov/Media/PDF/AnnualReport/2020_IC3Report.pdf.
[xxvii] Kaushik, N., (2020) “Designing MFA for Humans”, IDPro Body of Knowledge 1(3). doi: https://doi.org/10.55621/idpro.49.

Access Control

Introduction to Access Control (v4)

By André Koot

To comment on this article, please visit our GitHub repository and submit an issue.

Table of Contents

Abstract

As the name implies, Identity and Access Management (IAM) is split into two functions: managing identity information and performing access control. Arguably, if there was no access control requirement there would be no need for identity management. It is therefore the focus for IAM professionals. At its core, access control is ensuring users are authenticated to access protected resources. This is accomplished by managing user entitlements and satisfying the requirements of relying applications so that users can only access the systems and information they are entitled to access. This article looks at the history of access management, the expected current functionality, and the trends to be expected.

Introduction

Access control, as a concept, has a long history. But in order to investigate the current challenges and solutions, let's start by evaluating a very old, traditional model of classified government documents.

Information in documents stored in files should not typically be accessible to everyone. The information may be classified, and only people with a required clearance level should be able to access classified files. In a physical form, this control is relatively simple: a folder with highly classified information is visually classified by a 'Top Secret' or 'For Your Eyes Only´stamp.[i]

But this simple example already addresses different fundamental concepts of security. First, there's the information itself. The information can be classified as Top Secret, but that must be defined by someone with the correct level of authority, like the owner of the information, the document, or the folder. Then, it must be clear what the impact of the classification level is; the classification level is needed to differentiate different levels of access to and usage of the information. The owner of the folder will probably have some guidance as to what levels of classification can be applicable and what type of user can get access.

Second, there is the clearance level of an actor, the user of the information. In this case, the secret service agent will have been identified and vetted to be trusted in such a way that access to different security levels of information is allowed.

Third, the classification level and the clearance level will have to be mapped in order to ensure that only the person with the correct security clearance level can access the classified information. The owner will classify a document and will accept that a specific security level can only be accessed by a pre-defined trust level of an agent.

And fourth: before giving the folder to the secret service agent, the person who is responsible for storing and retrieving the file in an archive (the file manager or access controller) must verify if the agent who requests the file is in fact the rightful user. The access controller will, therefore, try to identify the agent; the agent has to prove the right to access. This verification can be done by showing the secret service badge and a signed letter to prove that the agent has permission to access the folder. The file manager will, of course, also have to validate that the signature on the letter is correct.

Only after these responsibilities have been fulfilled will the folder be handed over to the secret service agent. The hand-over is then registered in a journal.

The access controller will always oversee the access, and that’s been made easy by checking the stamp on the folder. Theft of information—e.g., data leakage—is also quite physical in this example: the folder is removed. A folder with the 'Top secret' stamp should also not be found lying around unobserved.

In this scenario, access control is quite simple: you can literally observe access infractions. Access is granted by physically handing over the folder to a person with the corresponding clearance level, indicated by a personal badge, and may be enforced further by restricting access to a specific location.

We can see the following topics:

1. Classification of information: this is an aspect of risk management
2. Classification of users: this is an aspect of identity management
3. Authorization mapping: this belongs to authorization management
4. Authentication: this verification is part of both identity management and access management
5. Access granted: this is access control

Since the advent of the computer, there has been a need to control access to systems, documents, and other protected resources. In the early era of computers, processes analogous to the old spy movie era were used to model access control mechanisms. Concepts like ‘owner of a resource’ and ‘reader of a resource’ were used. Programmers developed access control mechanisms like Discretionary Access Control (DAC) (“you may never bypass the access controller,” a feature that can still be found in the Windows NTFS file system) and Mandatory Access Control (MAC) (“you can only access the data in a specific location” such as a dedicated workstation in a specific room).[ii,iii] The fast growth of information technology resulted in a growing need to develop and improve access control. The increase in the number of users, the number of systems, and the exponential growth of the information processed makes it evident that the paper world metaphor is not sustainable in the digital world.

It was soon realized that the concept of trust levels—e.g., managing the clearance level of an individual document reader—is hard to implement. Because so many actors are playing along and there is no longer a physical security control in place (you cannot see the red lint). Instead, there can even be multiple copies of a folder or file in multiple locations, and theft no longer means that the data is gone, but data will probably be copied without the consent of the owner. What was physically easy to implement is not easy to implement in the digital world. But the lessons learned in identification, authentication, authorization, access control, logging, and auditing, have been kept.

Access to information, data, services, and systems, as well as access to physical locations, is governed by security policies. These security policies must be formalized and need to be enforced by the owner of the resource. In doing so, the owner will try to manage the risk involved in access, such as the risk of abuse of information, data leakage, theft, fraud, and other security threats. In order to be in control, the owner needs to have the assurance of the level of security capable of being achieved by the security controls that have been put in place.

Apart from the concepts of access control, ownership in itself is a complex topic. Looking at the concept of data ownership, many criteria to establish ownership can be identified. Someone can be the owner of information because:

- They created the data.
- They funded the data processing facility.
- The data is about this person (e.g., a medical record of a patient)

There can be many more criteria to identify the owner, but this is part of data governance and out of scope for this article. In the case of medical files, the object, the patient, has several inherent rights to the data, making this person partly accountable for the access decision.

Terminology

- Identification – Uniquely establish a user of a system or application.
- Authentication – The ability to prove that a user or application is trustworthy and has the authority to access a protected resource by validating the credentials of an access requester (a user, a process, a system, or a thing).
- Multi-factor Authentication (MFA) – An approach whereby a user’s identity is validated to the trust level required according to a security policy for a resource being accessed using more than one factor (something you know (e.g., password), something you have (e.g., smartphone), something you are (e.g., fingerprint).
- Authorization – Determining a user’s rights to access functionality with a computer application and the level at which that access should be granted. In most cases, an

'authority' defines and grants access, but in some cases, access is granted because of inherent rights (like patient access to his/her own medical data).
- Accountability – The obligation of a person to accept the results of one's actions, be they positive or negative. This person is probably also a type of owner.
- Protected Resource - A system, process, service, information object, or physical location that is subject to access control as defined by the owner of the resource and by other stakeholders, such as a business process owner or risk manager.
- Access Control – Controlling who can have access to data, systems, services, resources, and locations. The 'Who' can be a user, a device or thing, or a service.
- Access Governance – The assurance that all access has been given based on the correct decision criteria and parameters.
- Access Policy – Definition of the rules to allow or disallow access to secured objects.
- Access Requester – The person, process, system, or thing that seeks to access a protected resource.
- Access Supplier – The component granting access to data, systems, and services after the access policy requirements (set in the Policy Administration Point) have been met by the Access Requester.
- Policy Engine - It is a security component that validates whether an actor is allowed to access a protected resource, following the requirements in an access policy. A policy engine can be seen as a component that exists of a PDP and a PAP combined.
- Policy Enforcement Point (PEP) – The authority that will only let an access requester connect to the access supplier if the Policy Decision Point allows it.
- Policy Decision Point (PDP) – The policy engine validates access requests and provides attributes against the access policy (as defined in the Policy Administration Point).
- Policy Administration Point (PAP) – The location where the different types of owners define the access policy.
- Policy Information Point – The authority that refers to the (external) trusted providers of attributes that will be used in the Access Decision. An example is the credly.com service that administers Open Badges of certifications, such as CIDPRO™ or the Certified Information Systems Security Professional (CISSP).

Acronyms

- ABAC – Attribute-Based Access Control
- ACL – Access Control List
- AIAC – Artificial Intelligence-Supported Access Control
- CBAC – Context-Based Access Control or Claims-Based Access
- CIAM – Consumer Identity and Access Management
- CRM – Customer Relationship Management
- DAC – Discretionary Access Control
- MAC – Mandatory Access Control
- PBAC – Policy-Based Access Control

- PAP – Policy Administration Point
- PDP – Policy Decision Point
- PEP – Policy Enforcement Point
- RBAC – Role-Based Access Control or (less frequently) Rule-Based Access Control
- ReBAC – Relation-Based Access Control
- SCIM – System for Cross-domain Identity Management
- SoD – Segregation of Duties

AAA: Authentication, Authorization, Accountability

Just as we showed in the classified document example above, in order to get access, a validated identity is key. The ideas behind this paradigm can be summarized by the concepts of AAA.

Authentication

Authentication is the process of proving that the user with a digital identity who is requesting access is the rightful owner of that identity. It can be as simple as using a password or as complex as providing a digital certificate. Both the Access Supplier and the Access Requester must be able to manage and consume the results of the authentication process.

Challenge - Response

The user might provide proof of this rightful usage by providing a secret that only the access requester and the access supplier know, like a secret code or a password. The underlying mechanism is called Challenge-Response. The Access Supplier challenges the Access Requester to prove his or her identity, and the subject will have to respond in the way the Access Supplier expects. The simplest way to do a challenge-response is by asking for a password or pin-code. But also, the CAPTCHA feature on many websites is a form of challenge-response: prove that you are a human being.[iv]

Knowledge – Possession - Being

But other than a CAPTCHA challenge, a known secret can be shared. It may not be sufficient to assure the rightful access because by sharing a password or by finding a password lying around (on a piece of paper, for instance), others may pretend to be the rightful owner. This weakness of the known-secret model means that the trust level of an access requester who uses just a password may not be sufficient for some applications.

After identification and even authentication, there is a degree of uncertainty in identifying the rightful owner, which should result in further evaluation of the level of access. A low level of confidence may be enough to give access to public information, but it will probably be insufficient to provide access to classified information.

Adding more proof of identity can be done by demanding more specific and unique identifiers. These more trusted authentication means cannot be easily copied or easily shared or stolen (it is not impossible, but the cost of copying a secure physical token can be too high to make it economically unsound to forfeit). In practice, this is done by introducing additional factors, such as tokens, certificates, and biometric proof. Requesting these additional proofs of identity can be requested either at the start of a session at the first authentication or during a session after a previous low-trust authentication has been found insufficient for getting access to a secured resource. In this case, the low-trust access can be enhanced by performing a ‘step-up’ authentication, requiring additional factors: the first step during login could be using a password, and then a second higher-level step could involve the use of a token or biometric proof.

Authorization

Authorization, often a synonym for the phrase access control, is the next step in getting access after the phase of authentication. It is the act of granting access to a specific resource, such as a computer application or a specific function within an application.

Authorization is closely related to the concept of authority. Someone, such as an owner, is accountable and, because of the ownership, is mandated to authorize others to access the protected resource. This accountability does not imply that the other person becomes the owner, but it does mean that several permissions, such as ‘read’ or ‘delete,’ can be executed. The owner stays accountable throughout the lifecycle of the data. Some of the tasks of the owner can be delegated to others in such a way that, for instance, a line manager may, within the boundaries set by the owner, grant read access to a resource to an employee.

Mainstream Access Control Methods

Currently, many organizations have security policies embedded in various applications, operating systems, and networking components. These controls are implemented in the form of Access Control Lists (ACLs), Roles, and DAC business rules. But these controls have to be designed and implemented in every relevant component. And these controls have to be designed in a consistent manner. If, for instance, a Segregation of Duties (SoD) restriction is defined for a specific process, every system, application, platform, app, and network component must support the SoD rule. If one of the many components is lacking SoD control, then the organization is not in control.

This decentralized implementation of security policies makes it challenging to implement centrally managed organization-wide controls. It is likely that not all controls are similar and that the security policy and conformity must be verified for every system or platform access request.

Modern Access Control

In modern implementations of access control, a policy engine is used to evaluate access policies centrally, and policy enforcement should encompass the ‘risk level’ evaluation. The business process owner, or data owner, tasked with managing access risk, will define the policies for which they are accountable. In some cases, there are multiple ‘business owners,’ and each is responsible for their part of the corporate security policy. This assignment of business owners can result in continuously changing access control policies.

There is much development in this area, with applications no longer maintaining the ACLs of users. Instead, they rely on identity management authorization systems that will, based on one or more access policies, make the decision regarding a user’s access request. Different stakeholders in a company are responsible for different policies. All applicable policies must be evaluated before access is granted. This method of fine-grained access control is a type of MAC.

Accountability

Accountability is a key responsibility in access governance. Making sure that every access decision is accounted for by an authorized person implies that ownership must be addressed. The owner must be informed about all activities under their control in order to be successfully accountable for the data under their stewardship.

Registering all activities in access control is an essential quality requirement. This record can vary in complexity from logging every authorization request (like granting or revoking authorizations or roles to and from people) to logging changes of authorizations within roles. The existence of this register is essential to be truly in control of access. The same is true for the identification and authentication process. There must be assurance from the part of the login mechanism, the operating systems, and the IAM solutions applied to make sure that every access request is validated.

Specific Access Control Considerations

Access control is not only a business decision. Other considerations inform how this activity must take place, including how users will engage with the control mechanisms as well as legal implications for what is (and isn’t) required.

The Human Factor

The user who needs to cope with the security controls can themselves be a roadblock on the path toward effective ’control.’ User experience (UX) is a critical success factor in every information security project. If the security controls are too strict, users may be deterred, or they may try to circumvent the control. This avoidance on the part of the user is often seen in consumer access: if a customer portal is not built with a focus on the user, then consumers tend to go elsewhere. That is a missed opportunity, resulting in low conversion rates. Consumer Identity and Access Management (CIAM) solutions are developed to prevent this behavior.

The lessons learned in CIAM are also being implemented in workforce IAM: UX is starting to make an impact. For instance, if a user accesses a company intranet portal from their home location regularly in a prescribed way, like using a VPN, the access control system could validate this behavior as a factor in the authentication process. It could decide not to require the repeated use of multi-factor authentication since it is a trusted user making use of a known, trusted connection; it's a well-known context resulting in better control of access.

Legal Implications

Access control has historically been looked at as a way to support business processes and is part of a larger information security and risk mitigation policy. The question of legal implications directly tied to access control practices varies from business to business, from sector to sector, and from jurisdiction to jurisdiction. There is no unambiguous answer as to the direct legal requirement for most access control practices as these policies are often woven into a larger program that is driven in part by any number of laws, regulations, or standards. Part of the role of an access control program or system is to ensure that it is flexible enough to support the larger risk management programs of the business or organization. In this way, questions about legal requirements and compliance implications can be addressed organically, allowing the organization the confidence it needs to operate and move forward.

In separate articles in the IDPro BoK, different aspects of laws and regulations will be illustrated in more detail.

Current state of Access Control

Mainstream Access Control Mechanisms

Several mechanisms support the implementation of access control. This section covers the more common ones: Access Control Lists (ACLs), Role-based Access Controls (RBACs), and Attribute-based Access Controls (ABACs).

Access Control Lists

Access control to a protected resource is based on the classification level of the resource. Every resource will be classified by the owner (or a delegated person) in order to define the security level of the resource. Based on the security level, security controls must be put in place to ensure the correct level of access. The access available, i.e., the permissions that can be granted, are also known as entitlements (fine-grained permissions to access resources). One of the earliest and best-known implementations of entitlements is by using ACLs In an ACL, the owner of the file defines what users can have what type of access: read, write, update, delete, whatever the owner accepts as usage. This concept is easy to understand and easy to manage for individual objects. And if the number of objects is limited, controlling access via ACL's can be enough. But when the number of users and the number of objects grows, ACL's can be a restricting factor.

Every owner of a file will need to define the ACL for the object. This distributed method of control implies that central control of access is non-existent. But, from an auditing perspective: it's relatively simple to find out who has access to a protected resource since that is registered in the ACL of the resource.

The concept of ACLs will be explained in a future article in the BoK.

Role-Based Access Control

Managing ACLs can be a tedious task. Managing access to resources on a user by user or entitlement by entitlement basis faces issues as populations grow. At some point, the issue of scale meant that a new access management approach was needed. RBAC is an approach of granting access to resources on a group level instead of on an individual level. In order to realize this, an intermediate component needs to be in place after that of the access controller. A role manager or a role owner has to be able to map the role of a user to an entitlement to a secured resource. This mapping looks easy enough, but in practice, this means that this person needs to work with different other responsible persons in an organization to make sure that the authorizations are not conflicting with the business processes and organizational structures of the organization. In the access governance article, this concept and the complexity connected with the governance model is further explained.

In the example of an internal company website, every company employee is made a member of a group called 'Company Employees.' The resource—in this case, the main page of the internal website—is secured in such a manner that access is granted only if a user is a member of this group. Another example is the line manager who can make a new employee member of the role account manager and behold, the access permissions connected to the role account manager, are available to the new employee. This non-individual oriented way of granting access makes managing access a lot easier.

A system owner can also create 'roles' within an information system to prevent the need for managing individual entitlements. The system owner of a Customer Relationship Management (CRM) system can define a role for 'customer manager' and group system authorizations (such as reading a customer record from a database or filling in a form) to that role.

In RBAC, we can identify a multilevel role model. On the one hand, we can identify the grouping of identities organizationally or hierarchically, defining organizational or business roles. On the other hand, there is a grouping of authorizations or permissions on an application function or platform level called system or application roles. Connecting organizational roles to application roles creates a very efficient way of granting and revoking authorizations. But it is also very easy to complicate authorization management by nesting groups: for instance, employees working on the service desk can be made members of the group ’ServiceDesk'. This group then could be made a member of the group Windows Administrators. By doing this, it will soon become hard to find out who has the authorizations of a Windows administrator. That would be not just the group of people who are members of the Windows Administrator role but also employees who are members of the role of ServiceDesk employee. This nesting can frustrate the insight by no small means; many IAM projects fail by the lack of un-nesting possibilities. Nesting also limits the auditability of RBAC environments; groups have to be un-nested in order to evaluate authorizations and potential conflicting authorizations.

Implementations, pros and cons, will be explained later in a future article about RBAC in the BoK.

Attribute-Based Access Control

ABAC builds on the RBAC model by introducing additional controls based on business logic. A major failing of the RBAC model is its static nature. Once an entitlement has been granted, it generally is always available to an end-user, until it is manually revoked. This longevity means that users wind up carrying access with them from role to role if proper cleanup actions are not taken. To address this, ABAC expands on the model, taking into consideration different characteristics of users and users’ attributes at the moment of determining if access should be granted. As a result, an access management system can make a decision based on the entitlements of a given user, as well as the time of day, the location of the user (e.g., on network or remote, geolocation based on IP address) the type of device (e.g., personal, organization owned, desktop or tablet), and other worker metadata. ABAC can be used both in real-time to control access at the time of the transaction, or passively controlling the assigned roles and entitlements based on user metadata. Both approaches require strong input and support from resource owners, Role managers, and people or organization managers to understand the needs of the user as well as additional support from analysts to help define the business logic.

For example: The Customer Relations Management process owner could define that everyone with the attribute ‘Business Role = Account manager’ can access the resource only if attribute ‘Allowed Time = defined office hours’. Multiple variations of this dynamic access control philosophy will be described later in a future IDPro BoK article.

The Future Direction of Access Control

Access Control by means of ACLs and RBACs is relatively static; the combination between a user and his or her authorizations are set and do not vary easily, and other authorizations require changes. But people move between jobs, change devices, change location, or get new tasks in a new context. Also, the risk level assigned to a protected resource can change because of a change in context or a change in applicable laws and regulations. Relevant changes may include:

- Extended organizations, internationalization, collaboration and federation, flexible workforce, meaning that in daily operations, people outside the scope of the traditional HR-operations may need to get access.
- Moving data processing to the cloud - leading to the development of new protocols, such as SCIM (System for Cross-domain Identity Management (the first time the acronym was used, it was called Simple Cloud Identity Management, I suppose this was deemed too simple or restricting ☺).[v]
- New privacy regulations, such as the GDPR.[vi]
- The usage of mobile apps, using modern protocols like OpenID Connect requires a flexible access control topology.
- Enforcing end-user consent and control - developments like User-Managed Access (UMA). [vii]
- Move to API-based access to micro-services - leading to new access management architectures based on protocols like OAuth2.

These restrictions and changes show that a more dynamic method for managing access is needed. The future direction of access control takes this into account, and various developments can be identified.

Dynamic Authentication

Access control is not a static event. When a user starts a session accessing services requiring a low-risk level, then identification with a username and password combination may be sufficient. When later on in the session, another trust level is required. For instance, when performing a transaction, additional identification, like a token, might be needed.

In order to adapt to these session dynamics, authentication in itself should also be a continuous process through, for example, the new concept of behavioral biometrics. Examples of changing needs for trust in the identity:

- User switches context (such as location). This switch could effectively place the user in another trust zone, and the session should be re-evaluated
- A user opens an email attachment, which by itself requires a higher trust level. This action should enforce additional authentication, such as Multi-Factor Authentication.

Adaptive authentication is a secure, dynamic, and flexible form of authentication. It enables validating multiple factors to determine the authenticity of a login attempt before granting access to a resource. The factors that are used for user validation can depend on the risk associated with granting a particular user access and may involve adjusting the authentication strength based on the actual context.

Policy-Based Access Control (PBAC)

A dynamic, flexible method is required for access control to become effective and efficient in extended organizations in collaboration environments with a flexible workforce. Policy-based Access Control (PBAC) is the paradigm to provide this flexibility. PBAC, also known as Claims-based Access Control or Content-based Access Control, takes some of the business logic introduced in the ABAC model and enhances it by layering additional context evaluation and dynamic step-up capabilities

The context of an access requester can change dynamically. The dynamic nature of policy management and enforcement could require step-up authentication within a session to cater for the higher trust level needed if the defined risk controls require it. A policy engine will be responsible for checking if the user attributes and context information at the time that access is requested, comply with the access policies defined by the owners of the security policies. Context information might include time of day, geographical location, or device type. The scalability of access is also enabled by making it possible to collect attributes from different trusted and pre-defined attribute providers. As an example: this person can access the Risk Management reports, but only if this person has the CRISC certificate. ISACA provides this certificate, so a lookup in the ISACA registry could answer the question regarding the CRISC certification (the mapping of the Access Requester to the ISACA member is out of scope for this discussion).[viii]

The central component in this architecture is Policy Decision Point, which evaluates access policies and returns a response to the access request. The Policy Enforcement Point then enforces the response either by code embedded in the application or, increasingly, via an API gateway. The Policy Enforcement Engine is a discretionary component in the access request flow.

As a further natural development, AIAC and ReBAC have to be mentioned.

Relation-Based Access Control

A new concept in access control is ReBAC, or Relation-Based Access Control. ReBAC addresses the possibility of making access control decisions using the relationship between the access requester and the other identities who can potentially be affected by the access control decision. These access decisions can be deduced from (amongst other services) social media network relationships of the access requester. An attribute such as ‘reputation’ can be evaluated and considered. ReBAC relies on the availability of large, distinct data sets (incorporating data from HR/Sourcing & Access/entitlement/behavior) and on AI to conduct the evaluations and recommendations for access decisions.

The direction for ReBAC is not yet entirely clear, and the development is not mature enough for mainstream implementation. We foresee the potential for implementation as part of predictive role mining technologies for dynamic ABAC implementations.[ix]

Artificial Intelligence Supported Access Control (AIAC)

We can expect much more in this area when we add the concept of artificial intelligence (AI). With a robust environment that classifies sensitive resources, it's now possible to take a sophisticated risk management approach to dynamic access control whereby the identity manager solution will alert on access requests that exceed normal risk levels. AI will also monitor access control requests alerting on out-of-normal activity. As such, it can be an addition to current RBAC and ABAC implementations. This concept is not yet mainstream, and we can hardly predict the direction, but AI and machine learning may add some value.

User Control and Consent

Privacy laws and regulations create a new awareness of access to personally identifiable information (PII). These laws and regulations have driven the concept of data ownership and consent by customers, employees, or patients. Data owners expect to be in control of their personal information, and in many cases, laws and regulations are mandating this. Several technological platforms have begun to spring up to fill this data ownership gap. Solutions like User-Managed Access, by Kantara Initiative, have made their way in the new access paradigms. Facilitated by the further development of protocols like OAuth, implementation of the concepts is made easier.[x]

Conclusion

Mainstream access control mechanisms like RBAC and ACL's have a long tail and will continue to have valid use cases in many organizations. However, as companies, governments, and organizations begin to require communications and collaborations outside of their traditional four walls, other ways of controlling access are required.

Mainstream access control methods are not able to deliver the growing need for flexible access control in a changing world. Modern access governance requires modern access control methods. There is a clear need for dynamic access control. Interestingly, the tools are becoming available, and implementation need not interfere with the current best practices: adaptive authentication, and PBAC can be added to an existing identity and access architecture. It takes some planning, based on a roadmap. And of course, it requires implementing elements of access governance.

Author Bio

André Koot is IAM and Security Consultant at SonicBee. His IAM experience comes from a financial accounting and auditing background. This background of anti-fraud detection and prevention business processes led to research in the area of authorization and access control principles.

Change Log

Date	Change
2020-06-17	V1 published
2021-04-19	Author affiliation change
2021-09-30	Updated definition for authentication
2022-12-15	V4 published: clarification to Policy Engine definition; minor editorial updates

[i] Wikipedia contributors, "Classified information," *Wikipedia, The Free Encyclopedia,* https://en.wikipedia.org/w/index.php?title=Classified_information&oldid=1120242140 (accessed November 24, 2022).

[ii] Davis, Shannon, “A Look at Discretionary Access Control,” blog, TED Systems, 1 December 2020, https://www.tedsystems.com/look-at-discretionary-access-control/ (accessed November 23, 2022).

[iii] Rouse, Margaret, “mandatory access control (MAC),” TechTarget, December 2013, https://searchsecurity.techtarget.com/definition/mandatory-access-control-MAC (accessed November 23, 2022).

[iv] Wikipedia contributors, "CAPTCHA," *Wikipedia, The Free Encyclopedia,* https://en.wikipedia.org/w/index.php?title=CAPTCHA&oldid=1122595810 (accessed November 24, 2022).

[v] “SCIM: System for Cross-domain Identity Management,” http://www.simplecloud.info/ (accessed November 23, 2022).

[vi] “EU General Data Protection Regulation (GDPR): Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation),” OJ 2016 L 119/1.

[vii] Kantara Initiative, “UMA Specifications,” wiki page, last updated Jul 27, 2022, https://kantara.atlassian.net/wiki/spaces/uma/pages/29229182/UMA+Specifications (accessed 23 November 2022).

[viii] ISACA home page, https://www.isaca.org/ (accessed November 23, 2022).

[ix] “Data Mining and Predictive Analytics: Things We should Care About,” Inside Big Data, 24 November 2018, https://insidebigdata.com/2018/11/24/data-mining-predictive-analytics-things-care/.

[x] Wikipedia contributors, "Classified information.”

Introduction to Policy-Based Access Controls (v3)

By Mary McKee

Please see André Koot's Introduction to Access Control for a primer on access controls.

To comment on this article, please visit our GitHub repository and submit an issue.

Table of Contents

Abstract

The natural evolution of access controls has caused many organizations to adopt access management paradigms that assign and revoke access based on structured and highly reproducible rules.

One such paradigm is known as Policy-Based Access Control (PBAC), which is most differentiated by two key characteristics:

1. Where other access control paradigms often optimize for ease of granting user access to all relevant resources, PBAC optimizes for ease of extending resource access to all applicable users.

2. PBAC facilitates the evaluation of context (time of day, location, etc.) in granting access to a protected resource. Context is used to express who may access a resource and the conditions under which that access is permissible.

Shifting the focus of access controls from the user to the resource allows PBAC systems to be particularly resilient against shifts in organizational structure or regulatory obligations. Including context (such as an authorized user's location or device) allows for additional security controls to be expressed and extended within resource permissions, ensuring that all facets of access control are contained and auditable within a single structure.

Because PBAC accommodates a very precise expression of who may access a resource and under which circumstances, it lends itself to the automation of access provisioning and deprovisioning in a way that provides ease of management as well as increased security and adaptability.

Introduction

To effectively secure resources, access control systems must be designed to adapt to rapid shifts in technology, regulatory obligations, and organizational structure. As organizations embrace more sophisticated technology and seek protection from more sophisticated threats, access management strategies are evolving to address modern concerns.

Most early access management systems utilize what we now refer to as **Discretionary Access Control** (**DAC)**. With DAC systems (such as access control lists), administrators manually assign privileges to users according to their understanding of need, appropriate use, and organizational rules. As DAC systems grow in users, resources, administrators, and/or age, their reliance on ad hoc management leads to inconsistencies in application and understanding of access. As inappropriate access often goes unnoticed and insufficient access creates visible business challenges, DAC administrators are increasingly incentivized to be liberal with authorizations and conservative with access cleanup. As a result, DAC is often too costly, too inconsistent, and too inflexible for modern needs.

Contemporary access control systems aim to promote consistency and efficiency by granting access to resources through structured rules. Perhaps the best-known model for abstracting access control so that permissions are based on rules is known as **Role-Based Access Control (RBAC)**. Through RBAC, permissions are associated with “roles” assigned to users. This model effectively ensures that users with the same responsibilities are consistently granted the same permissions. It encourages governance by requiring that roles and their associated permissions be defined before they can be used.

Further, RBAC is suitable for use in federated authorization scenarios where resource permissions depend on the information provided by an external user authority. While these are improvements over DAC, RBAC permissions are not resilient against shifts in responsibility structure within an organization and are limited in how permissions can be defined. These drawbacks, covered later in this article, make it difficult for RBAC systems to ensure that users do not have more access than they need to perform intended business functions (also known as the *principle of least privilege*[i]).

Policy-Based Access Control (PBAC) is a more robust paradigm for managing permissions through structured rules in federated or non-federated contexts.

While the RBAC model intentionally bundles permissions, PBAC builds on a concept known as **Attribute-based Access Control (ABAC)** to automate fine-grained, decoupled permissions. Leveraging ABAC’s approach of calculating permissions based on user information such as a job code or employment status, PBAC provides increased precision by supporting appropriate access conditions (or context).

Terminology

- **Access control system** – a structure that manages and helps enforce decisions about access within an organization.

- **User** or **Subject** – a person or entity who may receive access within an access control system.

- **Resource** or **Object** – an asset protected by access controls, such as an application, system, or door.

- **Action** – a protected operation available for a resource, such as “view”, “edit”, or “submit”.

- **Permission** – a statement of authorization for one or more subjects to perform one or more actions on one or more objects.

- **Context** – conditions under which an action on a resource is authorized for a subject, such as time of access, location of access, or a compliance state.

- **Federated access controls** – an access control architecture that accommodates the separation of user/subject authority and resource/object authority.

- **Discretionary access control** – a pattern of access control system involving static, manual definitions of permissions assigned directly to users.

- **Role-based access control** – a pattern of access control system involving sets of static, manual definitions of permissions assigned to "roles", which can be consistently and repeatably associated with users with common access needs.

- **Attribute-based access control ("ABAC") / Claims-based access control ("CBAC")** – a pattern of access control system involving dynamic definitions of permissions based on information ("attributes", or "claims"), such as job code, department, or group membership.

- **Policy-based access control** – a pattern of access control system involving dynamic definitions of access permissions based on user attributes (as in ABAC) and context variables for permitting or denying access.

- **Principle of least privilege** – an information security best practice ensuring that users in an access control system do not have more access to resources than is necessary for their intended activities.

- **Segment** – a grouping of subjects that may be useful for authorizations, such as full-time employees, undergraduate students, IT administrators, or clinicians.

- **Abstraction** – the practice of identifying and isolating repeated aspects of operations or business logic so that they can be maintained in one place and referenced in many places.

PBAC vs. RBAC: A Comparison

To better understand PBAC structures, it may be helpful to examine how they differ from RBAC.

While the primary focus of RBAC permissions is the user, the primary focus for PBAC permissions is the resource.

RBAC asks, "What types of users do I have, and what may they do in my environment?". Controls are constructed with **subjects** (who is getting access), **permissions** (what is being accessed or used), and **roles** (what permissions can be assigned to a subject)[ii]. This looks like:

Subject		**Role**		**Permission**
Ada	as	Editor	may	Modify Documents

PBAC asks, "What types of resources do I have, and who/how may they be used or managed?" Controls are constructed with **subjects** (who is getting access), **actions** (what behavior is being discussed), **objects** (what resource is being accessed or used), and **context** (environmental or other parameters defining acceptable access)[iii]. This looks like:

Object		**Action**		**Subject**	**Context**
Documents	may be	Modified	by	Those with "Editor" job code	On managed devices

Both examples abstract subjects to ensure that all editors are granted the necessary permission. In the RBAC example, Ada acquires the permission because she has been assigned to the "Editor" role through a manual or automated process. In the PBAC example, Ada acquires the permission because the subject definition matches her employee record, though the subject definition could also be a manual process, such as the assignment of a group membership.

To make the most apples-to-apples comparison, imagine that an RBAC system adds Ada to an "Editor" role, and a PBAC system adds her to an "Editor" group membership that is referenced in access policies. Though these actions may seem nearly equivalent, the PBAC architecture offers the following advantages: the flexibility to support different situations (context), the ability to discretely handle changes without impacting other permissions (modularity), and the capacity to handle real-time permission evaluation (symmetry). Each of these factors promotes an organizationally consistent and defensible approach to access control, as illustrated by the following examples:

Context

Ada's employer may be subject to legal or compliance concerns that affect how resources may be accessed. For example, when national security regulation (such as export controls) restricts access from certain types of devices, relevant PBAC policies can be amended to include this stipulation.

If the company requires some form of training before resources can be accessed, this too can be articulated as context. A “certification status” attribute can be maintained for Ada based on records referenced from within or outside the authorizing organization. Ada’s permissions can require that this status is current at the time of access. Instead of laborious audit processes or managing infrastructure to revoke and reassign permissions as compliance states change, Ada’s access is automatically blocked when she is not compliant with training and automatically restored when she re-certifies her training. Similarly, if Ada must consent to terms and conditions for the access she has been granted, PBAC context can ensure that this has occurred in advance of any interaction with the resource.

For security reasons, Ada may be expected to only access company resources from safe-listed network spaces or with multi-factor authentication requirements that are more stringent than those of users with lesser permissions. By codifying and enforcing these requirements within the scope of the permission, Ada’s employer can easily reference, manage, and adapt all access requirements in a single place.

Modularity

Because permissions granted by PBAC policies are not inherently interconnected as they are with RBAC, they are highly modular and easier to manage with confidence. When an organization needs to add, remove, or modify controls on a resource, policies for that resource can be adapted exactly as needed without impacting other resources.

When permissions are bundled together, as in RBAC, accommodating new business scenarios requires a broad analysis of existing permission groupings. Often, administrators are forced to choose between a “close enough” access bundle that carries unneeded permissions with it or contributing to a proliferation of bundles that become increasingly difficult to understand and maintain.

For example, if senior leadership at Ada’s company selected her to edit sensitive briefings for their investors, it is likely that she would need access atypical for editors. An RBAC system admin charged with granting this access is likely to consider solutions such as:

- Giving all editors the access Ada now needs, thus over-privileging other editors.
- Granting Ada a senior leadership role in addition to the editor role, thus over-privileging Ada.
- Creating a new role for permissions specific to this need, setting a precedent of provisional role creation for ad hoc needs.
- Re-engineering roles to offer a cleaner solution for this business scenario, typically a costly exercise.

Organizations with evolving access needs will generally not find it practical to redesign RBAC roles each time an access need is not represented by an existing role. The alternatives – over-privileging or over-complicating – promote an increasingly lackadaisical approach to access management within the organization.

Symmetry

When there is a divergence between the criteria for granting access and criteria for revoking access in a system, it is common for the system to accumulate permissions that were at one time appropriate but would not be allowed under current policy. PBAC systems are not susceptible to this permission spread because access control decisions are made in real-time based on current attributes and context.

Since PBAC is an extension of ABAC, PBAC controls easily accommodate fully or partially automated access based on attributes. An institution may wish to automatically grant access to any current employee of a company, any employee who works at Office X, or any employee who works at Office Y and is not currently on personal leave.

Automating how access is assigned simplifies the tasks of automating continuous monitoring of permission validity and revoking permissions that are no longer allowable under current rules. This creates symmetry between provisioning and deprovisioning of access, minimizing system maintenance and remnant permissions.

PBAC is Practical

PBAC scales well because it is adaptable, and this adaptability can make it a practical option for organizations of any size. Time saved with streamlined RBAC roles can be quickly lost if the business impact of modifying a role (or its many associated permissions) is unclear. This can disincentivize active and responsible management of access controls and hamper growth in an organization of any size.

To illustrate how PBAC can be preferable even in a small organization, consider the following scenario:

JE Plumbing starts as a small business comprised of five plumbers and an owner who handles all administration.

Thanks to an excellent reputation and growing customer base, the owner is able to expand the staff to twenty plumbers, who are supported by a business manager, three sales representatives, and two finance specialists.

Over time, JE Plumbing sees an opportunity to expand the company's coverage area and offerings. To accomplish this, they set up two new locations overseen by two new business managers (one of whom was an internal promotion from a finance specialist position). They grow their residential plumber staff to seventy-five and hire twenty-five commercial plumbers. Finance and sales positions are replicated across the two new offices, growing that team from two to six. A dedicated marketing specialist is hired to cover all three sites.

An RBAC approach to this problem might start with two roles: an admin role for the owner and a technician role for her staff. As the company grows, a business manager might be trusted with the admin role, but new roles would need to be created for the sales and finance specialists. After doubling from two to four roles, the role count doubles again as the company splits the technician role into commercial technician and residential technician, splits the sales and marketing role into distinct roles, formalizes roles for business managers and customer service, and retains the original admin and finance roles.

Though this example looks at JE Plumbing's development at three points in time, it is unlikely that the company would implement such broad shifts overnight. To preserve security through incremental shifts in responsibility, a small business making strategic organizational adjustments with limited working capital should consider the absence of a role not included in this exercise: that of a full-time IT professional available to perpetually re-engineer access management structures and adapt each system utilizing them.

By contrast, a PBAC approach would start by looking at what resources JE Plumbing needs to secure: work orders, customer information, invoices, inventory, employee personal and licensing information, payroll data, and expense reports. Though responsibility for these functions changes as the company adds staff, the functions themselves remain the same. If the company expanded the nature of its business in addition to the scale, permissions could easily be added to support the new functions without interfering with existing functions.

This simple shift from expressing access controls from user-focused to resource-focused allows for access control complexity to grow linearly rather than exponentially. As a result, JE Plumbing can adapt permissions in step with organizational shifts without managing a ballooning number of roles.

In addition to being more sustainable, PBAC also creates opportunities for the company to reduce risk by setting the context for access. For example:

- When technicians can see all customer information, customers are at risk of privacy violations, and the company is at risk of an employee exfiltrating that information to help them start their own competing company. Perhaps technicians need to see addresses to navigate to job sites but only need to see information associated with open jobs assigned to them. Customer service may need to see phone numbers and email addresses for all customers but may not need address information.

- Only technicians making rounds need access to job information from out of the office, so restricting other users' access to internal IP addresses is an easy way to reduce the cyberattack surface for the company's systems.

- Overexposure of work order information encourages employee speculation about how the business is being run, which can result in misunderstandings or inappropriate disclosures about operational practices.

- When technicians can be assigned to jobs at a business manager's discretion, there is a risk of a technician being sent on the job with a lapsed license. Policy-based permissioning can require valid licensing before a job assignment can occur.

Although organizations with modest access management needs may initially choose to forgo PBAC features such as context limitations on access policies, committing early to PBAC architecture for access controls allows for an organic and natural maturation of access management rules over time - whether it be to accommodate more users, more resources, and/or a more sophisticated security or risk management posture.

When RBAC is Preferable

This article has primarily compared policy-based access controls to role-based access controls due to the prominence of RBAC as an access control strategy.

Some IAM professionals may be interested in implementing PBAC controls but must work with systems that can only support RBAC. In these cases, it is sometimes advantageous to rethink institutional roles in terms of resources or specific work functions rather than permission bundles that will be difficult to adapt over time. As long as an RBAC system accommodates multiple roles for a user, it should be possible to achieve some advantages of PBAC (like modularity) within that system.

When choosing between RBAC and PBAC, it may be helpful to consider that PBAC can be constructed to behave like RBAC more reliably than the reverse. For example, an organization that prefers to think in terms of "roles" may choose to represent group memberships as such, assigning those groups to many resource permissions to the same end effect - one action results in the application of a defined set of permissions. Conversely, options for applying a notion of context to RBAC permissions are often limited.

While the increased flexibility and scalability of PBAC make it a strong choice for protecting sensitive resources, it may be less approachable for casual users of an access management system. Systems with straightforward and fairly static access controls, especially those that delegate access management to end users rather than administrators (such as those where content creators can authorize collaborators), may find that the intuitiveness of a system like RBAC is more advantageous than the flexibility of PBAC.

Implementing PBAC

The key to building a successful access control environment is accommodating changing business requirements. To promote ease and precision of access management, the system should be neither too rigid nor too abstract.

To achieve this balance in a PBAC implementation, consider the following guiding principles:

Build Reusable Components

Managing abstraction in PBAC means isolating parts of your policies that may be applicable to other policies. The most obvious place where this applies is with user segmentation.

For example, if you are constructing a policy to say that:

Object		Action		Subject	Context
User profiles	may be	Updated	by	Business managers	For full-time employees

"Business managers" and "full-time employees" are very likely to be used again in other policies. Thus, creating a definition for these segments that can be used by one or more policies is wise.

The ideal way to avoid these definitions becoming too granular and rigid is through access management system implementations that allow for set logic - particularly intersections (membership in set A AND set B), unions (membership in set A OR set B), and complements (membership in set A, BUT NOT set B).

To expand on the previous example, if the policy above requires the following update:

Object		Action		Subject	Context
User profiles	may be	Updated	by	Business managers at the Detroit office	For full-time employees ***at the Detroit office***

The best way to solve this problem is usually[iv] to keep definitions of "business managers" and "full-time employees" and add a third: "Detroit office." The "Detroit office" definition can then be used to update the subject of your policy (granting access to the intersection of "business managers" and "Detroit office") as well as a context variable (scoping that access to the intersection of "full-time employees" and "Detroit office").

This approach makes it possible to achieve the same ease of assigning a permission to a group of individuals as you might in RBAC, with the benefits of avoiding interdependence between permissions, being able to cleanly segment objects as well as subjects, and supporting specificity through permission contexts (such as user groups, device identifiers, IP address ranges, or document classifications).

Facilitate Governance and Audit

A good access control system will allow auditors and business owners engaged in access governance to understand existing precedents in organizational access controls, analyze how they may need to be extended or modified, and ascertain the business impact of proposed changes.

When designing a PBAC system, it is important to make sure that subjects, actions, objects, and contexts are stored in a way that makes it straightforward to report on access from any of these perspectives. Business owners and auditors should have easy access to reports that answer questions about access users have, users able to access resources of interest, and allowable contexts for any actions defined for a resource.

The expressiveness of PBAC permissions makes it realistic to define all access considerations within policies. This flexibility is advantageous over implementing additional security measures (such as IP restrictions) outside of an organizational access control system. It allows for a single source of truth about circumstances under which access is allowed.

Being able to report on permissions in this way facilitates the examination of current rules for access to a resource. Good reporting may also include users who currently meet these criteria. Though PBAC is often used in federated contexts where identity (and other contextual) information for all potential users is not available to the resource administrator, such user reports can be helpful for spot-checking, especially in the context of a proposed change. Reports on who would gain or lose access under a proposed policy support business owners and auditors in refining controls to best facilitate organizational needs and security.

Embrace States over Events

Business processes are often developed with flowcharts, which are focused on events. This often leads to access management systems that are implemented on events that mimic flowcharts, such as assigning access when a new employee is hired.

Being based on observable attributes, PBAC policies tend to be more focused on states, such as an employee's current position. This offers several advantages:

- **Fewer states than events:** Access provisioning that is triggered when an employee first enters a position may need to account for nuances between external hires, internal transfers, and promotions. Unexpected events may occur, such as a canceled termination. Rather than tracking all potentially relevant business events, an access policy can simply apply to anyone currently holding the position.

- **Local process changes:** Access management teams are much more likely to be informed of changes to relevant states (e.g., employment, company policy, business functions) than to changes to events (e.g., how many processes can be used to hire staff, changes to the company network, infrastructure upgrades, etc.).

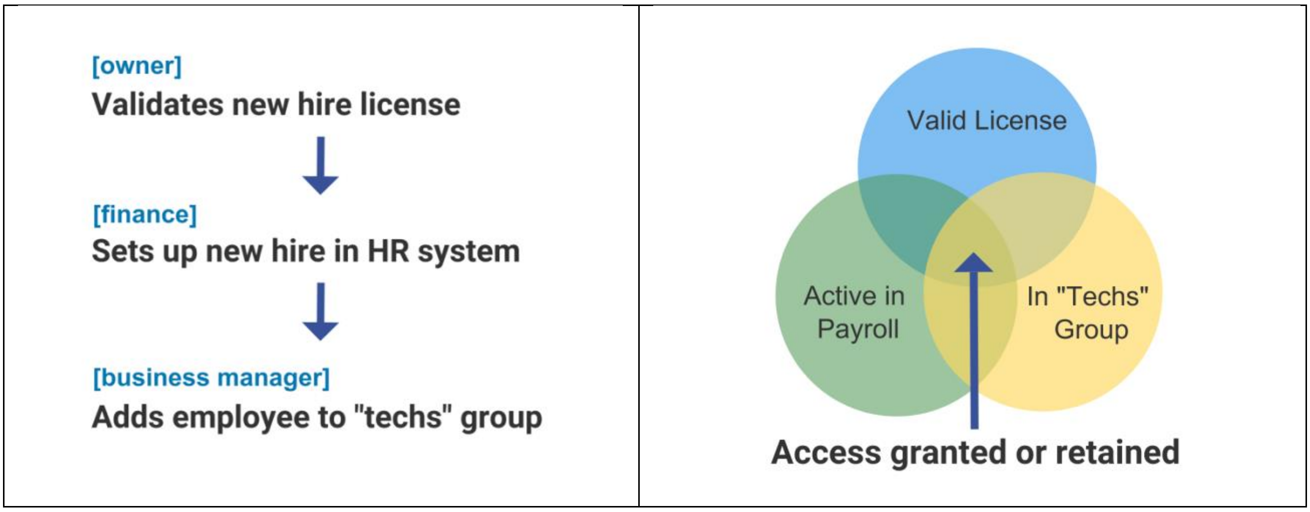
  When departmental processes shift in ways that affect the detection of events driving access, access management teams become responsible for investigating the resulting inconsistencies and may not be confident that their systems are functioning as intended.

- **States are more reconcilable:** Events occur at a point in time, which makes them more difficult to audit for appropriateness. For example, someone might have access through a legacy process that has since been revised (and should retain access) or because a deprovisioning was attempted (and should lose access) but was not completed. Without a current policy to compare against, it becomes very difficult to determine whether existing permissions are appropriate, further eroding trust in the system.

  Because states are continuously observable, compliance with policies defined by state can be easily validated, and the impact of proposed changes to such policies can be easily measured.

To workshop access rules that can generate robust PBAC policies, consider dropping the flowchart arrows and working only with circles representing conditions. Arranging these circles as a Venn or Euler[v] diagram allows for a discussion of acceptable conditions for access that will result in cleaner and more robust policies.

Event-based Permission Design	**State-based Permission Design**
Looks like: Flowcharts **Results in:** Rigid and sequential workflows, point-in-time validation, complicated deprovisioning logic.	**Looks like:** Overlapping circles **Results in:** Flexible and parallel workflows, continuous validation, harmony between provisioning and deprovisioning.


[owner]
Validates new hire license
[finance]
Sets up new hire in HR system
[business manager]
Adds employee to "techs" group
Valid License
Active in Payroll
In "Techs" Group
Access granted or retained

Support Separation of Concerns

More advanced guidance around PBAC may include references to standards such as OASIS' eXtensible Access Control Markup Language (XACML)[vi]. Such standards can be particularly useful when it is desirable to maintain separation between components of a PBAC system, such as federated systems, or when policies are based on sensitive data.

Consider the example of a scientific instrument subject to federal law requiring all users to be either a citizen or legal permanent resident of their country, and additionally with a clean background check performed within the last three years. To enforce this policy without exposing sensitive information like citizenship, immigration status, and background check results to the instrument, the managing organization could implement a separation of policy evaluation and policy enforcement such that the source systems for this data send the instrument a compliance status rather than the raw information needed to make a local access decision. In federated contexts, similar approaches are useful for reducing sensitive data exchange across organizational boundaries.

Conclusion

Access control systems promote and implement an organization's access control strategy as changes occur in users, personnel, responsibilities, organizational structure, and legal obligations. Most failures with access management are due to a system implementation that is too manual to scale or too brittle to adapt to changing business needs without costly and time-consuming re-architecture efforts.

While it is common to try to optimize access control systems for efficiency in *granting* access, a truer measure of a robust access control system is how reliably it can *revoke* access. Policy-based access controls support the security principle of least privilege by offering logical symmetry between access assignment and revocation. Defining policy for access allows access to be dynamically evaluated for validity and automatically revoked or reported as soon as that access becomes invalid under current policy.

Developing access controls from a resource-first perspective and adding a notion of context to these controls allows PBAC systems to maximize resource security over convenience of access assignment. While these systems can initially be more complex than other approaches, the atomic nature of policies and their relative resilience against the buildup of legacy permissions makes for a system that is much more maintainable over time as compared to more limited rule-based access management systems like RBAC.

Author Bio

Mary McKee began her career as a web application developer, eventually specializing in and leading teams dedicated to maturing processes in Identity Management and Cybersecurity. She now works as Senior Director of Engineering at Cirrus Identity.

Acknowledgments

The author would like to thank André Koot and Andrew Hindle for their thoughtful responses to earlier versions of this article, and Heather Flanagan, Christienna Fryar, Dave Wible, and Mary Ellen Wible for their feedback and support with its development.

Change Log

Date	Change
2023-10-27	V3 published; clarification to Embrace States over Events, Support Separation of Concerns, and author bio
2022-06-03	V2 published; Clarified scope as an introductory article; replaced section on static access controls; removed section on privacy
2021-04-19	V1 published

[i] "Least Privilege," https://us-cert.cisa.gov/bsi/articles/knowledge/principles/least-privilege (accessed February 10, 2020)

[ii] "Role-based access control," https://en.wikipedia.org/wiki/Role-based_access_control (accessed February 10, 2020)

[iii] "Attribute-based access control," https://en.wikipedia.org/wiki/Attribute-based_access_control (accessed February 10, 2020)

[iv] The examples in this section are meant to illustrate optimizing for set math capability within a context where both the identity provider (or user attribute store) and the service provider (or resource to be protected) exist within a common environment, and does not extend to federated contexts where a service provider may be interacting with one or more externally controlled identity providers. It is, however, worth noting that PBAC (/ABAC/CBAC) can easily accommodate these externalities.

[v] "Euler diagram," https://en.wikipedia.org/wiki/Euler_diagram, (accessed February 25, 2020)

[vi] "eXtensible Access Control Markup Language (XACML) Version 3.0 Plus Errata 01," https://docs.oasis-open.org/xacml/3.0/errata01/os/xacml-3.0-core-spec-errata01-os-complete.pdf (accessed May 20, 2022)

Strategic Alignment and Access Governance

By André Koot

To comment on this article, please visit our GitHub repository and submit an issue.

Table of Contents

Abstract

In today's digital age, for an organization to succeed, it must have a strong IT function. That IT function will not be at its best, however, if it is missing a close partnership with the business components of the organization. In many organizations, IAM is seen as an IT responsibility. While some IAM-related tasks and activities can be considered IT-related, others are not. Without a clear understanding of the different tasks and responsibilities in the field of IAM, the success of IAM-related programs will be limited.

This article argues for the need for explicit strategic alignment, also referred to as business-to-IT alignment, between IT efforts around IAM, particularly access management, and the business needs of an organization. Lack of this type of alignment leads to failed IAM projects and blocked business maturity growth.

Introduction

Many Information Technology (IT) departments are responsible for implementing IAM systems to support an organization's efforts to operate efficiently and effectively. Identity management systems are designed to automate the joiner, mover, and leaver processes (JML processes) for employees.[i] Access management systems, in turn, are designed to make it possible to request and grant authorizations in information systems and even physical access to facilities such as buildings or data centers. For IT to support the necessary processes and controls, they must understand the business drivers for the organization. IT in general, and IAM in particular, must serve the organization; strategic alignment is critically important and, unfortunately, challenging. Different day-to-day languages, cultures, and priorities obstruct the understanding on both sides regarding what has to happen and why for the business to succeed.

Terminology

- Alignment: the synchronization rate of processes and environments
- Governance: making sure that accountable owners are demonstrably in control
- Identity Governance and Administration: a solution for automating user management and authorizations in target systems, building on the organization's customer and human resource processes.
- Joiner-Mover-Leaver processes: The joiner/mover/leaver lifecycle of an employee identity considers three stages in the life cycle: joining the organization, moving within the organization, and leaving the organization.[ii]

Acronyms

- CEO: Chief Executive Officer; CFO Chief Financial Officer; CRO Chief Risk Officer; CTO Chief Technology Officer; COO: Chief Operations Officer
- RBAC: Role-Based Access Control
- IGA: Identity Governance and Administration
- JML processes: joiner, mover, and leaver processes

Understanding Strategic Alignment

Business-to-IT Alignment, also known as Strategic Alignment, has been studied since the 1980s. Following the Henderson and Venkatraman model, strategic alignment brings together a dynamic integration of IT planning and business development to shape or enable a holistic business strategy.[iii]

Ideally, IT enables the business to perform efficiently and effectively. IT can help solve business issues by providing logical, structured ways of working, integrating solutions, and making access and application integrations possible. For example, IT supports automating manual tasks, keeping records, integrating different information processing components and systems, and following security best practices. IT better understands what problems need to be solved when aligned closely with the organization's business drivers. In general, businesses are more successful when they incorporate the efficiencies IT can bring to the table.

In order to reach the necessary levels of strategic alignment, we first must consider the barriers. Often, the language used by the business to identify what's important is quite different than the language used in IT.

Business talks about		IT talks about
Customer satisfaction		System service level agreements (e.g., 99.999% availability)
Return on Investment (ROI)		Network architecture (e.g., hybrid, cloud, on-prem)
Legal and regulatory requirements (e.g., GDPR, CCPA)		Common Vulnerability and Exposure (CVE) Announcements[iv]
Market share		Latest container management technologies (e.g., Kubernetes)
Earnings before interest, taxes, depreciation, and amortization (EBITDA)		Access control mechanics (e.g., -rwxr-xr-x)
Financial bottom line (i.e., General ledger)		Network capabilities (e.g., bits per second, database structures)BPS
Interest rates		Data Center architecture and computing clusters
Consumer trust and business reputation		P1 (Priority 1 incidents)

(There is no implied horizontal correlation between the terms in the left and right columns).

Alignment Models

There are different methodologies that describe the necessary points of communication to support strategic alignment. Hendersen and Venkatraman, two IBM fellows, came up with this model for strategic alignment in 1993:[v]

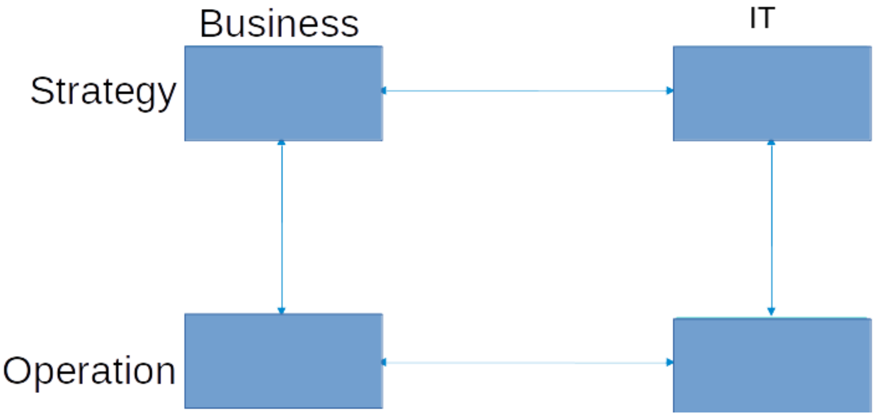


Figure 1: Simple Model for Strategic Alignment

This model suggests that business and IT stakeholders should communicate on both the strategic and the operational levels. This multidirectional communication ensures that the business processes are supported by fitting IT solutions. By pairing strategic choices with operational ones, the organization can minimize unnecessary changes in process and technology. For this model to work, however, the organization must address the fact that IT and the business often have different ways of working, cultures, languages, and jargon. These differences make strategic alignment difficult.

One critical characteristic of this model (and in the other models presented) is that communication between domains/cells can only occur across the horizontal and vertical lines, not diagonally. That means communication can only happen in formalized relations to prevent disrupting formal, mature procedures.

> *Case CEO:*
> *My old CEO was tempted to get a smartphone. All young marketers used those devices, so why not the CEO? But he also wanted to read his company email on the same smartphone. This expectation would not be a problem except for the fact that in 2008 enterprises were not supporting those devices in a standard way. The CEO directly ordered an IT engineer to make it possible: install the app, connect to the mail server, create a secure channel to the Internet, add certificates, etc. This non-standard change interrupted IT operations for three months.*

In the Amsterdam Information Model by Professor Rik Maes, Dr. Maes added additional components to implement information management and structure.[vi] :

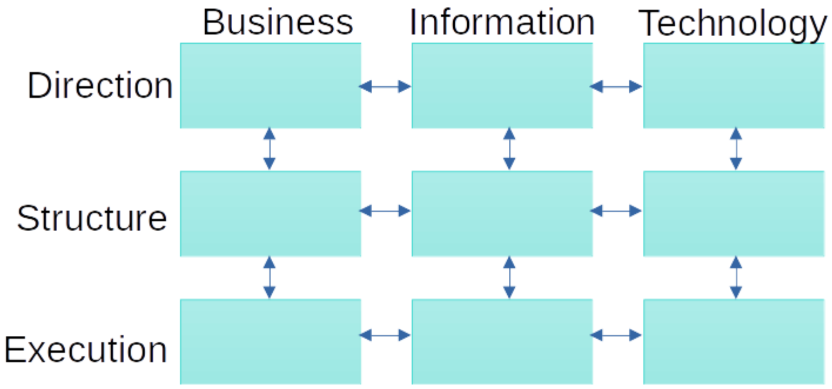


Figure 2: the Amsterdam Information Model for Strategic Alignment

The middle column, Information management, translates the business requirements into IT solutions (left-to-right translation). It also translates the features and functionality of IT components (platforms, services, applications) into business opportunities (the right-to-left translation). The information management function must overcome the issues indicated above, such as language and cultural differences. The information manager (or CIO) should understand and know how to converse with businesspeople and IT personnel. The information manager should be able to connect to the entire organization and act as the missing link in business-to-IT alignment.

The added horizontal middle layer also has a specific 'translation' role:

This layer can be seen as the architecture layer. It translates strategic concepts into day-to-day operations. Looking at the different columns within this layer, from left to right, we can identify the following architectural concepts:

- Business architecture (organogram/org-chart and business processes models, including Segregation of Duties (SoD), abuse of information prevention controls, etc.).
- Information architecture (data models, -flows, and interfaces).
- The IT architecture (including servers and networking, containerization, cloud, and security architecture).

In this model, we can position the CEO, CFO, and COO in the top-left area. These persons are accountable for defining the organization's business strategy, direction, and course. The head of IT, or CTO (Chief Technology Officer), would be positioned in the top-right area, accountable for IT strategy, like sourcing strategy and IT vendor management strategy. This assignment leaves the CIO in control of the middle column, responsible for the business-to-IT alignment.

Governance, ownership of control, would, in this model, be owned by the top-left area players.

IAM and Alignment

So far in this article, we have focused on the IT/business relationship in general. As IAM is traditionally considered part of IT, the challenges of strategic alignment are at the core of most failures of IAM projects. In many cases, IAM is very much an IT function. IAM includes basic "techie" tasks such as password resets, account management, user provisioning, and so on. IAM, however, is arguably more closely tied to business needs than any other aspect of IT. Authorization processes, in particular, regularly bridge the gap between IT operations and business requirements.

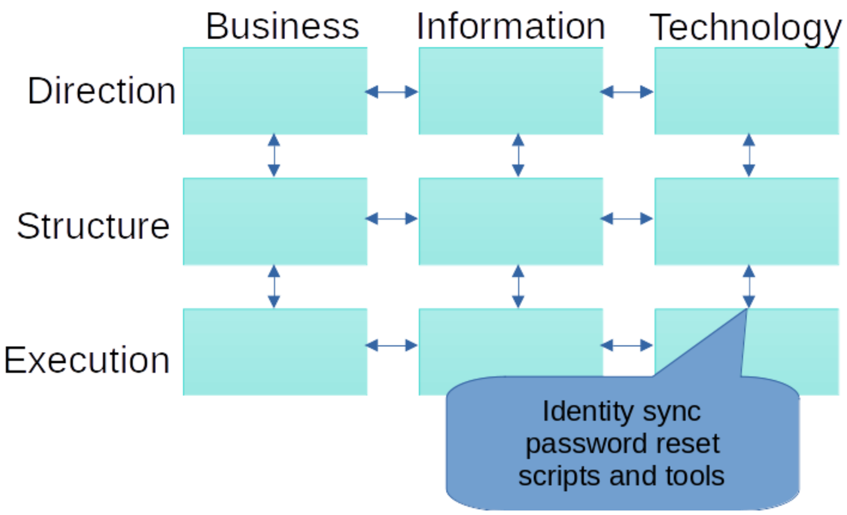


Figure 3: Amsterdam Information Model - IAM as an IT Function

IAM started as an IT responsibility. Creating interfaces and connectors, protocols, and adding certificates all fell in the realm of IAM and IT. The trigger for all identity transactions was often the HR department, but in daily operations, identity management belonged to IT as part of the general task of automating business processes. That has not changed. Most identity management in an organization is still seen as IT: bottom right.

Authorization management, on the other hand, is not as easily plotted. Authorization involves "determining a user's rights to access functionality with a computer application and the level at which that access should be granted. In most cases, an 'authority' defines and grants access, but in some cases, access is granted because of inherent rights (like patient access to their own medical data)."[vii] Authorization is directly tied to business practices, and yet the IAM group generally implements them.

Using the Amsterdam Information Model, we can identify where authorizations are most prominently defined. Authorizations are enablers for performing tasks in an organization and so are critical to the execution phases. Authorizations are derived from the organizational structure and business processes. Implementing authorization management must therefore be plotted on the Business Structure area in the model. For example, SoD rules are defined in a business process: one person may not be allowed to perform multiple successive tasks because that could create a risk of fraud, abuse of permissions, or data breaches. Tasks are defined in a process. That means a process owner, 'mid-left,' is accountable for defining these specific access control policies.

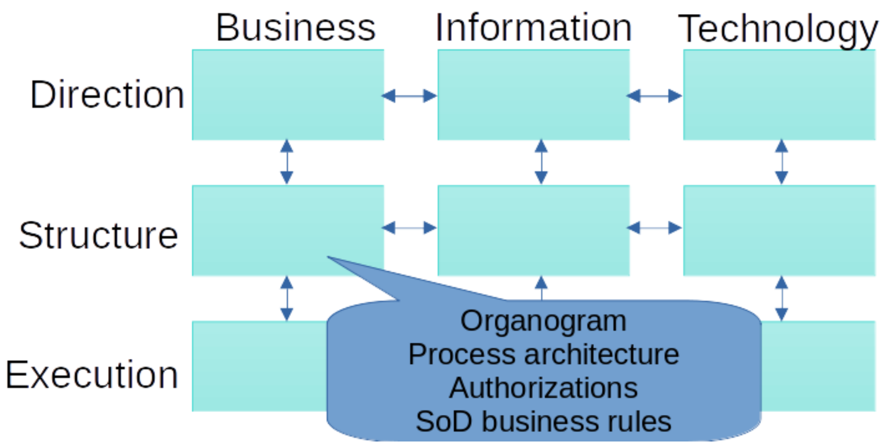


Figure 4: Amsterdam Information Model - Authorization as a Business Function

IT does not own or manage business structure authorizations. It's the responsibility of the 'business' owners, specifically the process owners, line managers, or data owners.

Managing authorizations–defining, granting, and revoking them–is one of the more challenging tasks for any organization. This task is where the concept of RBAC became handy. The concept was created in the mainframe era in solutions like IBM's Resource Access Control Facility (RACF) and the Access Control Facility 2 (ACF2) system. In the local area networking era, RBAC became the solution for managing this authorization complexity. In the nineties, dedicated identity management solutions started to appear, with authorization solutions exploring the concept of RBAC coming into existence at the turn of the century. These solutions evolved over time, eventually offering identity governance by adding attestation/recertification processes.

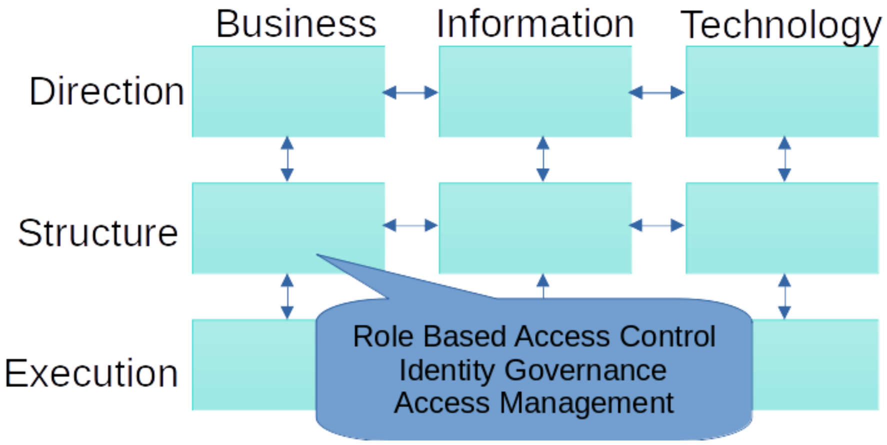


Figure 5: Amsterdam Information Model - RBAC and Identity Governance

These days, we see vendors moving to a spot in the center. Traditional Identity Management software vendors add authorization management solutions and traditional identity governance vendors add identity and workflow management capabilities. There are also 'new' entrants in the market, offering cloud-based solutions such as Identity Governance and Administration offerings.

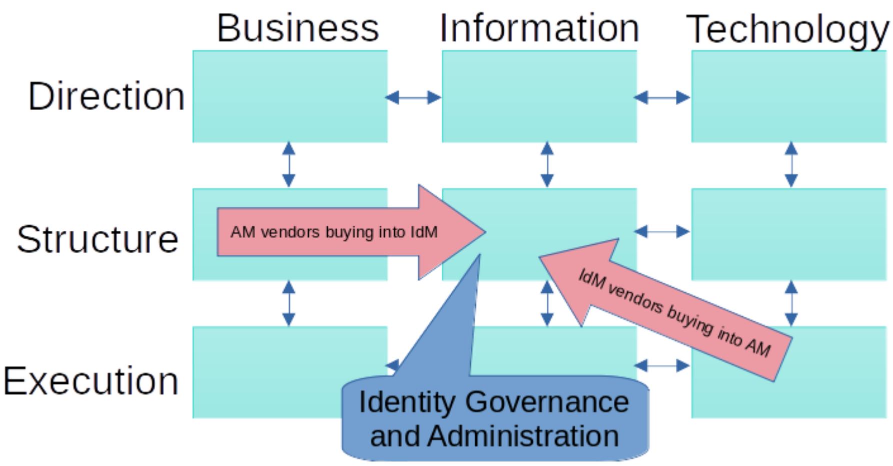

Business
Information
Technology
Direction
Structure
Execution
AM vendors buying into IdM
IdM vendors buying into AM
Identity Governance and Administration

Figure 6: Amsterdam Information Model - IGA

When evaluating Attribute Based Access Control and Policy Based Access Control models, the same strategic alignment change of responsibility can be seen. Several IT-oriented access control policies exist, such as the requirement to use TLS certificates and zero-trust networking. But other access policies are business oriented. Policies like SoD or privacy-related consent management have a clear relation to the business structure sector in the model.

An Extended Case Study

Information systems were generally developed to support the identity management process and to support authorization management; the current generation of IGA solutions performs their role admirably by supporting the business with reliable identities (based on the HR identity lifecycle) with reliable authorizations. And yet, there still is the issue, IAM is still seen as an IT responsibility. Let me explain this in a case:

> *Case Study - Accountability vs. Responsibility*
> *A financial institution supports its identity governance and RBAC requirements by using a modern IGA solution. The system is integrated within the IT landscape and connects several business applications for provisioning and reconciliation.*
>
> *An external auditor reported a high-risk issue concerning authorizations in the financial accounting system to the CEO.*
>
> *The CEO (Top-left) forwarded the findings to the CTO (Top-right), as the finding was about a system, and so the CEO believed IT had to solve the issue. The CTO forwarded the finding about the authorizations to the IGA product owner in the IT Service delivery department (Bottom-right). Unfortunately, the product owner cannot solve the issue.*
>
> *What went wrong?*

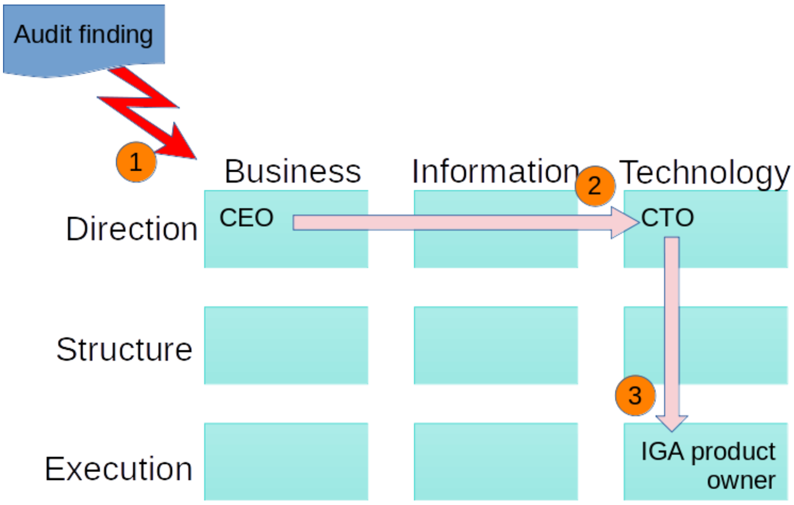


Figure 7: Amsterdam Information Model - default IAM communication

The product owner is responsible for the IGA system but not for the authorization decisions themselves; the product owner cannot fix the issues found by the auditor. In short, the product owner is responsible but not accountable for authorizations. Instead, the process owner for the financial business process should be tasked with resolving the issue.

Note that, based on the Amsterdam Information Model, there is no direct communication between the IGA product owner, who works at the operational level within IT (bottom-right), and the business process owner (center-left) in the business architecture layer. That communication would be a diagonal link and would interfere with regular, well-structured operations.

The advice was for the product owner to escalate back vertically to the CTO on the basis of lacking accountability. The CTO should then advise the CEO to assign the issue to a business process owner:

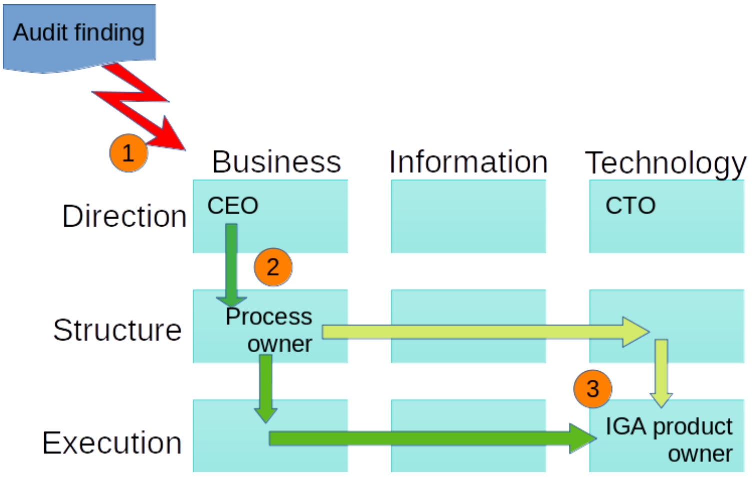


Figure 8: Amsterdam Information Model - Correct Communications Path

(Different paths for the necessary communication could be followed to make the required adaptations to the authorization model in the IGA system.)

The Way Forward

How do these models solve the issue of lack of stakeholdership in organizations? Does the alignment strategy solve the governance challenge?

First and foremost, the theory can demonstrate that access control, or authorization management, is not an IT responsibility. The 'business' is accountable for structuring and implementing authorization models and authorization management. IT can, at best, only support the business by implementing the tools that might help.

This makes the implementation of IAM a new challenge. Implementation is not just an IT project. Implementing an identity management solution can be done in an IT project style, but authorization management is not a project. Authorization management is the never-ending responsibility of managers and (business) owners.

And that leads to this conclusion: An IAM project cannot exist as an IT project. Implementing authorization management results in or requires organizational change and is therefore related to regular governance and control of business responsibilities.

Access Governance is what connects the business governance and control challenge to the IT solutions that are used to enable the organization to execute its mission. The easiest way to activate the business is to find someone who makes a decision on the topic of SoD or find someone who is a stakeholder in the approval process for access requests.

> *Case Study: SoD rules*
> *A financial institution is using a modern IGA solution to manage accounts and authorizations in Active Directory and miscellaneous information systems. This system depends on the concept of SoD. Using the SoD controls, it is impossible to assign two or more conflicting roles to the same employee. There are over 1200 SoD rules in the IGA system.*
>
> *When asked who had defined those SoD rules, the product owner in the IT department had no idea. While the product owner is responsible for making sure the system runs as expected, holding them accountable for the SoD rules is outside their area of responsibility; they may not even know all the parties involved in making those decisions.*
>
> *In an ideal world, the SoD rules would not be applied without an accountable business owner clearly identified. In this case, the financial institution has a large business project ahead of them to ensure the appropriate process owners have reviewed each rule.*

A good practice would be only to create roles and (business) rules if a person in the business domain can be assigned as the accountable stakeholder for the role or rule. Governance is not just relying on IT departments to solve issues but having someone accountable for managing the business and implementing the controls to manage the business.

Conclusion

In today's digital age, for an organization to succeed, it must have a strong IT function. That IT function will not be at its best, however, if it is missing a close partnership with the business components of the organization. The different parts must pull in the same direction to succeed.

IAM projects can only succeed with a strong business-to-IT alignment. As evidenced by the challenges associated with the organization-wide responsibilities around authorization management, IAM, perhaps more than any other IT-related function, must understand the needs of the business and enable those requirements in the identity systems.

Both parties are responsible for ensuring strategic alignment across the organization, being aware of and working to overcome the barriers of different cultures and jargon in each group.

Acknowledgments

The author wishes to thank IDPro Principal Editor Heather Flanagan for turning the original, hardly English language, text fragments into readable text.

Author Bio

André Koot is Principal Consultant at SonicBee in Amsterdam, The Netherlands. He is a member of the BoK committee of IDPro. André has over 30 years of experience in information security and over 20 years of experience in Identity and Access Management. He has a background in financial accountancy and business economics.

[i] Cameron, A. & Grewe, O., (2022) "An Overview of the Digital Identity Lifecycle (v2)", *IDPro Body of Knowledge* 1(7). doi: https://doi.org/10.55621/idpro.31

[ii] Bago (Editor), E. & Glazer, I., (2021) "Introduction to Identity - Part 1: Admin-time (v2)", *IDPro Body of Knowledge* 1(5). doi: https://doi.org/10.55621/idpro.27

[iii] Henderson, John C., and N. Venkatraman. "Strategic alignment: a process model for integrating information technology and business strategies." (1989), https://dspace.mit.edu/bitstream/handle/1721.1/49138/strategicalignme1989hend.pdf, and Dampney, C. N. G., & Andrews, T. B. (1989). Striving for sustained competitive advantage: the growing alignment of information systems and business. CSIRO Australia Division of Information Technology.

[iv] https://cve.mitre.org/

[v] Strategic alignment, Henderson and Venkatraman, 1993, reprint at

https://www.researchgate.net/figure/The-Henderson-and-Venkatraman-strategic-alignment-model-Reprinted-from-Henderson-JC_fig2_220220710

[vi] Amsterdam Information Model, 1999, reprint at https://www.researchgate.net/publication/242321998_A_Generic_Framework_for_Information_Management

[vii] Flanagan (Editor), H., (2022) “Terminology in the IDPro Body of Knowledge”, IDPro Body of Knowledge 1(9). doi: https://doi.org/10.55621/idpro.41.

Techniques To Approach Least Privilege

Matthew K. Carter

To comment on this article, please visit our GitHub repository and submit an issue.

Table of Contents

Abstract

This article will describe the lifecycle and techniques that access control practitioners should consider as they grant, validate, and refine permissions as they iterate toward least privilege. The article will compare just-in-time (JIT) approaches with long-standing permissions, balancing productivity with security. The article will explore the risks of using historical data to refine permissions. The reader will learn about refining least privilege in the context of an identity lifecycle and for a specific activity. The article will be agnostic in terms of cloud, hybrid and on-prem, as well as tools.

Introduction

Reducing excessive permissions is a continuous effort. Workforce members accumulate permissions throughout their employment, and job requirements change regularly. People take on temporary assignments, and organizations are typically better at granting permissions than taking them away. SaaS and IaaS providers are constantly changing the surface area of permissions that customers need to manage. It is a challenging balance to give employees, partners, and customers a sufficient level of privilege to digital resources without leaving an organization open to risk. The principle of *least privilege* is a hypothetical, best-case scenario of a human or non-human actor having only the permissions required to perform a task at the time it needs to be performed. Understanding techniques to create and refine permissions can help you approach least privilege and reduce the risk of an overly-permissive posture.

This article will discuss least privilege in the context of identity lifecycle and building policy for specific activities. We will examine the advantages of long and short-term permission assignments, considering techniques like just-in-time (JIT) permissions. We will utilize roles as a way of grouping together permissions related to identity and activities. This utilization is a natural extension of Role-Based Access Control (RBAC), though not all organizations use roles to model permissions in the same way. Roles provide a natural way to encapsulate multiple permissions to reduce maintenance versus assigning multiple permissions to a human or non-human principal. We will contrast least privilege applied to RBAC and Policy-Based Access Control (PBAC), but roles will be the primary mechanism for grouping permissions in this article.

Terminology

Least Privilege - "The principle that a security architecture should be designed so that each entity is granted the minimum system resources and authorizations that the entity needs to perform its function."[i]

Account Takeover - Account takeover is a form of identity theft and fraud, where a malicious third party successfully gains access to a user's account credentials.[ii]

Access Certification - Certification is the ongoing review of who has which accesses (i.e., the business process to verify that access rights are correct).[2]

Privileged Access Management - A mechanism for managing temporary access for accounts with high-risk permissions. PAM often involves check-out and check-in of a credential generated for a single use.

Just-in-time (JIT) access - a technique where a credential or a permission is granted to a principal for a temporary timeframe when they need the permission to perform an activity. Access is revoked once the activity is complete, limiting its usage.

Zero Standing Privilege (ZSP) - a state where JIT access is used for all permissions and no long-standing permissions are assigned to principals.

Cloud Infrastructure Entitlement Management (CIEM) - a categorization of technologies focused on managing the granting, verification, and refinement of permissions for cloud and hybrid technologies. CIEM is often seen as a component of Identity Governance and Administration (IGA).

Infrastructure-as-code - the process of managing and provisioning computer data centers through machine-readable definition files rather than physical hardware configuration or interactive configuration tools.[iii]

Least Privilege in the Identity Lifecycle

Least privilege can be applied at every stage of the identity lifecycle. Birthright entitlements should be continuously refined to help new employees to the workforce (joiners)[iv] be more productive on their first day while not giving excessive permissions that an inexperienced employee could accidentally misuse. Employees who change jobs (movers) inherit new permissions. They may require a ramp down of their previous job's permissions during their transition, which can cause delays in permission revocation until the transition is complete. These delays can put companies at risk of violating the principle of separation of duties (SoD) if the new job permissions create a toxic combination with the previous job role. Departing employees (leavers) still need limited access to company assets, such as access to paystubs and W-2s. Ensuring the former employee's post-employment credential has limited permissions may avoid damages.

One misconception is that striving for least privilege in the workforce is due to a lack of trust in employees. Least privilege actually protects employees and employers by limiting their respective exposure. A new employee is often granted a set of birthright permissions based on their job assignment. The permissions that are available to that employee should be continually refined to add or remove permissions to better align with employee needs as they progress in their tenure. A surplus of permissions can result in exploitation. An employee is more likely to notice an *account takeover* if they are actively using a permission, as they are more likely to observe changes to the resource.

In order to align the assigned permissions with the ever-shifting target of least privilege, organizations need to continually refine permissions granted through birthright and access requests. If these birthright permissions are managed through roles, the roles need to be analyzed for excessive permissions. If the roles do not apply consistently to the principals that the roles are assigned to, the roles should be refactored so that a role is representative of the activities that the principal needs to perform. A deficit of permissions will often cause productivity loss, so the risk of each permission needs to be evaluated to find the balance.

Self-service access requests can incorporate least privilege approaches to ensure that temporary lifespans for entitlements are used for one-time actions. Long-standing permissions granted through self-service access requests are reviewed during access certification along with birthright permissions to refine permission, regardless of when the permission was granted. Temporary access might involve Privileged Access Management (PAM) or JIT permission techniques described below.

During the *Access Certification* process, employees review who has access to resources. One guiding concept is removing unnecessary permissions that might create risk for an organization. This concept is one dimension of least privilege, where human and non-human entities are evaluated for what each has access to. Managers and application owners are responsible for refining permissions to find the balance between productivity and security. This risk evaluation is what Access Governance solutions are built to achieve. *Cloud Infrastructure Entitlement Management* (CIEM) solutions also provide tools to help refine permissions for workforce employees.

Unused permissions do not equate to unneeded permissions. Some activities are less frequent than a quarter, such as accessing tax documents, so avoid refinement based on static periods. Some activities may be less frequent than a year, such as activating a contingency plan, though hopefully, your company is rehearsing your business continuity planning.

Least Privilege for Activities

An activity, in this context, should be thought of as a set of resources and actions to perform a task. As an example, say you need to manage permissions for an *infrastructure-as-code* (IaC) process that creates multiple digital assets of different resource types to create an application. You also need to manage permissions to operate this new application after deployment. The inclination to execute the IaC process as "Admin"[v] is understandable, as introspecting and defining governing policies for an unfamiliar set of resources and actions can be time-consuming. However, the temptation to continually operate as a privileged user can result in long-standing over-permission that can be targeted by unauthorized privilege escalation.

An activity is often broken up into more granular actions and resources that are governed by the authorization system. For our IaC example, the process might contain create, modify, and delete actions for computing and data sources to set up and tear down the application. We will only consider the coarse-grained action-resource permissions in this article, for example, "create-compute" or "modify-database."[vi]

Two techniques for building least privilege roles for activities are **fail-then-add** and **record-then-replace**. Each technique provides a different balance between security and productivity by limiting the usage of privileged access.

For the **fail-then-add** technique, the infrastructure-as-code (IaC) process starts with no permissions. The IaC process is run, and when it fails due to authorization, that permission is granted. This sequence is repeated until the IaC process runs to completion. While this brute force approach may seem inefficient, the artifact role that it produces can be used for subsequent runs of the IaC process and reliably achieves least privilege for this activity. In order for the technique to be viable, you must have a clear feedback mechanism for the needed permission and transactional rollback capability. This technique also requires the practitioner to have a clear understanding of the required activities. Loosely adding permissions without a good understanding of the activities will lead to privilege creep, as revocation of superfluous additions rarely occurs after getting things to work.

The preferable second technique is a **record-then-replace** approach, where the IaC process starts with a privileged role like "Admin" that allows all actions for every resource type in the IaC process. An event is recorded for each action taken by the IaC process via a mechanism like audit logs. Once the activity completes, you can extract the actions from the recorded events and assign the necessary permissions to a new "least privilege" role. Subsequent runs of the IaC process are performed with the new least privilege role, replacing the privileged "Admin" role. Using this new least privilege role gives you an activity-specific role that can be used for other principals.

Basing least privilege on historical events like audit logs has a potential downside of incorporating unrelated or unauthorized permissions into the least privilege role if the unrelated or unauthorized activity is ongoing with that principal when the recording takes place. Check your recorded permissions to verify that extraneous permissions haven't crept into your least privilege role.

It's important to separate out setup and destroy activity from operational activity. Setup and destroy are activities that may include privileged permissions that are excessive for human actors once the non-human activity is complete. For our IaC example, creating the compute and data storage, then modifying its policy may be a setup activity, while running queries and mutations are operational activity. Setup permissions are limited for the non-human IaC process. When recording operations from your audit logs, stop the recording after setup to define the setup role. This prevents the modify-policy permission from being included in the operational role, leaving only query-data and mutate-data. An operator with modify-policy could grant themselves permissions, thereby violating the principle of least privilege.

Work with your digital resource providers to set up notifications of changes to permissions. If your role contains any kind of permission set based on expressions like wildcards that allow new permissions to be automatically included, a change in resource permissions could introduce new risks and push you further away from least privilege.

Just-in-Time Permissions

Let's consider the time component of least privilege. In general, a user principal having temporary access is more secure than long-standing access for the same permission. You will approach least privilege by only having the permission to execute an activity at the point-in-time the activity needs to be performed. Concurrent refinement of unnecessary permissions and a JIT approach to granting permissions bring us closer to least privilege. Keep in mind, however, that the overhead of managing temporary access and the productivity tax of having to ask each time may not make JIT a fit for every organization.

In a long-standing model, even if the role permissions are refined over time, the principal's effective permissions track with the role's permissions. The principal has the permission when they need it as the permission persists through the role assignment.

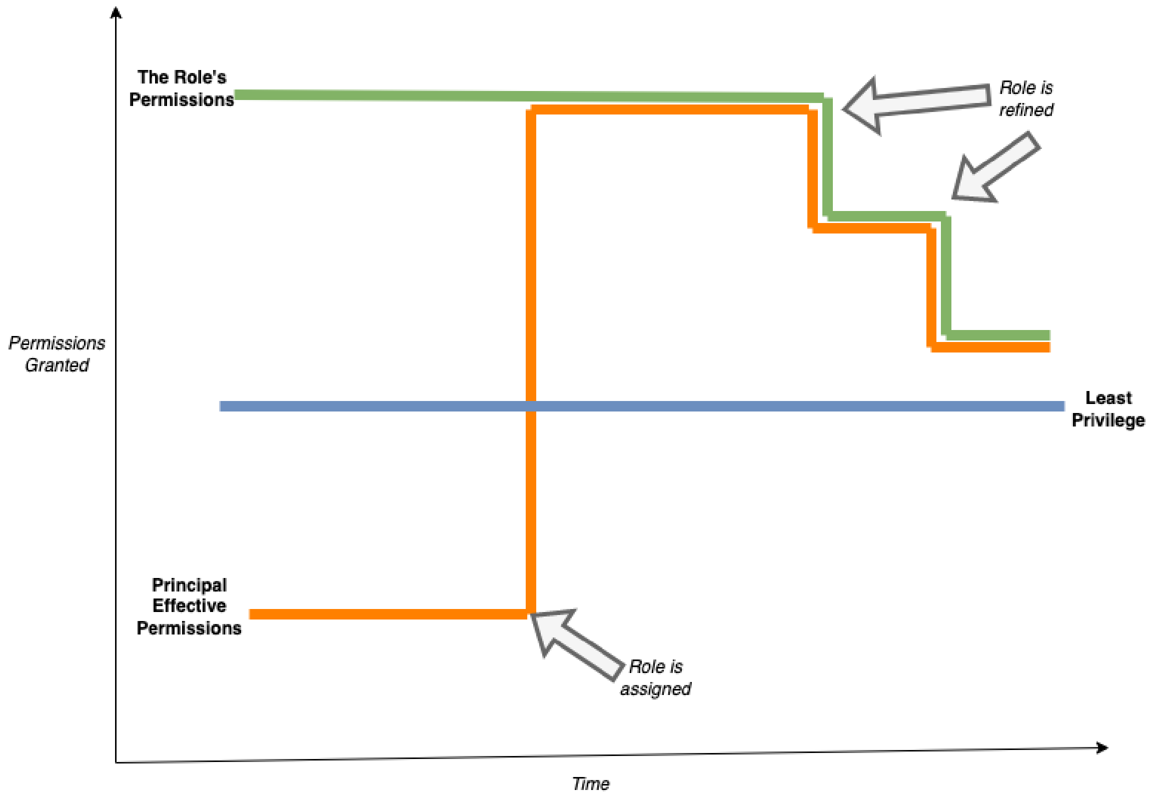


Figure 1:Long-standing Least Privilege Model

Least privilege is an activity that must be evaluated at specific points in time when a principal must take an action or access protected information. In a JIT model, permissions are granted only for the period that they are needed to perform the activity, then are revoked. By separating temporary privileged access from long-standing role assignments and the permissions granted by the roles, there is less creep of excessive permissions for those long-standing roles.

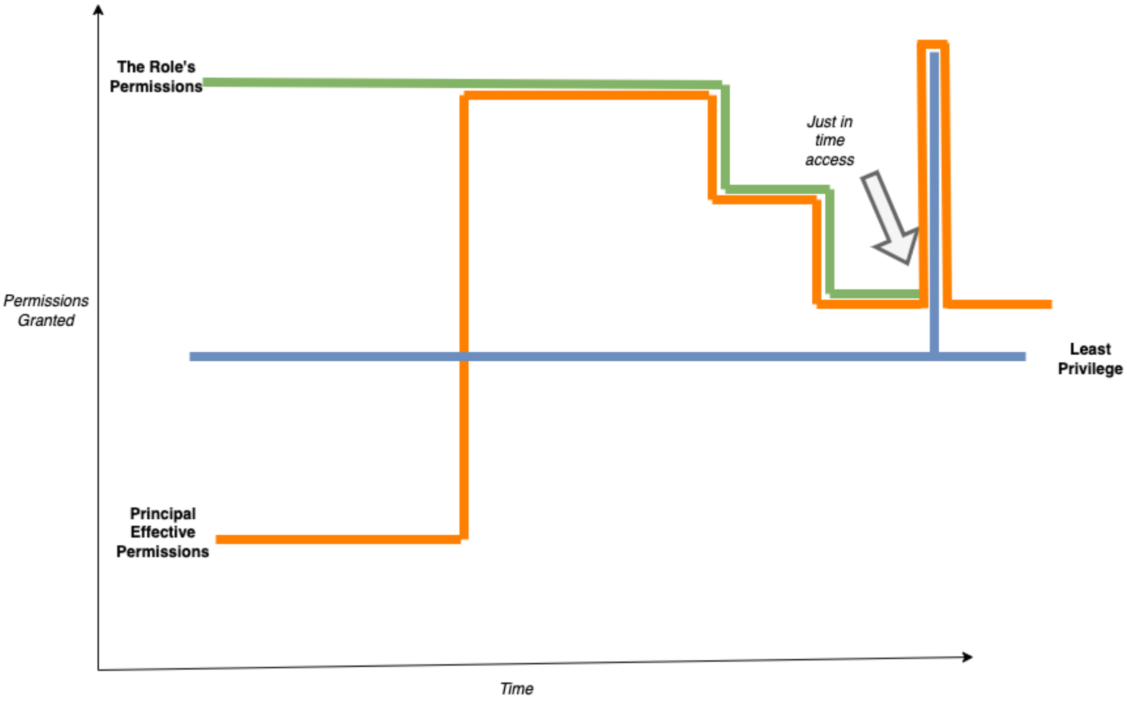


Figure 2: Just-in-time Least Privilege Model

This JIT approach is analogous to *Privileged Access Management* (PAM) systems, which typically allows one to "check out" and "check in" a credential used to access a shared (and often sensitive) system. Instead of checking out a credential, though, the JIT approach enables the actor to check out a permission to be granted to perform the action. Alternatively, the JIT approach may allow the principal to check out the ability to assume a role with the necessary permissions to perform the activity. The permissions granted for that JIT access should also be continually refined.

The risks with long-standing permissions or role assignments are related to unauthorized privilege escalation. If a credential of the principal is compromised or another principal is able to assume the role with the permission, a privilege escalation breach occurs. A JIT approach can mitigate the risk of the privilege escalation, as the principal of the compromised credential would not have a long-standing permission. The principal requires the additional step of checking out the role or permission. This mitigation assumes that the same compromised credential cannot be used for the "checking out" of the role or permission needed to escalate privilege. Thus, best practice dictates that the JIT system requires an additional authentication factor. For example, if typical operations utilize a fingerprint biometric, the privilege escalation might require a hardware device token.

There is a balance between security, productivity, and convenience to consider when implementing least privilege. If the cost of building and maintaining refinement and JIT exceeds the impact of privilege escalation in your systems, you may choose to accept the risk of long-standing or unneeded permissions granted to principals. There is a productivity risk of being too surgical with permissions and interrupting work. Employees that have to constantly check out permissions to do their job may grow weary of the tax and find ways to circumvent the control.

Least Privilege Relation to Policy-Based Access Control

Typically, Policy-Based Access Control (PBAC)[1] lends itself well to least privilege as its rules tend to be more granular than RBAC with the specification of specific resources and actions in. For example, the following natural language statement is representative of a PBAC rule:

Allow read content if the reader's clearance is higher than the content's classification

This statement grants the read-content permission based on a conditional comparison of an identity attribute, the reader's clearance, to a resource attribute, which is the classification of the content. This rule could be updated to approach least privilege, perhaps by specifying a smaller population of readers or specifying which instance of the content server. However, this negates some of the value of PBAC as you have to have rules for each enumerated instance of the content server. Least privilege becomes a balance with the centralized policy decision nature of PBAC and maintainability that comes from having rules that can apply to multiple abstractions.

PBAC lends itself to modeling least privilege in various dimensions. For example, to refine the content example toward least privilege, you might add a network expression that adds additional constraints on where readers can access content, or combine a risk engine score in a deny-override rule.

> *Allow read content if the reader's clearance is higher than the content's classification and client.ip in a specified range*
> *AND*
> *Deny if read-content risk is greater than low*

Refining a policy-based approach to access control may inherently require less refinement than an RBAC model over time. It does, however, require rigor toward auditing PBAC rules that may grant unnecessary access for a population or has a path that isn't reachable. Access governance is less mature in PBAC than in RBAC, so there may be less choice from commercial offerings in this area.

[1] More on Policy-Based Access Controls is available Mary K McKee, "Introduction to Policy-Based Access Controls (v2)" IDPro Body of Knowledge 1(8). doi:https://bok.idpro.org/article/id/61/

Summary

Least privilege is an ever-shifting target that can act as a “north star” for your access governance teams to strive for in order to reduce the risk of unauthorized privilege escalation. Continuously refining permissions assigned during birthright, self-service access requests, and specific activities can limit the accumulation of privileged access that can be misused over time. Incorporating JIT strategies to grant permissions for short durations to achieve a temporary task reduces long-standing permissions. Organizations should consider the productivity risk from the over-refinement of permissions or the overhead of having to ask for permissions too frequently before investing in tools or processes. Monitor your provider’s permission model to ensure that newly introduced permissions do not introduce risk from your policies that use wildcards. As you commit to a role-based or policy-based access control model, your techniques for least privilege will vary, but the concepts will be consistent. Continuously evaluating these factors over the lifecycle of all identities and policies will reduce the surface area that can be exploited.

Author Bio

Matt Carter has worked in the identity and cloud security industry since joining Netegrity in 2000. Mr. Carter has worked in several roles, including product management, pre-sales, and implementation at companies like Oracle, AWS, and Axiomatics. Currently, he is a CIAM sales specialist at Okta. Matt has been active with IDPro as a certification test item writer and serves on the Book of Knowledge committee.

Matt Carter lives in greater Boston with his wife, two dogs, two cats, and has three kids in college. He is into pickleball, kayaking on the Charles river, and science fiction. Fun fact: Matt once ran with the bulls in Pamplona...in flip-flops.

[i] NIST Information Technology Laboratory, “least priviledge,” Computer Security Resource Center glossary, https://csrc.nist.gov/glossary/term/least_privilege (accessed 6 September 2022).

[ii] Flanagan (Editor), H., (2021) “Terminology in the IDPro Body of Knowledge”, *IDPro Body of Knowledge* 1(8). doi: https://doi.org/10.55621/idpro.41.

[iii] Wikipedia contributors, "Infrastructure as code," *Wikipedia, The Free Encyclopedia,* https://en.wikipedia.org/w/index.php?title=Infrastructure_as_code&oldid=1100109083 (accessed September 6, 2022).

[iv] More on Joiner, Mover, and Leaver is available in Cameron, A. & Grewe, O. (2022) “An Overview of the Digital Identity Lifecycle (v2)”, IDPro Body of Knowledge 1(7). doi: https://doi.org/10.55621/idpro.31.

[v] "Admin" - shorthand term for a privileged user or role that has full control over a digital environment. The scope of "Admin" may vary, but represents a set of permissions that would allow a person controlling it to manipulate or damage assets and should be tightly controlled.

[vi] An organization's constraints for provisioning a resource like compute can be very specific in terms of policy. For instance, an organization may only want to allow a database to be created in a particular region, of a certain size, and with specific features enabled.

Digital Identity

Account Recovery (v3)

Dean H. Saxe, Sr. Security Engineer, Amazon Web Services

To comment on this article, please visit our GitHub repository and submit an issue.

Table of Contents

Abstract

All systems that require authentication of users share a common problem: users are human. Users forget or lose their credentials, lose, reimage, break, or sell hardware with embedded credentials (e.g., a phone or laptop). Account access is lost when users lose access to an email address their account is bound to. In some systems, credentials expire and need to be reissued. The common theme is that users need alternative mechanisms to restore access to the accounts whose credentials are unavailable.

The following article establishes a framework for evaluating Account Recovery mechanisms and establishes recommendations for Account Recovery in consumer, education, enterprise, and government spaces by identifying the benefits and risks of common mechanisms. Given the variety of concerns – privacy, security, and access continuity - in different domains, the reader of this document is expected to apply the guidance herein alongside their domain expertise and judgment to design, develop, and deploy Account Recovery mechanisms for their online systems. Due to the intersection between Account Recovery actions and Customer Service teams, the author strongly recommends that the reader also consult the article "Managing Identity in Customer Service Operations" in the IDPro Body of Knowledge.

Terminology/Glossary

- **Account Owner –** An entity that "owns" or claims responsibility for an account. Generally, an account is issued in the name of the owner(s) or their delegate(s) in the case of enterprises.
- **Account Recovery (AR) -** The process of returning account access to an account owner when they lose, forget, or cannot otherwise produce the account's nominal credentials. This may be accomplished in person, remote, or in a hybrid format.
- **Account Takeover -** Account takeover is a form of identity theft and fraud, where a malicious third party successfully gains access to a user's account credentials.[i]
- **Agent (also "Customer Service Agent")** - The person responsible for communicating with and solving problems on behalf of your customer or end-user.
- **Credentials -** Any attribute or shared secret that can be used to authenticate a user.
- **Knowledge-Based Authentication (KBA) -** A method of authentication that uses information known by both the end-user and the authentication service but is not necessarily a secret.
- **Multi-Factor Authentication (MFA) -** An approach whereby a user's identity is validated to the trust level required according to a security policy for a resource being accessed using more than one factor (something you know (e.g., password), something you have (e.g., smartphone), something you are (e.g., fingerprint).[ii]

- **Personal Data -** Personal data are any information which are related to an identified or identifiable natural person.[iii]
- **Social engineering -** Social engineering is a method of manipulating people so they give up confidential information, such as passwords or bank information, or grant access to their computer to secretly install malicious software.[iv]
- **Threat Modeling -** Threat modeling is an analysis technique used to help identify threats, attacks, vulnerabilities, and countermeasures that could impact an application or process.[v]
- **Username -** An identifier unique to the authentication service used in conjunction with a credential such as a password or FIDO authenticator to authenticate a user.

Account Recovery

Defining AR

What is AR? You'll see one definition above, but a fuller description follows. AR is a mechanism or collection of mechanisms that are used to maintain continuity of access to a user's services. AR operates by providing an *alternative authentication mechanism* to *reestablish authentication credentials*, such as through re-identification of the user. A key property of any AR mechanism is that it must meet or exceed the security of the nominal authentication mechanism for the account that it serves to recover. If this property is not met, users may choose to execute the AR mechanism rather than remember their credentials. This also opens the door to AR being used as an account takeover mechanism.

A real example of the abuse of AR mechanisms happened to the author. Our family had shares of an American company; the shares were managed through an online portal. Each year I had to log in to collect the tax forms, but I could never remember the password. The service's AR process required two pieces of readily available information: my mother-in-law's maiden name and my wife's date of birth. Each year I would log in with these pieces of known information, collect the documents I needed, and logout. The password was not required, nor did the AR process require a password reset or notify the account holder of the access!

An Iron Triangle of Account Recovery

As an owner of a resource, I have to decide the balance of three concerns - Privacy, Access Continuity, and Security - that meet my needs within the constraints of the service I'm accessing. In an iron triangle, I can move away from any vertex toward another to obtain relatively more of one concern (e.g., privacy) at the cost of another (e.g., security or access continuity).[vi]

In the stock example above, the system design focused on high access continuity exclusively to the detriment of security - the account is easy to access by malicious actors who could execute transactions - and privacy - the account owner is fully identified by the stock service, as is the nature for most financial systems.

In contrast, my current bank focuses on access continuity and security - it is hard to gain access to my account online due to strong authentication requirements, and (relatively) easy for me to regain access to my account by visiting a branch in person with government identification. The bank is obligated to identify me based on my government-provided identity documents (e.g., passport, driver's license) for conducting certain transactions and uses this same in-person authentication of my government-issued credentials to restore access to my account if required. *This is an act of authentication!* The driver's license looks normal, unaltered, anti-fraud elements are in place, the expiration date is valid, the image looks like the person standing in the bank, the document is machine-readable and matches the person, etc.; thus, I can conduct a transaction. (Note that this is not a fraud-free mechanism of authentication. However, the risk of a scalable attack in the physical world is significantly less than a purely online service.)

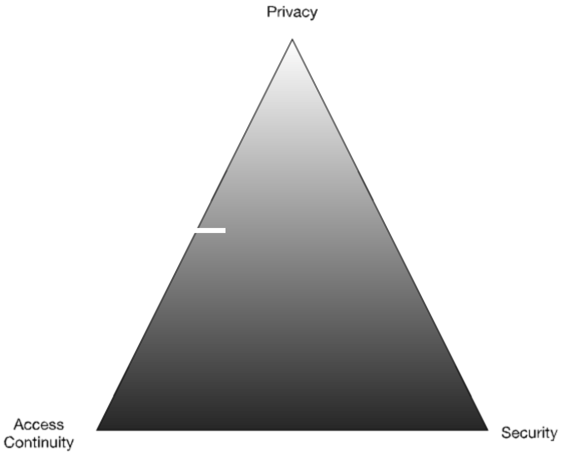
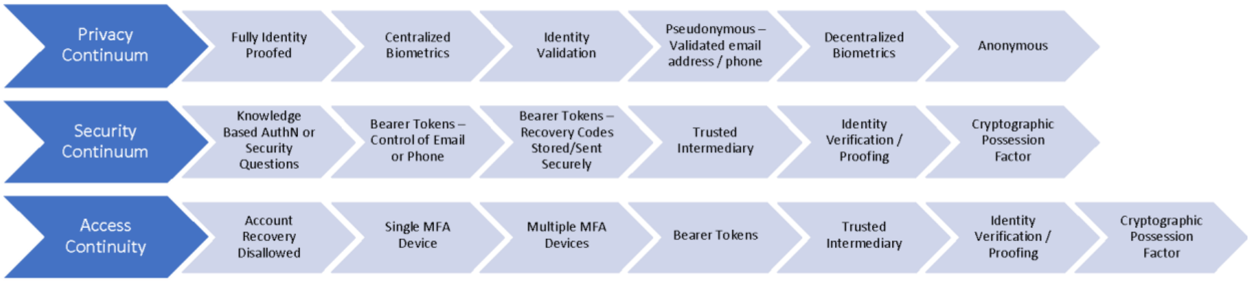


Figure 1: The Iron Triangle of Cybersecurity

Finally, Reddit, a social news aggregation site, balances all three concerns. My email was validated on signup by forcing me to close the loop by clicking on a one-time use URL. Reddit allows me to use multiple MFA devices, and I can recover my account through a backup code. But if the backup codes are lost, the password unknown, and MFA devices are not available, I'll lose access to my account without recourse.

Which one is correct? Potentially all of them, depending on the threat model.

Given these constraints, how can we apply this iron triangle to designing registration, authentication, and account recovery systems? Below are three continuums representing each vertex; movement toward the arrow is correlated with a higher score on the continuum toward the vertex in the triangle (values are relative, not absolute).



Privacy Continuum	Fully Identity Proofed	Centralized Biometrics	Identity Validation	Pseudonymous – Validated email address / phone	Decentralized Biometrics	Anonymous	
Security Continuum	Knowledge Based AuthN or Security Questions	Bearer Tokens – Control of Email or Phone	Bearer Tokens – Recovery Codes Stored/Sent Securely	Trusted Intermediary	Identity Verification / Proofing	Cryptographic Possession Factor	
Access Continuity	Account Recovery Disallowed	Single MFA Device	Multiple MFA Devices	Bearer Tokens	Trusted Intermediary	Identity Verification / Proofing	Cryptographic Possession Factor

Figure 2: The three continuums of an Iron Triangle of Access Continuity. Moving from left to right on each continuum leads closer to the appropriate vertex of the triangle.

In a nutshell, Identity architects can use the iron triangle to first identify where in the triangle the use case is situated and second to identify the trade-offs that are made to meet the needs of the use case.[vii] However, the devil is in the details, and those details will differ wildly across different identity ecosystems.

Consumer AR

Consumer use cases are focused on end-users of commercial systems open to the general public. Depending on the nature of the consumer relationship, there may or may not be any in-person interactions, which can limit the mechanisms used for reestablishing credentials for the user.

The risk associated with consumer accounts varies widely depending on the service. While both banking and social networking need to operate AR mechanisms for their users, the risk of compromise of each account type is significantly different. There is also a different set of information available to these different consumer services to enable AR.

Enterprise AR

In the enterprise, the focus is usually on access continuity – minimizing user downtime - and security for AR processes. AR is generally straightforward for on-premises employees: Present yourself and your enterprise or government identification to the IT Help Desk and reset your credentials. This is a form of identity proofing for AR. However, as more corporate employees work remotely, this in-person mechanism may no longer work. In these cases, enterprises must look at remote mechanisms, which could include remote identity proofing, using a trusted intermediary (e.g., supervisor) to vouch for the employee, and intermediate the process of AR, using a quorum of trusted intermediaries to vouch for the employee, etc.

Education AR

Similar to enterprises, the focus for education is on access continuity. On-campus staff and students can use in-person services for account recovery. Remote students and staff may use similar mechanisms to enterprises, adapted to their unique environment.

Government AR

Due to the wide variations in government systems and services, there is little consistency in this realm. Implementers should be observant of local, national, and supranational laws, regulations, and cultural norms when working with account recovery in this space.

AR Mechanisms

Below we review common AR mechanisms. However, we would be remiss to not include as the first and primary mechanism *Make Losing Access Difficult.* In other words, if we do not first start with a focus on maintaining access continuity for our users in the happy path, we will see more requests for AR**. Identity architects must consider the AR use cases as a primary concern when designing authentication systems and not treat AR as a second-class use case.**

Make Losing Access Difficult

How do services make access continuity easy and losing access difficult? At the most basic level, services should nudge their users into making good decisions. This can include:

- Baselining contact information – does the user have access to their email, phone, or other contact channels? If not, is there a backup mechanism to reach the user? Did your identity system close the loop, ensuring access to the primary contact information to complete account registration?

- Baselining authentication mechanisms – Your users may have one or many devices used to authenticate to different services. Can the user access their authentication mechanism(s) such as FIDO authenticators, OTPs, and a phone number for SMS? Do the devices still work? Are the device and/or mechanisms still supported?

- Back-Up Authentication – How will your users authenticate if the primary authenticator is unavailable? The canonical example is a user who is flying – they have internet access but may not have SMS messaging. How will these users authenticate if the service requires an SMS OTP? Best practices should include encouraging multiple authentication options per user, such as multiple OTPs, FIDO authenticators, and backup codes. The loss of one does not trigger an AR event or limit the availability of the service. Limiting users to a single MFA mechanism *ensures* that that user will need to execute AR if the device is lost, broken, or temporarily unavailable. *This is a user experience that should be avoided!*

- Remind users to set up one or more AR mechanisms early in the account lifecycle, and nudge users to baseline those mechanisms regularly. *Users without an AR mechanism may not be able to recover accounts.* If the user has not configured AR, use significant changes (e.g., exceptional growth in usage of a cloud service), security checkups, or other dashboards to drive user actions.

- Use synced passkeys. Synced passkeys enable the process of credential recovery in addition to the existing account recovery mechanisms. Credential recovery for synced passkeys, e.g., those synchronized to a platform such as Apple or Google, or a third-party passkey provider, such as 1Password or Dashlane, is facilitated by the user's passkey provider. Functionally, credential recovery operates by enabling the user to bootstrap a new device into the provider's ecosystem after losing all prior access. Mechanisms are non-standard and likely to vary between providers, therefore, the security of these mechanisms must be assessed on a provider-by-provider basis.

Identity providers should also guide their users to avoid single points of failure on the user side. For example, if the user places their credentials in a password safe and recovery codes are stored in the same safe, loss of access to the password safe eliminates at least one recovery pathway. Although we cannot always prevent users from shooting themselves in the foot, we can try to limit the damage that the user can do to themselves.

User Notifications

Before diving into the mechanisms of AR, we must pause to talk about user notifications as an important component of the AR process user experience. All actions that impact the user’s ability to maintain access continuity must be reported to the user. These include, but are not limited to:

- Changes to the account email address
- Changes to the account phone number(s)
- Changes to the account credentials including, but not limited to
  - Passwords
  - MFA devices / mechanisms
  - Reset or re-issuance of recovery codes
- Removal or addition of trusted intermediaries
- Account recovery (success or failure)

Due to the time-sensitive nature of these messages, they should be broadcast to all available channels which the use has consented to, such as email, SMS, and push notifications. Notification should be sent to the prior email address and/or phone number during a change request, allowing the user an opportunity to identify a fraudulent change and revert the change before further damage occurs.

Bearer Tokens

Bearer tokens, when used for AR, can be thought about as paper tickets to a concert or sports event. The tickets (or bearer tokens) are used once to access a service in lieu of the user's normal credentials.

These bearer tokens take a few forms:

- Alphanumeric codes sent via email or SMS in response to an AR request
- Magic links, a form of passwordless login, sent via email or SMS in response to an AR request
- Recovery codes obtained prior to losing access and stored as physical or digital copies in a safe place, such as a fireproof safe.
- Recovery code sent to the user via postal mail or private delivery service

Grouping these mechanisms as bearer tokens allows us to reason about their usability and security together. The assurance level of a bearer token is directly correlated to how it was delivered. Recovery codes obtained in an authenticated session are generally higher assurance than one-time codes or magic links; however, this is dependent upon how they are stored by the user.

Benefits

- An easy user experience that requires no specialized knowledge or hardware. After triggering an AR event, such as by entering a username into an AR workflow, the user cashes in the bearer token for the ability to reestablish credentials with the service.

Threats and Mitigations

- Bearer tokens may be used by whoever bears them – this makes them easy to use and abuse, such as through phishing.
    - Minimize the validity window of all bearer tokens.
    - Keep state – is the user on the same device and same browser as when the request was triggered? Has the IP changed? What other data can be collected to ensure the user has not been phished for this information.
- The risk of bearer tokens also encompasses the risk of the medium by which they are sent to the user. These threats cannot be mitigated by the identity provider.
    - Email is subject to interception, such as by phishing, leading malicious actors to access the bearer tokens sent to the email address.
    - SMS is subject to interception, such as through SIM swapping attacks and SS7 vulnerabilities.

  - Email and SMS mechanisms are subject to threats against the providers and their infrastructure, as well.
- Users fail to copy recovery codes, fail to store the recovery codes securely, or lose the recovery codes.
  - Providers can recommend mechanisms for storage and management of codes, but the user may not follow the guidance.
- Users lose access to their email or phone number or enter incorrect values which the user cannot access.
  - Verify the user has access to the email or phone number when they are submitted to the IdP.
  - Baseline the continued access to the email and phone number over time.

Knowledge-Based Authentication / Security Questions

Both Knowledge-Based Authentication (KBA) and Security Questions are used as recovery mechanisms by having the user "prove" they are the legitimate owner by answering questions known only to the user. Unfortunately, both KBA, based on public information databases or recent user transactions, previous passwords, and security questions, based on preconfigured questions and answers provided by the user, are relatively weak recovery mechanisms.

KBA mechanisms often utilize information such as home addresses, loan dates/amounts, and credit report data to weakly identify the human owner of an account. However, due to numerous data breaches, this information is insufficiently secret and should not be depended upon as a recovery mechanism for accounts with any significant value.

Information used for KBA may often be available to family members or other parties close to the user, reducing their efficacy.

Similarly, security questions often have predictable or easily identifiable answers. Questions such as favorite color have low entropy (according to this study, 64% of Americans choose one of four favorite colors, blue (29%), green (21%), purple (8%), and red (8%)), while questions about a favorite sports team or high school mascot may be readily discoverable through social media.

As a low assurance mechanism, KBA and security questions are only recommended for the lowest-risk operations as a last resort.

Benefits

- KBA and secret questions are easy to use, when they work.

Threats and Mitigations

- KBA data may be obtained from breach corpuses, public databases.

    - Don’t use KBA for account recovery.
- Insufficient protection in cases of domestic violence or intimate partner violence where the KBA data may be known.
    - Don’t use KBA for account recovery.
- Customers may not remember details to answer KBA questions. A customer’s inability to remember details such as financial transactions will trigger false negative matches for legitimate customers. Conversely, a user who answers all questions correctly may be a fraudster.
    - Don’t use KBA for account recovery.
- Security questions and answers may be forgotten. Users may fail to recall the answers, misspell answers, misuse capitalization or punctuation, all of which could cause the user to fail authentication.
    - Baselining of security questions and answers to ensure access continuity.
- Security questions and answers are alternative passwords and suffer the same risks as any password authentication scheme.
    - Users must never be asked to share KBA data or security questions and answers with CS agents to eliminate this risk.
    - Follow password storage guidance for all security questions and answers.

Identity Verification / Identity Proofing

In some use cases where privacy of the individual’s identity is not the overriding concern, systems may use identity verification or identity proofing to establish the real-world identity of a human, often based upon government (driver’s license, passport), enterprise (employee badge), or educational credentials (university or school ID) issued by a trusted authority. Early in the account lifecycle, perhaps as a requirement to establish the account, the user’s identity is verified, binding the identity to the user account. This may take place in person (e.g., at a bank, registering for a trusted traveler program, at a university during registration, at an employer on the employee’s first day), or remotely. Since these require in-person interactions, they cannot easily be automated and provide a higher barrier to entry for fraudulent access. In the remote use case, a common modality is to ask the user to take an image of their identity document and a selfie or short selfie-video. The identity documents are reviewed for signs of tampering or other fraud markers. The image on the identity document is compared with the selfie or video, which is usually tested for liveness by asking the user to do certain behaviors such as look up, down, left, right, before confirming that the human at the keyboard is the same human on the identity document (to some level of certainty).

Benefits

- Establishes a binding between the natural person and the user account that cannot be broken. Even if the user replaces their passport, identity verification can be re-executed to verify that the human is the “owner” of the account they are trying to recover (within certain confidence intervals).

- Resistance to scalable fraudulent mechanisms, though this depends upon the specific mechanisms used.
- May be highly automated with Artificial Intelligence/Machine Learning (AI/ML); however, many providers still use manual review of less common identity documents first before using them to train AI/ML systems.

Threats and Mitigations

- Users are uncomfortable sharing identity documents with online services. For example, the United States Internal Revenue Service (IRS) used ID.me to provided Identity Proofing services in 2022, resulting in a significant backlash.[viii]
    - Provide clear information on how the data provided will be used and stored.
    - Provide an alternative mechanism for users who are unwilling or unable to provide identity documents for remote ID Proofing. In-person, identity proofing, for example.
- Fraudulent documents
    - Today, there are no common criteria to assess identity document verification/proofing services against one another.
- Presentation Attacks – presenting a static image or video of the real person, rather than the person attempting fraudulent identity verification
    - Images and video selfies should use mechanisms of liveness detection to ensure the images are real and being captured in real-time.

Trusted Intermediary

Common in corporate settings, users are able to recover access through a trusted intermediary, such as the user's manager. The general use case is that when an employee loses access and needs to reset a password or configure a new MFA device, the helpdesk or the user's manager (or skip-level, etc., though this brings diminishing returns) can authenticate to a recovery service to help the user reestablish corporate credentials. Individual processes may vary depending on the familiarity of the user with the trusted intermediary. For example, a direct report to a manager may have the manager mediate recovery without presenting any identity information. The same user who approaches the helpdesk for a password reset will have to present a corporate badge or similar identity information before executing the reset. In a services industry, a sales manager or technical account manager may be the trusted intermediary for their customers if access is lost. The process may be completed in person, over the phone, or via video conference.

Facebook uses a trusted contacts model to create a self-service recovery mechanism.

Multiple intermediaries can be used, as well, in a quorum (*m* of *n*) based solution. Quorums are useful for higher assurance use cases to eliminate the threat of social engineering or a single malicious user using the AR process to gain access to unauthorized accounts.

Benefits

- Distributes the work of AR amongst many possible trusted users, allowing for a high level of access continuity.

Threats and Mitigations

- Malicious “trusted” intermediary takes over a targeted account.
    - Require quorums
    - Don’t pass recovery tokens, URLs, etc., through the trusted intermediary. Allow the intermediary to trigger sending the token to the subject of the AR action via email, SMS, or other mechanisms. (Be careful, this could look like phishing!)

Possession Factor

Similar to the bearer token discussed above, a possession factor – such as the ability to sign a transaction with a specific private key – can be used as a recovery factor. However, the average user should not be expected to generate and manage their own keys securely. The addition of FIDO2 security keys and passkeys creates a secure mechanism for creating and managing account-specific key pairs. When used as a first-factor device (e.g., the passwordless flow), a security key or passkey can be registered as a “recovery key” for the account.[ix] Only the owner in possession of the key and with the biometric or PIN to unlock it can recover the account. Applications on a mobile device can be used as a possession factor when unlocked with the user’s biometric or PIN code. This can be done using common protocols, such as FIDO passkeys, or using a bespoke mechanism.

Last, self-sovereign identity (SSI) can use a similar mechanism. By proving ownership of a specific private key associated with the user’s DID document, the owner can conceivably recover an account.

Benefits

- Ease of AR if the possession factor is registered early in the lifecycle and can be made available when needed by the user.

Threats & Mitigations

- Loss of the cryptographic key or its storage medium.
    - Implementers must consider the relative frequency of loss of a phone, for example, vs. a hardware key vs. a public key generated on the user’s disk. This may be mitigated through passkeys synced via a cloud service.
    - Allow for multiple possession factors per account.
    - Periodically remind users to check their ability to recover with the possession factor(s)

Customer Service

The final mechanism for AR is through a customer service mechanism, such as customer service for an enterprise. Customer service may use one or more of the mechanisms identified above to process an AR request. For additional information on using CS for AR, see "Managing Identity in Customer Service Operations" by Arynn Crow and JP Rowan.[x]

No Account Recovery

In some scenarios, no account recovery may be the secure and private option. While not recommended for most use cases, not supporting any account recovery is seen in practice and may be the preferred option for some high-security services in order to minimize the risk of account takeover.

Conclusion

Account recovery is a mechanism to support authentication for your service. Building an AR service requires service owners to consider what they, and their customers, value: access continuity, security, or privacy, and build mechanisms to support AR that balance these three concerns. Which AR mechanisms are chosen will additionally depend on the support environment that the service is deploying into: education, enterprise, government, etc. Each has different abilities available to them that may enable stronger AR mechanisms. However, all AR mechanisms share one thing in common: users must register for them implicitly or explicitly if they are to regain access to lost accounts. Therefore, AR is more than just a technical solution to be implemented; it is a user experience and human behavior problem to be solved.

Acknowledgments

- Arynn Crow
- JP Rowan
- David Brossard
- Paul Figura

Author Bio

Dean H. Saxe is a Senior Security Engineer with the AWS Identity team and a founding member of IDPro. He can be reached at dean@thesax.es or on Twitter @n3rd1ty.

Change Log

Date	Change
2023-10-27	V3 published; addition of info on passkey recovery
2022-06-03	V2 published; clarifications added to AR mechanisms
2021-04-19	V1 published

[i] Flanagan (Editor), H., (2021) "Terminology in the IDPro Body of Knowledge", *IDPro Body of Knowledge* 1(7). doi: https://doi.org/10.55621/idpro.41

[ii] Ibid.

[iii] Ibid.

[iv] Ibid.

[v] Ibid.

[vi] Caccamese, A. & Bragantini, D. (2012). "Beyond the iron triangle: year zero." Paper presented at PMI® Global Congress 2012—EMEA, Marsailles, France. Newtown Square, PA: Project Management Institute, https://www.pmi.org/learning/library/beyond-iron-triangle-year-zero-6381

[vii] Bucci, Steven. "The Iron Triangle of Cybersecurity." Security Debrief. February 23, 2011. http://securitydebrief.com/2011/02/23/the-iron-triangle-of-cybersecurity/.

[viii] Thimot, Tom, "The IRS/ID.me debacle: A teaching moment for tech," Venture Beat post, 15 April 2022, https://venturebeat.com/2022/04/15/the-irs-id-me-debacle-a-teaching-moment-for-tech/.

[ix] The astute reader will note that this is the same mechanism proposed by the FIDO Alliance for recovering from loss of a security key. At this time, there is no way to backup a security key, therefore registering multiple keys is the specified mechanism of account recovery.

[x] Crow, A. & Rowan, J. P., (2021) "Managing Identity in Customer Service Operations", IDPro Body of Knowledge 1(4). doi: https://doi.org/10.55621/idpro.65.

Defining the Problem – Identity Proofing Challenges

Russ Reopell, Sandy Christopher, and Lorrayne Auld

To comment on this article, please visit our GitHub repository and submit an issue .

Table of Contents

Abstract

Identity proofing, process by which a credential service provider collects, validates, and verifies information about a person, is a critical step for many identity systems. This article explores identity proofing in general and why current practices are challenging. While the article is largely informed by the identity proofing examples within the United States, the concepts are globally applicable.

Introduction

Whether you're purchasing merchandise online or requesting financial or medical services from the federal government or health care providers, being able to prove you are who you claim to be and are indeed entitled to the goods and services you are attempting to access has become a crucial and required fact of everyday life. This article helps readers understand the difficulties and challenges they may face in registering for online goods and services.

Terminology

Applicant: A subject undergoing the processes of enrollment and identity proofing.

Binding: Associating an authenticator with an identity.

Claimant: A subject whose identity is to be verified by using one or more authentication protocols.

Claimed Identity: An applicant's declaration of unvalidated and unverified personal attributes.

Credential: An object or data structure that authoritatively binds an identity—via an identifier or identifiers—and (optionally) additional attributes to at least one authenticator possessed and controlled by a subscriber.

Credential Service Provider (CSP): A trusted entity that issues or registers subscriber authenticators and issues electronic credentials to subscribers. A CSP may be an independent third party or may issue credentials for its own use.

Enrollment: Also known as Registration. Enrollment is concerned with the proofing and lifecycle aspects of the principal (or subject). The entity that performs enrollment has sometimes been known as a Registration Authority, but we (following NIST SP.800-63-3) will use the term Credential Service Provider.

Identity: An attribute or set of attributes that uniquely describes a subject within a given context.

Identity Evidence: Information or documentation the applicant provides to support the claimed identity. Identity evidence may be physical (e.g., a driver's license) or digital (e.g., an assertion generated and issued by a CSP based on the applicant successfully authenticating to the CSP).

Identity Proofing: The process by which a CSP collects, validates, and verifies information about a person.

Identity Provider (IdP): The party that manages the subscriber's primary authentication credentials and issues assertions derived from those credentials. This is commonly the CSP as discussed within this article.

Knowledge-Based Authentication (KBA): Identity-verification method based on knowledge of private information associated with the claimed identity. This is often referred to as knowledge-based verification (KBV) or knowledge-based proofing (KBP).

Registration: See Enrollment.

Remote*: In the context of remote authentication or remote transaction*, an information exchange between network-connected devices where the information cannot be reliably protected end to end by a single organization's security controls.

Subscriber: A party enrolled in the CSP identity service.

Why do we need identity proofing?

Today, many companies and government agencies rely heavily on accurately identifying, credentialing, monitoring, and managing user access to information and information systems across their enterprise to ensure they know who is accessing their data. One of the challenges of digital identity is associating a set of online activities with a specific entity. There are numerous situations where it is important to reliably establish an association of a digital identity with a real-life subject. Examples include obtaining health care and executing financial transactions. There are also situations where the association is required for regulatory reasons (e.g., the financial industry's Know Your Customer (KYC) requirements, established in the implementation of the USA PATRIOT Act of 2001) [i] or to establish accountability for high-risk actions (e.g., changing the release rate of water from a dam).

Identity proofing establishes that a person is who they say they are based on the validity of one or more pieces of identity evidence. The more due diligence incorporated into the identity-proofing process, the higher the confidence that the applicant is who they claim to be. For example, one would place little confidence in self-asserted identity ("I say I am Santa Claus, therefore I am Santa Claus"). However, suppose I claim to be Mother Nature and can provide written and corroborated identity evidence proving I am Mother Nature. In that case, there is a much higher level of confidence placed in that identity. If I provide all that documentation to the CSP in person, you can be sure I am who I claim to be.

What is identity proofing?

Identity proofing is the process used by a *credential service provider (CSP)* to collect, validate, and verify the identity evidence provided by an applicant to establish a subscriber's digital identity. The *identity provider (IdP)* manages the subscriber's primary authenticators and, in federation agreements, issues assertions derived from the subscriber's account. When an applicant is identity proofed, the expected outcomes are:

- The *claimed identity* (a set of unvalidated and unverified personal attributes) is resolved to a single, unique identity within the context of the population of users the IdP/CSP serves and has been validated to exist in the real world.
- All supplied identity evidence is validated to be correct and genuine (e.g., not counterfeit or misappropriated).
- The CSP/IdP verifies that the claimed identity is associated with the real person who supplied the identity evidence.

When conducting an online transaction, a digital identity represents the person trying to access the digital service.

How is a Digital Identity created?

A digital *identity* is created based on a positive verification of an applicant from the identity proofing process. Identity proofing starts during the initial enrollment/registration process and may be updated at various stages of the digital identity lifecycle where life events warrant it. Figure 1 shows the Digital Identity Lifecycle and the events that take place during the creation, ongoing maintenance, and the suspension or expiration of a digital identity.[ii] Identity proofing can be performed remotely via the Internet or in person at a physical building with individuals hired and trained to perform proper proofing.

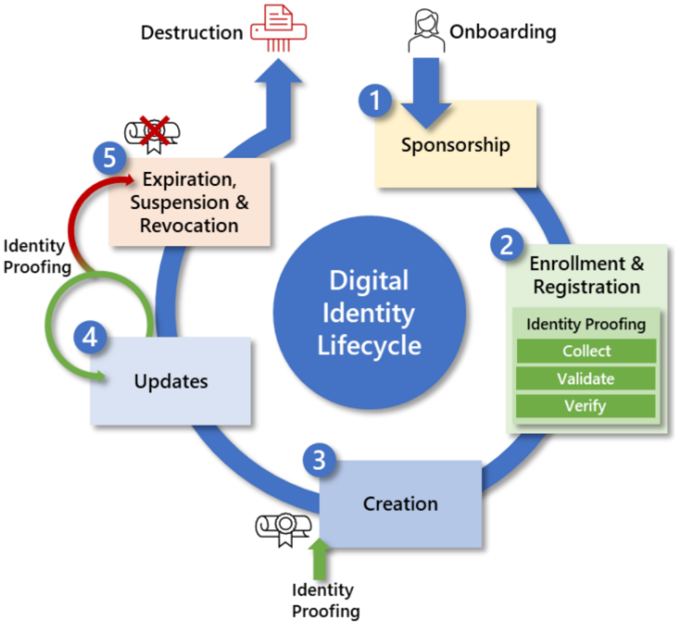


Figure 1 - Identity Proofing in the Digital Identity Life Cycle

Identity proofing is thought to be done once, at the time of enrollment/registration. But that may not be the only case and may be required at various stages of the digital identity lifecycle where life events warrant it. As illustrated in Figure 1, the following are the digital identity lifecycle processes:

1. Sponsorship: The onboarding process to obtain a digital identity. This process may require the applicant to either have or create an account with the CSP prior to sponsorship. This is the first step in the digital identity lifecycle.
2. Enrollment and Registration: The process through which an applicant applies to become a subscriber of the CSP and the CSP validates the applicant's identity. This is generally done via an in-person or remote identity-proofing process.
3. Creation: After a successful Identity Proofing event, the CSP provisions a credential by binding the credential to the subscriber's digital identity.
4. Updates: The act or process by which a requirement to be identity proofed after the initial digital identity is established. Examples of identity-proofing updates include:
    a. Per policy, an organization may require identity proofing of their users every three years, such as a government employee who needs to renew the certificates on their smart card.
    b. Change in name or gender may require the subscriber to be identity proofed again.
    c. The subscriber may initially have been identity proofed at a lower assurance level but, based on required access to higher-risk transactions, the subscriber may be asked to be identity proofed at a higher level of assurance.
    d. There are several scenarios, including times of emergency or transactions between strangers, when one may need to be identity proofed to ensure that that digital identity still belongs to that real-life person who was identity proofed at enrollment.
5. Suspension/Revocation: Revocation is the process of permanently changing the status of a credential to invalid (e.g., the credential has been compromised or the status of the sponsor has changed). There may also be an expiration of the credential bound to the subscriber, which may either trigger another identity-proofing event to renew the credential or surrender the credential housed on a smart card to the CSP. Reasons for suspending or revoking a credential include:
    a. Lost/stolen device.
    b. Death of the subscriber.

What is the difference Between In-Person Proofing and Remote Proofing?

In-person identity proofing is when individuals are required to present themselves and their documentation directly to a person. Remote identity proofing is used when individuals are not expected to present themselves or their documents in person and, instead, provide it online. In either case, this traditionally involves validating and verifying presented data against one or more corroborating authoritative sources of data.

Why is remote identity proofing hard and what are the challenges?

Historically, IdPs/CSPs who offered remote identity proofing services typically relied on *knowledge-based authentication (KBA),* where applicants were asked static questions about themselves and expected to be the only ones to know the answers to such questions, such as job history, credit report data or credit history, their mother's maiden name, their date of birth, etc. IDPs/CSPs used data collection companies, such as the credit bureaus, Lexis/Nexis, SEON Technologies, Silent Eight, and others, as authoritative sources of identity information to verify the applicant's responses. If applicants responded correctly to these questions, the credit bureaus would provide a scoring to indicate the assurance of that identity based upon the answers provided. The CSPs, in turn, used those scores in determining the acceptable level of assurance that the identity was verified. However, due to recent data breaches, massive amounts of personally identifiable information (PII) have been stolen and made available from multiple sources, including those on the dark web. Reports of fraud activity clearly show that significant amounts of PII have fallen into the hands of criminals and are being used for identity-related crimes, such as stealing services, assets, or benefits. The recent Twitter, LastPass, and AT&T data breaches, as reported by the Identity Theft Resource Center, are good examples of these types of compromised identity data.[iii] As a result, solely relying on the use of KBA is insufficient for corroborating an individual's claimed identity.

Successful remote identity proofing is contingent on the user having technical knowledge of the process and what is needed to accomplish it successfully (e.g., the user has a smartphone and the ability to use it to capture images/pictures and has valid identification that can be verified with the issuing authority). Online remote identity proofing is difficult because the validation and verification process can be cumbersome and challenging. Identity documentation may not be available, or the documentation provided by the applicant may be insufficient. Further difficulties arise when not all applicants have a smartphone or government-issued identification card that can be remotely validated. Some may find the identity validation and verification process can be too time-consuming or difficult. This increased user friction causes applicants to get frustrated and abandon the service.

The U.S. Government Accountability Office (GAO) released a remote identity proofing report that identified four out of six federal agencies that are still relying on PII-related KBA.[iv] The GAO report cites high costs and implementation challenges for certain segments of the public as reasons why some agencies have not adopted alternative identity-proofing methods to KBA. For example, the lack of a mobile phone for some applicant populations was given as a key implementation challenge. Organizations still using KBA should evaluate the value of their KBA solutions and, where possible, replace them with a more dynamic KBA. Additionally, the European Union Agency for Cybersecurity, ENISA, which is dedicated to achieving a common level-high of cybersecurity across Europe, also published a remote

I.D. proofing report in March 2021.[v] In their report, they've identified similar gaps with a lack of awareness and understanding of the remote proofing process, the variation in quality and completeness of identity evidence across the many European countries, and the desire to use physical presence as the benchmark, which, while tempting, cannot be reasonable when considering the variables introduced in remote proofing.

Over the last few years, there have been multiple government efforts to offer the public secure and private online access to participating government programs both here in the U.S. and abroad. The goal was to make managing government-provided benefits, services, and applications easier and more secure for the populations they were designed to serve. Whether agency applications and services would need to integrate with a single government authentication service is still in question. A single authentication entity for government services would require users to first be redirected to this central authentication service via secure protocols to register, be identity proofed, and assigned an authenticator (either remotely or in-person). Once the user has been identity proofed and acquired an authenticator, the authenticator could be presented to any Government online application or service that accepts them, provided they meet the required identity assurance level of that application or service. Gaining consensus across multiple agencies of the one government to use a common authentication service has proven to be much more difficult than anticipated.

Another remote proofing challenge is that there are too many misperceptions about why personal information, especially biometrics, is being requested and used. Many citizens do not trust the government to protect their personal information and question how it is being used. As a result, many people are reluctant to share their personal information for fear that the information will be used for more than the specified purposes. By not carefully explaining why data is being collected, how it is being used, and whether or not the data is stored or destroyed after remote identity proofing is complete, individuals may not provide the required information and will therefore fail remote identity proofing.

According to concerns expressed by the GAO report, additional work is needed to ensure that a fraudulent image, such as a photo of a mask, is not being provided in lieu of a live image — a threat known as a "presentation attack." Keeping up with ever-evolving threats to remote identity proofing and implementing the proper security controls to mitigate those threats is an ongoing challenge.

Challenges with remote identity proofing extend to other countries as well. The United Kingdom (U.K.) was among the first to try remote identity proofing, but it has been plagued with performance issues. One of their key problems was centered around the datasets used by the identity providers when trying to confirm a user's identity. Applicant data used for verification did not match what was on the government's systems, resulting in the U.K. government not being able to create and manage the system. Due to these problems, private industry is taking over the effort with the first task addressing the issue of the mismatched datasets used by the identity providers.

Summary

Today, many organizations and government agencies rely heavily on being able to accurately identify, credential, monitor, and manage user access to information and information systems across their enterprise to ensure they know who is accessing their data. There are numerous situations where it is important to reliably establish an association of a digital identity with a real-life subject. Identity proofing establishes that a person is who they say they are based on the validity of one or more pieces of identity evidence. The more due diligence incorporated into the identity-proofing process, the higher the confidence that the applicant is who they claim to be.

Historically, those who offered remote identity proofing services typically relied on knowledge-based authentication (KBA), where applicants were asked static questions about themselves (such as their mother's maiden name, the street they grew up on, or their father's date of birth) and expected to be the only one to know the answers to such questions. However, vast amounts of data about an individual have been stolen in data breaches and are readily available to purchase online. This stolen data can be used by fraudsters to then obtain access to your bank account, receive your stimulus check, or your tax returns. It is due to this high increase in stolen identities that organizations are finding that they no longer trust that digital identity and must improve their remote identity-proofing efforts to more effectively thwart fraudsters.

The use of online remote identity proofing services is difficult because the validation and verification process can be cumbersome and challenging. Identity documentation may not be available, or the documentation provided by the applicant may be insufficient. Further difficulties arise when not all applicants have a smartphone or government-issued identity card that can be remotely validated. Some may find the identity validation and verification process can be too time-consuming or difficult. This increased user friction causes applicants to get frustrated and abandon the service.

Authors

Lorrayne Auld
Principal Cybersecurity Engineer, MITRE Corporation
Lorrayne has over 25 years of experience in the area of identity and access management, secure web, portal, and Public Key Infrastructure (PKI) technologies supporting the Federal Government. She has worked both as a hands-on integrator and as a cybersecurity engineer providing guidance to the government. She has helped multiple agencies with their Identity, Credential, and Access Management (ICAM) strategies, implementation guidance, and best practices.

Lorrayne serves as the focal point for researching, understanding, and applying ICAM emerging technologies while ensuring ongoing growth within this area. She also serves as the senior advisor to the ICAM capability area as well as a mentor to junior staff. She has spoken at conferences on higher assurance identity proofing and next-generation authentication technologies. Lorrayne is a member of Kantara, IDPro, Women in Identity, and the FIDO Alliance.

Sandy Christopher
Senior Communications Advisor, MITRE Corporation
Building on 20+ years of leading communication and change, Sandy delivers holistic communication programs that measurably engage stakeholders and achieve business goals. Throughout her career, Sandy has worked with executive leadership to create strategic communication plans that align employees with the priorities of the organization. She is an innovative problem solver with extensive domestic and international communication experience on a wide range of issues, including organizational change, crisis communications, healthcare, information technology, ethics, operational risk, quality, deregulation of the utility industry, human resources, environmental, and financial services.

Russ Reopell
Principal Cybersecurity Engineer, MITRE Corporation
With over 25 years of experience in identity and access management, Public Key Infrastructure (PKI) technologies, and web services focused on identity, authentication, and authorization, Mr. Reopell has supported the Federal Government, Department of Defense, and Telecommunication companies. He began his career as a programmer and quickly became involved in the design, development, integration, and testing of various Air Force and Naval support systems. In the early 80s, he began working on information security systems and helped deploy security solutions in federal and commercial spaces until finally focusing on Identity, Credential, and Access Management (ICAM) strategies, implementation guidance, and best practices.

Russ worked closely with other MITRE staff and served as MITRE’s ICAM Capability Area Lead for many years. Russ was the go-to person across MITRE to assist with or guide staff in the design and integration of ICAM capabilities to the many sponsors MITRE supports. He is responsible for researching, understanding, and applying ICAM emerging technologies and helped to grow work in this ever-evolving area. Russ is a member of IDPro and enjoys mentoring junior staff to increase their knowledge as well as pique their curiosity about the many exciting innovations in the ICAM space.

[i] Dow Jones, "Understanding the Steps of a "Know Your Customer" Process," Risk and Compliance Glossary, n.d., https://www.dowjones.com/professional/risk/glossary/know-your-customer/ (accessed 27 March 2023).

[ii] For more on the digital identity lifecycle, see Cameron, A. & Grewe, O., (2022) "An Overview of the Digital Identity Lifecycle (v2)", *IDPro Body of Knowledge* 1(7). doi: https://doi.org/10.55621/idpro.31

[iii] Identity Theft Resource Center, *2022 Data Breach Report*, January 2023. https://www.idtheftcenter.org/wp-content/uploads/2023/01/ITRC_2022-Data-Breach-Report_Final-1.pdf (accessed 24 March 2023).

[iv] U.S. Government Accountability Office (U.S. GAO), *DATA PROTECTION Federal Agencies Need to Strengthen Online Identity Verification Processes*, May 2019. https://www.gao.gov/assets/gao-19-288-highlights.pdf (accessed24 March 2023).

[v] European Union Agency for Cybersecurity, *REMOTE ID PROOFING Analysis of methods to carry out identity proofing remotely,* March 2021. https://www.enisa.europa.eu/publications/enisa-report-remote-id-proofing (accessed 24 March 2023).

Identifiers and Usernames

By Ian Glazer

Table of Contents

Abstract

An identifier is the way an identity management system or other entity refers to a digital identity. The identifier used by the system, however, likely differs from the identifier used directly by the user and will definitely differ from identifiers in another domain. This article reviews the concept of identifiers as they relate primarily to people, both from a user's perspective and a system's perspective, and their impact on the systems that use them.

Introduction

What are identifiers and usernames?

In the physical world, we use a variety of ways to identify a person or a thing. From serial numbers to mailing addresses to license plates to nicknames, humans select a specific thing from a collection of similar things via an identifier. In the online world, this behavior is no different. Computer systems and people who work with them use identifiers to distinguish between similar items. More formally, and in the context of identity management, we can think of an identifier as the way an identity management system refers to a digital identity.

However, the person associated with that digital identity may not use the same identifier that the system uses. In fact, it is highly likely that they do not. Likely the person uses a human-friendly identifier. For the sake of differentiation, let's call the way a person in control of a digital identity identifies themselves to a system a username.

Why consider identifiers and usernames?

How systems refer to digital identities and how people refer to their digital identity in a system are crucially important. Identifiers and usernames are one of the most commonly used components of a digital identity management system. They have implications for usability, security, customer satisfaction, and system operations, and enable (or prevent) cross-system correlation and user account management. They have applicability in business to employee (B2E), business to business (B2B), business to customer (B2C), and business to business to customer (B2B2C) use cases.

Failing to consider identifiers, and especially usernames can have direct, negative impacts on the projects and systems you are working on.

Types of Identifiers

Identifiers come in two varieties: internal and external. Internal identifiers are the means by which a system refers to a digital identity. Formats of internal identifiers can vary greatly. One common format of internal identifiers is a universal unique identifier (UUID.) Specified in the IETF RFC 4122, UUIDs come in 4 variants or versions.[i] Many systems use UUID version 4 (often referred to as UUID4), which are randomly generated identifiers. An example of a UUID4 is: d5372288-697b-42bf-928a-562aca0deeaf.

But not all internal identifiers are UUIDs. Systems can use other means of uniquely identifying a specific thing from a collection of similar things. Examples include identifiers that have specific meaning to the system but are meaningless outside of the system, such as the following identifier, “005o0000000s4Hu.”

The second variety of identifier is an external identifier. An external identifier is the means by which a person in control of a digital identity refers to that identity when interacting with a system. These include but are not limited to a telephone number, email address, nickname, or handle.

In some cases a system identifier can be used by both internal and external purposes. Since email addresses must be unique within an organization i.e. a company, or domain i.e. college or physicians, the member’s ‘net id’ i.e. first part of their email address, will be used within corporate systems as a user’s identifier. A net-id could be comprised of first initial, second initial, last name and a number that ensures uniqueness.

Terminology

- Internal identifier: the way an identity management system refers to a digital identity
- External identifier: the means by which a person in control of a digital identity refers to that identity when interacting with a system
- Username: a common term used for an external identifier

Aspects of Usernames

When considering what the format of usernames should be, an identity practitioner must consider the five guiding principles of usernames. The practitioner should consider username format in greenfield situations as well when new B2C or B2B2C services are being created, at the very least. Often, especially in B2B scenarios, usernames have formats established in previous generations of systems, and those formats take on an almost mythic quality. It is not reasonable to simply change username formats, and needless to say, changing username formats, especially in an enterprise B2B setting, is not an undertaking one should take lightly.

Cloud applications are a potential area of username confusion. For a multitenant application, usernames should be common, i.e., the username for a digital identity in one application is the same as that used in another application. However, in some cases a user sets up an account in a SaaS application and selects another username. If this application is subsequently interfaced to the identity management environment a transformation mechanism will be required. API gateways or identity provider services maintaining multiple usernames are options.

The five guiding principles identity practitioners should consider are that usernames:

- Are not a secret
- Must be classified as public data
- Must be memorable
- Must be unique
- Must be recoverable

Secret

There is an instructive lesson in the United States' Social Security Number (SSN) as an anti-pattern for usernames.[ii]

SSN was meant as an internal identifier. Originally it was something the Social Security Administration would use to tie a human to their earned wages and eventually to their entitlements; it was something that they would use for their business processes. They shared this internal identifier with people and their employers to make business processes run. However, the use of this internal identifier grew. Businesses began to use SSN as a way for people to identify themselves to the business; in essence, business turned this internal identifier into a username. This secondary use was based on the idea that only the person would know their SSN and thus, because it was secret, the holder of the SSN would be assumed to be the correct person. And this is where things went wrong.

The need for this specific secret permeated so many of our business processes throughout our economy. This need has created a massive amplifier for damage when data brokers and others have breaches.

The lesson of SSN is that usernames cannot be secrets. If you share an internal identifier with a party outside of your organization, you have turned that internal identifier into public information, and thus it cannot be a secret.

If you have a username or an internal identifier that has to be treated like a secret, then you do have an authentication mechanism on your hands, not a username. And this means that it needs to be treated akin to a password.

As a pointer to an advanced topic for a later date, consider this- biometrics, broadly speaking, cannot be secrets. A person cannot keep their fingerprints, facial geometry, or irises secret. Because of this, a system or process can use biometrics as external identifiers. But because they are "just" identifiers, some degree of authentication is required to ensure the person actually intends to present their biometric. This degree of uncertainty is why liveness detection and attention detection are so crucial. For example, it is insufficient to accept a fingerprint biometric without also checking that the finger is real, has blood pumping through it, and isn't a fake made of gelatin.[iii]

Public

It is not enough to make sure that usernames are not secrets. Identity practitioners must also classify usernames as public in one's data classification scheme. This action applies to employees, partners, and customers alike.

Classifying usernames as public does not mean attributes related to the individual are public. Such attributes cannot reasonably or safely be used in a username. Consider a simple four-level data classification system:

- Public: this data can be shared across organizational boundaries freely and with a low level of concern.
- Restricted: this data is essential to business process and likely cannot leave organizational boundaries. Only data subjects, employees, and contractors can have access to this data.
- Confidential: this data is crucial to business operations. Significant harm may occur if this data transits organizational boundaries.
- Secret: this data is extremely organizationally sensitive. Only a small select group of people and systems can have access.

In an ideal world, an airline or hotel loyalty number (another kind of identifier) is likely classified "Restricted." Usernames must be classified as "Public." Airline or hotel loyalty identifiers demonstrate the problem of an identifier that is "public" but contains attributes that have value.

In addition, classifying usernames as public reinforces the idea that identifiers cannot be secrets.

As a clarification, the recommendation is that the username should be classified as public data, in a data classification system. That does not mean that usernames should be publicized (e.g., listed on a public site) – that is a self-inflicted enumeration attack.

Memorable

Part of the canon of US literature is Herman Melville's Moby Dick. And its first sentence reads "Call me Ishmael." Ishmael, the username, is not the most important part of that sentence – the "call me" part is. The power to name something is the power to control it. And by naming himself Ishmael takes control over himself, away from the Reader and away from the author.

In support of self-determination, people have to give themselves names, and in the digital world, this is crucially important. Usernames need to be self-generated in B2C and B2B2C settings, which is to say the person should have the power to create their preferred username. It is important to also consider self-generated usernames in B2B and B2E settings as well.

Many enterprises have a standard username format and they bring that preference to B2C and B2B use cases. A classic username format is First Initial, Last Name. For example, Sally Smith would get a username of “ssmith” and if that wasn’t unique a random number would be added. This habit-based username format, although reasonably effective, doesn’t support a desire for self-determination which is so crucial in B2C use cases.

Failure to support memorable usernames means increased account recovery calls, more on-screen help, and more customer support needs. And it also leads to duplicate identifiers because people often forget the identifier they used to register.

When building a username scheme, one needs to provide choice to the user. If asked for email as username and then on the next screen the user says, ‘do not use email to talk to me’, then there is significant cognitive dissonance. In order to provide choice, consider supporting multiple username schemes such as email addresses and user-created nicknames. Supporting multiple schemes adds a level of complexity, but the user empowerment that brings with it engenders self-determination and customer satisfaction.

Unique

Usernames need to be unique. Internal identifiers need to be unique. Neither statement should be controversial, but there is nuance here.

It is not enough to say a username must be unique; one must consider the scope of uniqueness. Is the username unique:

- at the individual service level?
- at the tenant level (if you are multi-tenant)?
- within a namespace with a service or set of services?
- globally across all of your services?
- universally?

Is there a clear picture of the scope being designed for? Even if there is, that picture may change; practitioners need to consider if the future might include merging internal systems or have to support various merger and acquisition activities in the future.

Also, uniqueness has implications depending on the type of identifier. Usernames and internal identifiers do not have to have the same scope of uniqueness. For example, while an internal identifier needs to be globally unique, a username might be unique only in a subset of systems in the enterprise. Internal identifiers have to be unique at the service-scope, e.g., unique in a specific enterprise service. To mitigate potential data subject reidentification, then those identifiers ought to be globally-scoped unique. Meanwhile, a person might use their email address to log into multiple systems - a service-level scoped unique username.

In addition, do not, in the same system, make the username and the internal identifier the same value. In some regards, this was one of the mistakes the US made with the Social Security Number.[iv] Practitioners should not make them the same value if only because changing either later can be enormously challenging. Furthermore, a common username scheme of choice is an email address, and these can change over a person's life based on life events such as marriage and divorce. Accommodating such changes to the username in a scheme where the username and internal identifier are the same requires that all systems with the "old" username/internal identifier need to be aware of the change and updated; in a complex environment, that task may be nearly impossible.

A final consideration is username reuse. Yahoo email allows people to use email addresses that were once used by someone else. Phone numbers are regularly reused. In this case, the username may still be unique but the person in possession of that username has changed. This transitional period is a difficult situation to be in if for no other reason than the new possessor of the email or phone looks like an attacker in many cases.

Recoverable

Usernames need to be recoverable, which is to say, that there needs to be a way to get a person back to their digital identity. Recovery means re-attaching the person to the digital identity; it does not necessarily mean they will use the same username over again.

In this regard, recovery is more than just reminding the person of what email address they used to log in. Consider telling a person that the email address they used was their old work email address that they cannot access. That leaves the person little recourse but to call the help desk… or move on to a new service.

Recovery is a re-association and to do this safely, it often requires re-proofing the individual is who they claim to be. Especially in B2C scenarios, such a re-proofing process requires considerable thought as it has significant security and customer satisfaction implications.

Conclusions

Identifiers are necessary to an identity system, with internal and external identifiers serving different purposes. While the two types of identifiers can be the same, the IAM practitioner should consider this with caution. External identifiers, also known as usernames, should consider these five guiding principles:

- Usernames should not be considered a secret.
- Usernames must be classified as public data.
- Usernames must be memorable.
- Usernames must be unique.
- Usernames must be recoverable.

Each principle has implications for the identity practitioner to consider as they develop an identity management system. Constructing a username framework is part of the ‘identity orchestration’ task.

[i] Leach, P., Mealling, M., and R. Salz, "A Universally Unique IDentifier (UUID) URN Namespace", RFC 4122, DOI 10.17487/RFC4122, July 2005, <https://www.rfc-editor.org/info/rfc4122>.

[ii] Carolyn Pucket, "The Story of the Social Security Number, Social Security Bulletin, Vol. 69, No 2, 2009, https://www.ssa.gov/policy/docs/ssb/v69n2/v69n2p55.html.

[iii] John Leyden, "Gummi bears defeat finger print sensors," The Register, 16 May 2002, https://www.theregister.co.uk/2002/05/16/gummi_bears_defeat_fingerprint_sensors/.

[iv] Pucket, see *Expanding Uses of the SSN*, https://www.ssa.gov/policy/docs/ssb/v69n2/v69n2p55.html.

Practical Implications of Public Key Infrastructure for Identity Professionals (v2)

By Robert Sherwood

To comment on this article, please visit our GitHub repository and submit an issue.

Table of Contents

Abstract

Public Key Infrastructure, or “PKI,” is a technology that enables authentication via asymmetric cryptography. It is widely deployed for some vital security use cases on the Internet, especially for the authentication of servers via Transport Layer Security (TLS).

Despite its wide use in some scenarios, there are significant challenges in deploying PKI for more widespread use among smaller organizations or consumers.

Identity Professionals who need to deploy a PKI or have inherited a deployed PKI from someone else have several important considerations, including lifecycle management of keys and certificates, choosing the appropriate way to encode user identifiers, and understanding cross-PKI trust.

Introduction

In high-risk environments containing extremely sensitive data, every participant must have high confidence in the identity of every other participant. Public Key Infrastructure (PKI) is one of the most long-lived and widely deployed authentication technologies in these high-security environments. Despite the difficulty in deploying PKI for end users, which we discuss below, PKI was the only high-assurance credential available in the commercial market for many years.[i] It is still considered the gold standard of credential assurance by many experts. Military and government environments have used PKI to provide secure authentication in sensitive environments since the late 90s.

Despite the widespread adoption of PKI in government environments, PKI has yet to see the same success in commercial settings. Later in this article, we will discuss some of the reasons for the lack of widespread adoption.

Despite the difficulties, PKI can be a feasible alternative to passwords for some enterprises, thanks to the implementation of smartcard-based authentication in many operating systems and browsers. Enterprises have renewed interest in smartcard login to eliminate passwords for privileged users in high-risk environments and scenarios.

This article includes analysis and guidance for the deployment of PKI for both human users and machines.

Terminology

- Asymmetric Cryptography: Any cryptographic algorithm which depends on pairs of keys for encryption and decryption. The entity that generates the keys shares one (see Public Key) and holds and protects the other (see Private Key). They are referred to as asymmetric because one key encrypts, and the other decrypts.

- Automatic Certificate Management Environment (ACME): A communication protocol for automating lifecycle management of PKI certificates. Significant providers like Let's Encrypt leverage ACME to support issuing TLS certificates for web servers.
- Certificate Authority Trust List (CTL): A client maintains a list of trusted Certificate Authorities created and managed by the software provider or local administrators. The client will only trust certificates issued under one of the CAs in the CTL, so the CTL serves as a "safe list."
- Certificate Management System (CMS): A system that provides management and reporting layers for certificate issuance and revocation. A CMS integrates CA products with Identity Governance and Administration (IGA) systems as well as Service Desk systems.
- Certificate Policy (CP): A document that defines the high-level policy requirement for a PKI. RFC 3647 identifies a PKI's policy framework and describes a CP's contents and outline. An enterprise operating a CA will often publish its certificate policy to external parties so they can determine whether to trust certificates issued by the CA.
- Certification Practices Statement (CPS): A CP identifies the requirements for managing a CA and issuing PKI certificates. A CPS describes how a CA implements those requirements. The CPS uses the same outline as the CP, defined in RFC 3647. Unlike the CP, enterprises rarely publish their CPS in unredacted form.
- Certificate Revocation List (CRL): A certificate authority will publish a list of revoked certificates, called a CRL so that clients can verify that a certificate is still good.
- Certificate Signing Request (CSR): When requesting a certificate, the requesting entity provides a copy of the public key, their identifiers, and other information in a specially formatted binary object called a CSR.
- Classical Computer: A computer that uses binary encoding and Boolean logic to make calculations in a deterministic way. We use the term Classical Computers in contrast with Quantum Computers.
- Cryptographic Module: A hardware or software component that securely performs cryptographic operations within a logical boundary. Cryptographic Modules store private keys within this boundary and use them for cryptographic functions at the request of an authorized user or process.
- Cryptographic Module Validation Program (CMVP): A program allowing cryptographic module developers to test their modules against the requirements defined in FIPS-140. The computer security resource center under the United States National Institute of Standards and Technology (NIST) maintains a publicly available list of validated modules.
- Electronic Identification, Authentication, and Trust Services (eIDAS): European legislation gives legal standing to electronic signatures under eIDAS. This legislation also documents providing legally binding digital signatures with X.509 certificates to comply with Qualified Signature requirements.
- Elliptic Curve Cryptography (ECC): An asymmetric cryptosystem based on calculating points along elliptic curves.

- Encryption: Processing data using a cryptographic algorithm to provide confidentiality assurance.
- Federal Agency Smart Credential Number (FASC-N): A unique identifier associated with a smart card. FASC-N is used in the US Federal Government PIV standard to support Physical Access.
- Federal Information Processing Standard (FIPS) 140: A NIST standard defining "Security Requirements for Cryptographic Modules."
- Groups: A set of identities with defined permissions. In this specific context, a group contains many individuals, but the group identity is opaque, and no information is available regarding which group member took an individual action.
- Hardware Security Modules (HSMs): A hardware cryptographic module that generates and protects cryptographic keys.
- Identifier: The way a system refers to digital identity. PKI Certificates support both internal and external identifiers. See Ian Glazer's article, "Identifiers and Usernames."[ii]
- Internet Key Exchange (IKE): A subordinate standard under IPsec specifying how to use X.509 certificates to establish symmetric keys for an IPsec tunnel.
- Internet Protocol Security (IPsec): A standard for communication between two machines providing confidentiality and integrity over the Internet Protocol.
- Key: In a cryptosystem, a Key is a piece of information used to encrypt or decrypt data in a cryptographic algorithm.
- National Institute of Standards and Technology (NIST): A US Government agency that defines and publishes various standards. One department within NIST, the Computer Security Resource Center (CSRC), publishes the Federal Information Processing Standards (FIPS) series. While these standards are only mandatory for US Government Agencies, many countries recognize them as de-facto global standards.
- Non-person entities (NPE): Any unique combination of hardware and software firmware (e.g., device) that utilizes the capabilities of other programs, devices, or services to perform a function. Non-person entities may act independently or on behalf of an authenticated individual or another NPE.[iii]
- Online Certificate Status Protocol (OCSP): A protocol that allows a client to query the Certificate Authority or a Validation Authority for the status of an individual certificate rather than downloading a CRL.
- Path Discovery and Validation (PDVal): The process to determine whether a certificate is valid and trusted by the validator.
- Personal Identification Number (PIN): A numeric secret commonly used to unlock a private key container in software or hardware.
- Personal Identity Verification (PIV): A US Government program designed to enable strong authentication for all government employees and contractors, based on Public Key Infrastructure.
- Private key: A key that a single entity exclusively and privately controls. It corresponds to a public key that the entity may share for data encryption or signature verification.

- Public key: A key that an entity publicly distributes. It corresponds to a private key that the entity exclusively and privately controls.
- Public Key Certificate: A certificate containing a public key, one or more identifiers for the private key holder, an identifier for the Certificate Authority, and additional metadata to support security requirements.
- Public Key Infrastructure: A set of tools, standards, and related policies designed to manage trust based on public/private key pairs and certificates.
- Registration Authority (RA): An individual, system, or business function which provides registration and identity proofing for entities receiving certificates and manages the certificate issuance and renewal process. The most important responsibilities of an RA include identity proofing and binding the private key to the identity.
- Revoke: Revocation is the announcement that clients should no longer trust an individual certificate.
- Roles: A set of permissions. A role must be associated with an individual user, and the user gains the associated authorization when they are associated with the role.
- RSA: An asymmetric cryptosystem based on large prime numbers. The acronym RSA stands for the three principal inventors, Ron Rivest, Adi Shamir, and Len Adleman.
- S/MIME: A standard for constructing and sending digitally signed or encrypted messages using asymmetric cryptography.
- Secure Socket Layer (SSL): A deprecated standard for encrypting data in transit; TLS has superseded it.
- Server-based Certificate Validation Protocol (SCVP): A protocol that allows a client to query a server to determine whether a certificate is valid and trusted. The server does not need to be associated with the issuing CA. SCVP does two things; (1) it determines the path between the end entity and the trusted root, whereby the client doesn't need to trust any intermediate CAs. (2) it also performs delegated path validation according to policy.
- Signature: Processing data using a cryptographic algorithm to provide integrity assurance.
- Subject Alternative Name: One or more identifiers for a certificate subject that certificate issuers can use to carry application-specific identifiers such as email address or User Principal Name (UPN).
- Subject Distinguished Name (Subject DN): A unique identifier for the subject within the scope of the Certificate Authority. Issuers structure the subject DN like an LDAP entry name.
- Transport Layer Security (TLS): A cryptographic protocol designed to provide confidentiality and integrity of communications between two endpoints.
- X.509: An ISO standard from the X.500 series that defines the basic rules for encoding public key certificates.
- Validator: An entity that verifies a certificate and confirms that the other party controls the private key in the transaction.

Basics of PKI for Identity Practitioners

What is PKI

PKI stands for "*Public Key Infrastructure,*" a set of interlocking standards and technologies that support the secure exchange of public keys for *asymmetric cryptography* use cases.

Originally developed as part of the X.500 series of specifications for electronic directory services, the *X.509* standard proposed a way to link a public key into a universal, hierarchical directory designed to support OSI networks.

The OSI protocol is, for all intents and purposes, dead. However, the X.500 specification lives on in simplified form as LDAP, and X.509 has found a second life in the modern Internet.

PKI is woven deeply into the fabric of the Internet, and it supports the following critical Internet capabilities:

- *TLS* as a general encryption layer for application protocols
- *S/MIME* as a standard for secure email
- *IPsec* as a standard for virtual private networking (VPN), which depends upon PKI via the Internet Key Exchange (IKE) extension
- Some commercial software or services, such as Adobe Acrobat, Microsoft Word, or DocuSign, support digital signatures for non-repudiation or integrity protection. In Europe, qualified signatures and time stamps have official legal standing, recognized in the Electronic Identification, Authentication, and Trust Services (eIDAS) framework.

This article is not a general primer on PKI; it provides a minimal overview of PKI as it relates to identity Management and identifies critical issues relevant to identity Practitioners. Interested readers are referred to the references section at the end for more detail.

Here are some excellent resources to learn more about PKI in general:

Books:

- Applied Cryptography, by Bruce Schneier, is a classic guide to the cryptographic technology underlying PKI and its applications. For those who want to know everything about this subject, this is the place to start.

Online Resources

- The US Federal government has deployed PKI widely for both logical and physical access. IDManagement.gov maintains information about the Federal PKI here: https://playbooks.idmanagement.gov/fpki/

- Bruce Schneier, the author of Applied Cryptography, maintains a fascinating and helpful blog here: https://www.schneier.com/

Standards:

- X.509: The original specification for PKI certificates. This document must be purchased.
- RFC 5280: The Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile standard specifies a subset of the X.509 standard for use on the Internet.

How do a 'Private Key' and a 'Public Key Certificate' Provide Authentication Assurance

Public and Private Keys

Private and *public keys* are random numbers, but not just any random number.

- In the *RSA* specification, keys are derived from a large prime number.

- In *ECC,* keys are related to points along a particular elliptical curve.

By taking some data, such as text or an image, and plugging the data into a specific equation with one of the numbers (keys), you create a scrambled version of the data that only the other number (key) can unscramble. This concept is the basis of asymmetric cryptography.

The *private key's* owner must retain and closely guard it since a foundational assumption in PKI is that only the authorized user controls the private key. The public key, by contrast, can be widely shared.

A sender can scramble a message using the public key to send a message only the private key's owner can read. Because the private key is the only key that can unscramble and read the message, the sender knows that the message can only be read by the private key owner.[iv]

The owner of a private key can use it to scramble a message, and a recipient can only unscramble the message with the public key. The recipient can be sure that the private key owner sent the message and that it has not been modified in transit.[v]

In asymmetric cryptography, “*encryption*” refers to scrambling data with the public key, and “*signature*” refers to scrambling data with the private key.

In practice, signature and encryption are much more complicated, involving cryptographic hashes or intermediate symmetric keys. For our purposes, it is sufficient to understand that private keys sign and public keys encrypt.

![Illustration of Encryption and Signature path for a document](PKI.jpg)

Despite the widespread use of PKI for highly secure credentials, asymmetric cryptography does not directly provide authentication! Authentication protocols that leverage PKI credentials depend on signature or encryption.

In public-key authentication schemes, the user is whoever has control of the private key. When a system wishes to authenticate a private key owner, it requires them to use the private key they own. The user can sign something with the private key that the system can verify with the public key or decrypt something with the private key that the system encrypts with the public key.

The user can provide a signed message for the authenticating system to verify, or the authenticating system can generate and encrypt data that the user can only decrypt with their private key. In both scenarios, possession of the private key, demonstrated by the ability to use the private key to decrypt or sign data, proves the user's identity.

Public Key Certificates

So far, we have seen how to authenticate a user if you have their public key. The challenge that remains is finding a reliable way to exchange public keys.

We have solved this problem in several ways with different protocols and systems. Many modern authentication protocols, including FIDO, Verifiable Credentials, and passkeys, leverage public/private keys and asymmetric cryptography. Every protocol requires an out-of-band process that links the public key with a public identifier. In some contexts, notably in the SSH public-key authentication protocol or "Web of Trust" based systems like PGP, pre-existing relationships or pre-provisioned authorizations are sufficient.

In the context of complex modern business processes where you are unlikely ever to meet the majority of people you interact with, users cannot simply exchange keys directly. After all, in the absence of some way to verify the identity of the individual providing you with a public key, you have no way to distinguish between your intended counterparty and an imposter.

PKI solves this issue by relying on the concept of a "trusted third party." In a PKI, a central trusted authority vouches for identities according to a documented process. This centralized authority introduces scalable trust by allowing users to verify the identity of previously unknown users or systems. The users rely on the centralized authority to enforce an identity registration and lifecycle management process. The mechanism used in PKI to convey this assurance is the public-key certificate.

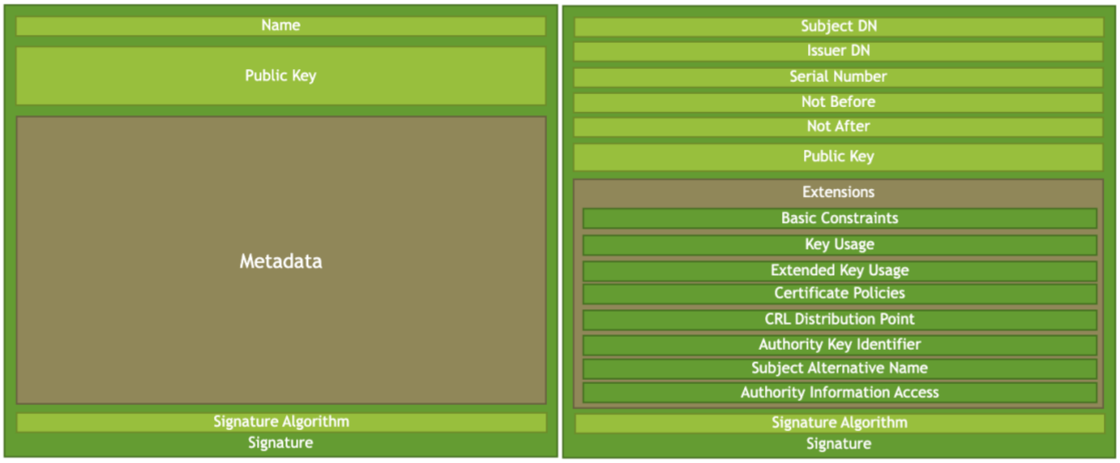
For every participant to have confidence in distributed business transactions, each participant must have confidence in the identity of every other participant. For asymmetric encryption to support business applications, the public key must be connected, or "bound," to the participant's identifier. In PKI, public key certificates are the artifacts that connect a public key and an identifier.

A *public key certificate* contains several critical pieces of information. For authentication purposes, the following three fields are the most important:

- The public key

- One or more identifiers associated with a user

- Information about the “trusted third party” that vouches for the association between the key and the identifier.



Figures 1 and 2: Key components of a PKI certificate that support identity including Name, key, metadata, signature algorithm, and signature. Additionally, a detailed listing of several possible elements of a PKI certificate.

A public key certificate is a file with a prescribed structure defined by the X.509 v3 standard and refined by RFC 5280. It contains the user's public key, their identifiers, and important metadata about the certificate itself.[vi] The file is digitally signed using the private key of a trusted third party, called a "Certificate Authority."

Who Can Get a Certificate

Any business process participant who can generate and store a private key and associated public key may receive a certificate. The most common recipients of certificates are listed here:

- Humans: A human being can receive a public key certificate that names them individually.

- *Non-person entities:* Examples of non-person entities include devices like routers, software services like web or email servers, IoT devices, and other non-human entities like software providers who digitally sign software packages.

- *Roles:* Sometimes, a person may act in a role, such as "Software Release Manager" or "Doctor on call." A certificate authority can issue a certificate that identifies the user's role, allowing them to authenticate in the persona of that role. Role certificates are issued to individuals and contain a personal identifier for the person holding the private key to maintain individual accountability. Everyone with a role certificate has a unique private key.

- *Groups:* In some exceptional cases, several people share a private key. In this case, a certificate authority can issue a certificate to a group. The certificate will identify the group, and the group members will take additional security precautions to ensure that only authorized members use the private key.

How Are PKI Certificates Like Other Credentials, and How Are They Different?

Users can authenticate themselves with a private key and corresponding PKI Certificate, like other credentials.

- The trustworthiness of the credential depends on the identity proofing and issuance process as much as it depends on cryptographic math. As with other credentials, the identity assurance level for authenticated users is low if the proofing or issuance processes are insecure or the user does not protect the private key.
- Like other credentials, a private key and certificate are a single authentication factor that enterprises can supplement with additional factors. Typically, we consider a key and certificate "something you have" and often supplement it with a PIN, Password, or biometric.

PKI credentials have many unique properties not shared by most other authentication credentials.

A public-key certificate file contains all the information necessary to authenticate the subject:

For most other credential types, each authentication challenge requires the involvement of the credential issuer. When a user enters a password, the authenticating system must check it against the directory or database where the user created their account. By contrast, PKI authentication can occur without directly interacting with the issuing Certificate Authority. The user generally activates the private key with a secret, such as a

PIN or a password, but his secret is entered directly into the software or device containing the private key; the user does not provide it to the Certificate Authority.

Public key certificates are long-life credentials:
Certificates may be valid for a much longer-term than is typical for other credential types. It is common for a certificate authority to issue a public key certificate to a user with a three-year lifetime. This extended lifetime is acceptable because the private key credential is not user-selected and is too long to be easily memorized or copied by humans.

Key protection affects the overall security of the PKI credential:
Like any other authentication secret, the user must protect a private key from third parties to prevent the third party from impersonating the user. Recall that in public-key cryptography, the user is whoever controls the private key. For this reason, it is essential to ensure that private keys cannot be copied or taken without a user's awareness and permission. Because private keys are usually very long and appear random, they cannot be memorized and must be stored.

Several technologies are available to protect private keys, including *Hardware Security Modules (HSMs)* or personal tokens such as the YubiKey Security Key or SafeNet eToken Smart Card. The United States *National Institute of Standards and Technology (NIST)* has published a standard, *Federal Information Processing Standard (FIPS) 140,* and has implemented the *Cryptographic Module Validation Program (CMVP)* to ensure that HSMs implement proper cryptographic algorithms and key protections for private keys.

The security properties of PKI credentials mean they can provide a higher level of identity assurance than other kinds of credentials. Governments reserve the highest levels of assurance defined by governments for PKI certificates stored on smart cards. This security comes at a price in terms of direct costs and additional complexity.

PKI credentials can support additional use cases beyond interactive authentication:
While passwords, OTP, and other credentials are limited to interactive authentication, PKI credentials are suitable for transactions that are not immediate and interactive. One example is a digital signature, where the recipient of a signed message must know the signer's identity, but the signer may not know who will verify the signature. Encryption is another case where the encryptor of the sensitive data must ensure that the intended recipient is the only one who will have access even when data is exchanged out-of-band and asynchronously. PKI can enable capabilities such as S/MIME, Qualified Signatures, and others that cannot be supported by credentials that only provide authentication.

Factors and Problems Limiting PKI Adoption

The roots of PKI extend back to the 1970s, and the earliest versions of the Secure Sockets Layer (SSL) standard cemented its use as the basis for secure communication in the mid-1990s. However, despite its maturity and widespread use for some specific use cases, it has yet to see broad adoption for authentication of individuals, either for business-to-consumer or business-to-employee use cases. There are many reasons why PKI has yet to see widespread adoption outside these narrow use cases, though the technology and vendor support has improved. The following are some of the most significant challenges hindering adoption:

Enterprise key management is challenging:

For PKI to be a trustworthy and secure authentication approach, the private key must be controlled exclusively by the authentication subject. As we said earlier, the user is whoever controls the private key. There are two ways to ensure that the intended user is the only one with access to the private key. The authentication subject must generate the private key within a protected software environment, or the CA must generate the private key on the subject's behalf and then pass it to the subject using a secure transfer mechanism. Both processes are complex and challenging to automate without extensive tooling.

Internet software providers have focused on providing automation for critical technical use cases, such as TLS for Web Servers. Protocols like *Automatic Certificate Management Environment (ACME)* and services like Let's Encrypt provide zero-touch key management and certificate rotation for web servers. These services do not support the management of certificates issued to humans.

Vendors, meanwhile, have implemented sophisticated, proprietary solutions for the automation of key management. Microsoft Active Directory Certificate Services can provide key management and certificate services for machines and human users in an Active Directory environment. The Entrust Certificate Authority provides a client-side tool to manage the lifecycle of keys and certificates for clients. However, these tools and others like them are tied to a specific product and are part of a closed, proprietary system.

Other providers, like KeyFactor or Venafi, can provide certificate lifecycle services for a mix of CA products. However, these tools are proprietary and may require significant integration efforts.

PKI has poor usability:

As discussed above, key management is a complex organizational and technical issue with its share of challenges. Unfortunately, many PKI implementations require end-users to manage much of that complexity. Notably, users must initiate the key generation and request process. Once a user generates a private key and the CA issues a certificate, the user must configure all of their tools (operating system, web browser, mail client, etc.) to use the private key generated by the user and manage the list of trusted certificate issuers.

Sophisticated enterprises with dedicated engineering teams should be able to handle this complexity on behalf of the user community. Still, this complexity is difficult to manage even in highly controlled environments. This complexity is unmanageable for most small businesses and home users.

One way to address this user challenge is to have a designated administrator or security officer who assists users in generating their private keys and initializing their tokens. This approach is widespread in large enterprises and can also be feasible for smaller companies.

In high-security environments, users and administrators generate private keys on a hardware security module. This hardware requirement adds device driver installation and management issues to the other problems confronting users attempting to use PKI for authentication. Some platform vendors have implemented platform-level API (e.g., Microsoft CAPI). Still, support for this API is not universal, with some applications implementing proprietary or platform-neutral key storage systems that do not integrate with the host OS.

As with many IDM technologies, enterprises should observe the 80/20 rule. IDM professionals should ensure that critical or widespread user applications support your PKI implementation and accept alternative credentials for essential legacy applications.

Public key enablement of applications is hard:
We have discussed the difficulty of using PKI for authentication from the perspective of Authentication Subjects. Enabling applications to consume PKI credentials is even more challenging in some ways:

1. The list of trusted certificate issuers must be maintained and synchronized across all applications where the user may need to authenticate.
2. Applications must validate certificates, which requires the applications to access a public HTTP site or LDAP directory to obtain Certificate Revocation information.
3. A local user profile must be created in the application based on an identifier present in the certificate or entered by the user during a manual registration process.

There is no concept of provisioning or de-provisioning built into PKI by default, so applications must implement this capability through a separate integration with the Registration Authority (RA). Since it is common for users to authenticate with a site directly, CAs rarely offer this capability. Identity professionals should leverage existing directory technologies, such as Active Directory, to support user profiles for multiple applications.

For internally-facing enterprise applications, an IGA system may manage these aspects. Across enterprise boundaries or in a B2C context, this additional complexity makes PKI credentials difficult and expensive compared to other authentication technologies.

Certificate trust path discovery and validation are complex and existing implementations have inconsistent behavior:
In the previous section, we discussed applications needing to validate the certificates. This validation is complicated, even when administrators configure applications to use a static Trust List of known good issuers.[vii] To complicate things further, PKI supports a form of federation through cross-certification, discussed below in more detail. In this section, we will note that determining whether a trusted partner issued a certificate in a federated or cross-certified environment is very challenging.

Path Discovery and Validation (PDVal) is complex. Different vendors implement it inconsistently. One application may treat a certificate as valid, while another application may reject the same certificate, depending on the underlying certificate validation library. Some third-party solutions support consistent PDVal across products, but they must be implemented and integrated with each endpoint. This burden has made enterprises leery of implementing PKI on the server side.

Unique Considerations for Identity Practitioners

Ensure that PKI is the Right Fit for Your Requirements

Deployment of PKI involves several complexities and difficulties outlined in this document. However, PKI is a powerful tool that can offer strong authentication and support other use cases, such as email signing/encryption, that are impossible with other strong authentication credentials. When considering the deployment of PKI, ensure that the use cases you can support justify the added complexity for your environment and your users.

For TLS and link encryption, PKI may be the best or only choice, but that does not necessarily mean that you should implement your own local PKI. A third-party PKI service provider is an excellent alternative for most organizations.

The Importance of Planning

If you determine that an internally managed PKI is the correct choice for your organization, planning is critical for a successful PKI deployment. While the need for planning is not unique to PKI, the complexity of a PKI environment can make retroactive cleanup much more complex than careful up-front planning and deployment. As with any Identity Management technology, planning is critical to success.

IGA and PKI

Enterprises that leverage Identity Governance and Administration tools may need to expand their toolkit to accommodate PKI credentials. Existing IGA tools can manage accounts and privileges but may not track the PKI credentials associated with the managed accounts. It is important to recall that a PKI certificate and private key represent self-contained credentials that are still valid even if the underlying account has been deactivated or deleted. Unless the certificate is revoked or has expired, external applications may still accept a PKI credential as valid.

The challenges of managing non-human accounts such as machines, IoT devices, Bots, or other entities also apply to certificates issued to non-person entities. Refer to "Non-human Account Management (v3)" by Graham Williamson, André Koot, and Gloria Lee for information about unique issues related to these entities.[viii] The section below, 'Machine Identities and Certificate Management Systems,' discusses machine identities in more detail.

Many CAs include management capabilities to address these challenges. Some third-party (Certificate Management System) CMS products interact with multiple CA products to provide a single pane of glass for certificate management in a multi-vendor multi-CA environment. A later section discusses these products in more detail.

Lifecycle Management of PKI Certificates Compared to Other Credentials

Modern cryptographic algorithms ensure that private keys cannot be easily guessed. For example, a *classical (non-quantum) computer* would need about 300 trillion years to break a 2048-bit RSA key, while the same computer would require an average of five sextillion seven hundred eighty-three quintillion + years to guess a 128-bit ECC key.

However, the security of an overall system rarely depends exclusively on math.

The overall security of a PKI system includes several variables, including unreliable humans. CAs issue end entity certificates for a relatively short time, such as 90 days for public SSL certs or up to years for human subscriber certificates. CA certificates may be valid for as long as 20 years. This lifetime is much longer than a typical password or other credentials because the private key is never directly presented during authentication. The CA must store its private keys in a Hardware Security Module to ensure that an attacker cannot copy them.

Because CAs issue certificates with a fixed lifetime, key management can become a significant challenge. Enterprises should deploy a Certificate Management System (CMS) to monitor certificates and automate the renewal process or provide notification when a renewal is required. Most CA products include a rudimentary management console, but CMS products can offer a single pane of glass to manage multiple CAs from different vendors. CMS systems can also provide Service Desk support tools for assisting in smartcard registration and forgotten/locked PIN issues.

As with any other type of credential, a certificate may become invalid before it expires for various reasons. A user may leave the organization, change roles, or lose access to the private key. PKI provides for the revocation of public-key certificates in this case. The list of "no longer trusted" certificates is called the Certificate Revocation List (CRL). In every certificate, the CA publishes a URL where the CRL can be found. Alternative protocols, such as the Online Certificate Status Protocol (OCSP), offer other means of checking the validity of a certificate. Most web browsers have implemented proprietary revocation-checking techniques.

A third technology, called Server-based Certificate Validation Protocol (SCVP), has been developed and documented in a standard but has not been widely implemented. It is mentioned here for completeness, but most enterprises can disregard SCVP.

Because a certificate passes between the subject and a third party without involving the original issuer of the certificate, it is imperative that applications correctly validate certificates and check the revocation information.

Options for Identifiers in Public Key Certificates

The primary purpose of the certificate, as described above, is to link a public key with a user identifier. Of course, a user may have several identifiers for different use cases. Rather than issuing separate certificates for each user identifier, the PKI specification supports including multiple identifiers in a single certificate.

The primary user identifier in a certificate is the Subject Distinguished Name (Subject DN). The Subject DN must be structured like an LDAP Distinguished Name. Typically, there will be a "Base DN" shared by all certificates issued from a Certificate Authority and one or more "Relative DNs" which differentiate certificate subjects. Common relative DNs include "Organization" and "Organizational Unit." Finally, the certificate lists a subject's unique identifier. This identifier can take several forms, such as "Common Name," usually a user's Legal Name. In large PKI deployments, users with frequently seen names may have other identifiers embedded or appended to their names to distinguish between users with the same legal name. The common name is not the only possible identifier for a user in a Subject DN. Certificates can also use UID or email address to identify a certificate subject.

Because the Subject DN must mimic an LDAP Distinguished Name, it is very restrictive. For this reason, certificates often use an additional field instead. The “*Subject Alternative Name”* field is a much more flexible option to encode different user identifiers. It allows multiple names to be encoded and does not mandate a particular structure. Common uses for the subject alternative name field include:

- Email address to support S/MIME digital signature and encryption

- UPN to support smartcard login on the Windows platform
- Hostname to support TLS connections

The Subject Alternative Name does not impose any constraints on the type of identifiers that can be encoded. So, in addition to all of the previously listed identifiers, private communities of interest may insert identifiers that have strictly local meaning into this field. An example is the *Federal Agency Smart Credential Number (FASC-N),* which is part of the US Federal Government's Personal Identity Verification (PIV) standard.

Generally, Enterprises should use the subject alternative name for the user or machine identifiers. The Subject DN must be unique but should not contain multiple identifiers or non-standard ID types.

Machine Identities and Certificate Management Systems

While PKI has not seen widespread adoption as a credential for people,
it is dominant as a credential for machines, thanks to its use in TLS. TLS is not only used to provide secure access to web servers in end-user browsers; it is also widely used as a tunneling technology in machine-to-machine or site-to-site communication.

Virtualization and containerization technologies and the use of cloud providers have exploded in recent years. For this reason, the number of PKI-based machine identities is increasing exponentially. Managing and tracking the keys and associated certificates is becoming a significant challenge.

A Certificate Management System is an increasingly critical tool for enterprises to deploy to avoid service outages due to expired certificates, especially for enterprises with hybrid-cloud-based infrastructure or multi-vendor server environments.

Federated Authentication and PKI

We have already seen a critical difference between PKI and other credentials - a user can authenticate to an external application without involving the issuing authority in every transaction. This property can simplify authentication flows but places a greater burden on external applications since they validate the certificate themselves.

For an external application to consume certificates issued by your certificate authority, the application must trust your certificate authority. There are two basic ways for an application to trust an external certificate authority: explicit trust using certificate trust lists or implicit trust based on cross-certification.

Explicit trust is the most commonly used approach. In this model, applications explicitly managed trusted issuers within static Certificate Authority Trust Lists (CTL). The location of the trust lists will vary from product to product and system to system. A Java virtual machine, the Windows Operating System, and most web server software all maintain individual trust lists. Synchronizing them all in an Enterprise environment can be a very complex challenge. If you leverage an internal PKI, a CMS product can automate much of the management of disparate trust stores. Vendors typically configure standard software packages to trust the more prominent commercial Certificate Authorities. This is another good reason to acquire certificates from these sources.

Implicit trust relies on a technique known as cross-certification. In cross-certification, a CA will issue a certificate to another CA, typically operated by a different organization. From the perspective of an application validating certificates, the external CA appears to be another internal CA connected to the CA issuing the cross-certificate. The promise of cross-certification is that it dynamically allows applications to discover trust relationships between independently operated certificate authorities. In practice, however, the inconsistent behavior of PDVal implementations in software validating trust relationships between CA has prevented the promise of cross-certification from being fully realized. For most enterprises, cross-certification is not a helpful tool for federated authentication.

Finally, Identity Federation technologies can simplify the implementation of cross-domain trust by providing assertions across enterprise boundaries rather than relying directly on crossPDVal. The certificate can be validated within enterprise boundaries using relatively straightforward reliance on trust lists and local revocation publication. An SSO product can provide a federated token to external applications. This will address the interactive authentication use case but will not solve the challenges associated with other use cases that PKI can support, such as secure email encryption and signature or digital signatures for documents.

Conclusion

PKI is a powerful but complex tool for highly-secure authentication. It is likely already used within your environment for NPE or machine identities. Identity professionals should investigate the tools and processes used by individual programs to minimize redundancy of effort and cost.

Carefully weigh the benefits of the use cases within your environment before committing to deploying the technology to end users. If you choose to deploy PKI, avoid the temptation to introduce local or proprietary extensions, and stick to widely supported standards.

If an enterprise identity management environment is needlessly complex, it significantly complicates PKI deployment. Before deploying PKI, or any other complex authentication technology, ensure that identity management tools and practices are rationalized and streamlined within the enterprise environment.

If you introduce PKI for end-users, consider deploying a Certificate Management System to track the lifecycle of keys and certificates across your entire domain.

Author Bio

Robert Sherwood is the Principal Consultant at Credentive Security, a boutique consulting firm focused on Identity Strategy and Architecture.

Change Log

Date	Change
2021-09-30	V1 published
2022-12-15	V2 published; content significantly updated

[i] Note that credential assurance is distinct from identity assurance. Identity assurance measures how well you verified the identity of the account holder and how securely you connected the identity to the credential at the time of issuance. Credential assurance measures how confident you can be that the credential subject has maintained control over the credential, and that the credential has not been compromised.

[ii] Glazer, I., (2020) "Identifiers and Usernames", *IDPro Body of Knowledge* 1(1). doi: https://doi.org/10.55621/idpro.16.

[iii] Williamson, G. & Koot, A. & Lee, G., (2022) "Non-human Account Management (v3)", *IDPro Body of Knowledge* 1(7). doi: https://doi.org/10.55621/idpro.52

[iv] Technically, the sender generates a symmetric key, encrypts the message with the symmetric key, and then encrypts the symmetric key with the intended recipient's public key.

[v] Technically, digital signing appends a 'hash' to the document that can be deciphered by the sender's public key - ensuring the sender's identity.

[vi] International Telecommunications Union – Technology (ITU-T), *X.509 : Information technology - Open Systems Interconnection - The Directory: Public-key and attribute certificate frameworks*, October 2019, https://www.itu.int/rec/T-REC-X.509.

[vii] For example, see this article on how browsers handle revocation checks: https://www.ssl.com/blogs/how-do-browsers-handle-revoked-ssl-tls-certificates/.

[viii] Williamson, Koot, and Lee, "Non-Human Account Management (v3)."

A Peek into the Future of Decentralized Identity (v2)

Leo Sorokin

Table of Contents

Abstract

As digital transformation sweeps across the globe, it has affected everyone – from citizens to employees, from corporations to governments. Digital identity is a foundational enabler for business processes in the digital economy. Decentralized identity is the next evolution of digital identity capabilities and brings with it an opportunity to streamline how people interact with other institutions, physical objects, and with one another. This paper considers the future world of decentralized identity and offers clarity around the benefits of decentralized identity, terminology, sample scenario, and a sample technical implementation, while also addressing some of the limitations of this model. This paper further grounds the reader in the current state of decentralized identity capabilities while outlining the evolution of identity practices from past to present.

Introduction

Digital identity is rapidly gaining criticality in our world as organizations digitally transform. Identity plays a pivotal role in a digital transformation and can empower both governments and businesses to provide secure whilst restricted access to data for any stakeholder whether employee, partner, customer, or citizen. Digital identity is becoming a vital component of security in a world with data proliferation on a myriad of devices and a network perimeter that is ever-more challenging to define.

One active area under development in the identity space is the concept of *decentralized identity*. Decentralized identity is a fundamental shift from *account-based credentials* toward *verifiable credentials* and is a major philosophical as well as technical change in the way identity-related information is acquired and presented. The World Wide Web Consortium (W3C) is working on publishing standards around *Verifiable Credentials* and *Decentralized Identifiers*.[i],[ii] However, as with any technology standard, it must be broadly adopted by the community for it to be useful at scale.

Today, a person's digital identity (and associated personal data) is strewn across many online services, with access to such services being primarily performed via a username and password. Such an account-based credential is usually provisioned directly by the service provider, or by a large and rather centralized identity provider (IdP), such as Google, Facebook, or Twitter with which a service provider application will federate. This account-based federated model, however, has some significant limitations: the IdP may stop offering its services to third-parties; the identity supported by this IdP may be compromised thus impacting every service provider application that uses that identity; the IdP may track an individual's activities across multiple services; and an IdP may decommission the account being used for authentication. There are many challenges with the federated identity model, but going back to identity silos where each service provider provisions and manages its own set of credentials for its users, resulting in users having to manage dozens of such account-based credentials is not ideal either.

Decentralized identity strives to place the individual at the center of digital identity experiences by attempting to insert the individual at the center of identity data exchange. At its simplest, decentralized identity attempts to map physical wallets and the physical cards within them to a very similar concept in the digital world – a digital wallet with digital cards.

Today, there are many that are very excited about the potential of this model as well as many that are skeptical. Although decentralized identity and the concepts underpinning it attempt to solve the challenges we have had with digital identity over the past few decades, it is still too early to predict how individuals, governments, and corporations will approach it, and how each of these actors will be able to derive value from it.

Decentralized Identity Benefits

A decentralized identity system can be used to replace a traditional username and password during a typical *authentication* sequence. This is perhaps the first use-case most will think about. However, authenticating in a passwordless manner is possible today even without any decentralized identity components. As such, the true value of decentralized identity can be more easily understood during *authorization*. During authorization, the service provider may mitigate risk by requiring the individual to present one or more digitally signed attestations commensurate with the level of risk that specific transaction entails and the level of value being obtained. This capability could be leveraged to increase trust between the parties, improve the user experience for the individual, while at the same time lowering costs for the business.

The purpose of decentralized identity is to empower individuals to own and control their digital identity and how their identity data is accessed and used. The premise behind decentralized identity decouples it from the notion of a username and password or the traditional account-based model. A digital identity is not yet another username and password-based account that is provisioned and maintained by a third party. With a decentralized identity model, the individual can be both authenticated and authorized to perform a transaction with one service, and then present the same identity information to another entity with which the individual might prefer to interact. In addition, the individual can become their own identity provider, which is more difficult to accomplish with the centralized or federated models we have today.

Decentralized digital identity and the *personal data* associated with it should enable the individual to have more control over how that data is accessed and used. As a byproduct of this philosophy, personal data should be presented by the individual to service providers on an as-needed basis, with specific terms of use. This principle is fundamental to decentralized identity. In a decentralized identity ecosystem, there is no one single central authority; value is exchanged in a more peer-to-peer manner. Since the individual controls and owns their personal data, they are the ones to enable other parties to access it by granting them specific permissions. This is in stark contrast to today's reality where personal data may be shared and stored by third parties outside the individual's control with the individual having no means of specifying the terms of use under which the identity-related information is shared.

In a decentralized identity environment, it may be possible to possess a digital card for a drivers' license, credit card, or even a passport, and have them available on a mobile device. In another scenario, it may help when traveling abroad while having to visit a doctor. Today, it would be very cumbersome and not practical to share medical history and medications with a doctor, other than through a simple verbal explanation. However, with a healthy decentralized identity ecosystem of issuers and verifiers, it would be possible to share important medical information in a digital privacy-preserving manner, thus enabling the doctor to make a better medical decision and provide the patient with a much better service. An additional example is a mortgage lender that may need the homeowner to provide proof of active property insurance. To that end, the homeowner can present the property insurance information to the mortgage lender and the lender can periodically verify the current status of the insurance policy on its own without unnecessarily burdening the homeowner with having to constantly present this documentation to the lender for verification on a recurring schedule. While centralized or federated identity might also support these use cases, decentralized identity might be better suited for them.

Decentralized identity may enable new business models and value exchange. It may pave the path for fully digital-only experiences that remove the requirement for individuals to present themselves in-person to perform high value transactions. Decentralized identity may also enable a better in-person user experience in a variety of situations without requiring a person to carry a physical wallet at all. There are also potential benefits for businesses to streamline how they might verify and build trust with their customers. There is definite potential here, but only time and the market will tell if the great expectations for decentralized identity will be fully realized in practice over the long term.

Decentralized Identity Terminology

The following are the primary components involved in a decentralized identity experience. These definitions have been simplified to make it easier to understand the actors and how they interact:

- *Self-sovereign identity* is a term that describes a digital movement that is founded on the principle that an individual should own and control their identity without the intervening administrative authorities.
- *Verifiable credentials* are attestations that an issuer makes about a subject. Verifiable credentials are digitally signed by the issuer.
- *Issuer* is the entity that issues verifiable credentials about subjects to holders. Issuers are typically a government entity or corporation, but an issuer can also be a person or device.
- *Holder* is the entity that holds verifiable credentials. Holders are typically users but can also be organizations or devices.
- *Verifier* is the entity that verifies verifiable credentials so that it can provide services to a holder.
- *Verifiable presentations* are the packaging of verifiable credentials, self-issued attestations, or other such artifacts that are then presented to verifiers for verification. Verifiable presentations are digitally signed by the holder and can encapsulate all the information that a verifier is requesting in a single package. This is also the place where holders can describe the specific terms of use under which the presentation is performed.
- *User agent* or *digital agent* is the software application that holders use (typically a mobile device app) that receives verifiable credentials from issuers, stores them, and presents verifiable credentials to verifiers for verification.
- *Identity hub* or *repository* is the place where users can store their encrypted identity-related information. An identity hub can be anywhere – on the edge, on the cloud, or on your own server. Its purpose is to store personal data. Some implementations may allow other entities to access the identity hub of the user if the user specifically grants such access. You can think of an identity hub as the individual’s personal data store.

- *Decentralized Identifier (DID)* is an identifier that is created and anchored in a decentralized system such as a blockchain or ledger and can represent any entity in the ecosystem – an issuer, a holder, a verifier, and even an identity hub.
- *Digital cards* represent verifiable credentials that users collect over time and are stored as part of the user agent or the identity hub of the user. It’s somewhat simpler to refer to them as digital cards rather than verifiable credentials when speaking about them.
- *Digital wallet* represents a digital metaphor for a physical wallet and is generally represented by the combination of the user agent and the underlying capabilities of the computing device, such as secure storage and secure enclaves on a mobile phone. The digital wallet contains digital cards.
- *dPKI* is a decentralized public key infrastructure and is usually implemented via an immutable blockchain or ledger – a place where DIDs can be registered and looked up alongside the associated public keys of the DID and its metadata. dPKI can be described more generally as the *verifiable data registry*, as the dPKI is just one of many possible implementations for a verifiable data registry. While this paper refers to dPKI, the reader should be aware that a verifiable data registry need not necessarily be “decentralized”.
- *Universal resolver* is an identifier resolver that works with any decentralized identifier system through DID drivers. The purpose of a universal resolver is to return a DID document containing DID metadata when given a specific DID value. This capability is very useful because DIDs can be anchored on any number of disparate dPKI implementations.

The figure below highlights some of the terminology just outlined with the major actors and their relationships. It also represents the sample scenario we will cover later in this document.

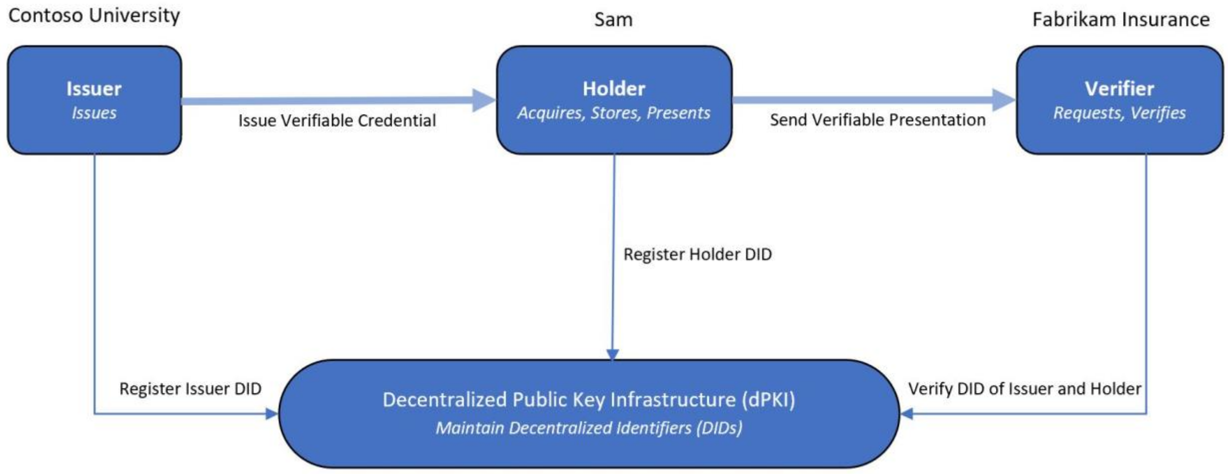

Contoso University
Sam
Fabrikam Insurance
Issuer
Issues
Issue Verifiable Credential
Holder
Acquires, Stores, Presents
Send Verifiable Presentation
Verifier
Requests, Verifies
Register Holder DID
Register Issuer DID
Decentralized Public Key Infrastructure (dPKI)
Maintain Decentralized Identifiers (DIDs)
Verify DID of Issuer and Holder

Figure 1 - Verifiable Credential Issuance and Presentation

It is essential to note that no personally identifiable information should be stored on the decentralized public key infrastructure. Personal identity data is stored as part of the individual’s digital wallet or identity hub in a secure location.

Usually, the holder will present verifiable credentials to verifiers during a business transaction in real-time, like the way we currently present our passport at a border crossing. However, in more advanced scenarios, some implementations may enable the holder to grant a verifier-specific access to data in the holder’s identity hub. That way, the verifier can access data that the individual has allowed access to, instead of the individual having to manually present verifiable credentials to the verifier on a recurring schedule. Nevertheless, the more traditional approach still requires the holder to present verifiable credentials to the verifier explicitly, but the verifier will have the ability to periodically check the status of the credential, such as whether or not the credential has been revoked by the issuer, on its own without burden to the holder.

Now that you are armed with an understanding of the terminology, let’s take a closer look at a sample scenario.

Decentralized Identity Scenario

The example below is meant to provide an end-to-end use-case of the value and utility of a decentralized identity ecosystem. It is not a comprehensive or exhaustive description of all that is possible with decentralized identities as it represents just one possible decentralized identity flow.

Suppose Sam wants to purchase vehicle insurance from Example Insurance, but to get a good rate, Example Insurance requires proof that Sam is a graduate of ABC University. In our decentralized identity scenario, the actors are as follows:

- Sam as the verifiable credential subject and holder.
- ABC University as the verifiable credential issuer.
- Example Insurance as the verifiable credential verifier.

The following sequence of steps represents a flow where the end-goal is for Sam to receive a digital diploma from ABC University and then present it for verification to Example Insurance in order to claim the automobile insurance discount:

1. Sam receives an email from ABC University congratulating Sam on graduating while also providing a QR code Sam can use to scan with Sam’s mobile phone. Sam has an app on Sam’s phone that is registered to handle such a request. This app represents Sam's *digital wallet* that will hold all the *digital cards* that were collected over time. Sam scans the QR code, the digital wallet app launches, and Sam is informed that in order to receive Sam’s digital diploma Sam needs to sign-in to the ABC University website.

2. In our case, Sam presses on the link and enters Sam's existing credentials to authenticate on the University's website or if Sam didn't have such a credential, Sam may be asked to come in person to the registrar's office to do ID proofing and receive their credentials. Once Sam provides their existing credentials, Sam is informed that Sam can go ahead and *accept* this digital card from ABC University. Once Sam accepts the card, Sam is asked to secure this operation with a biometric, such as a fingerprint, face, or even a PIN. After Sam performs this action, the card is now securely stored in Sam's digital wallet. Sam can inspect the card, view the data that the card has about Sam (which was attested to by the university), such as Sam's full name, major, graduation date, and issue date. Also, Sam can view the activity that this card was involved in, such as when it was issued, to whom it was presented, and how it was used - all of this can be done from the digital wallet app on Sam's phone. Each such activity can be considered as a *digital receipt* or *verifiable history* that Sam can use to track who has (or had) access to the data for this card. These digital receipts are stored locally along with the card in Sam's digital wallet, which is always under Sam's control. More generally, we can also refer to this digital card as a *verifiable credential*.

3. Now, to claim Sam's discount, Sam navigates to the Example Insurance website on Sam's mobile phone and notices the *Verify Credentials* button. This is a deep link and when Sam presses it, the digital wallet app opens with a permission request. The permission request indicates that Example Insurance needs to receive a ABC University alumni digital card for Sam to get Sam's discount. Note that Sam doesn't have to authenticate to Example Insurance with a username and password nor use a federated IdP. Sam can simply present the digital diploma Sam already possesses in Sam's digital wallet. In our scenario, Sam only presents Sam's ABC University alumni digital card to Example Insurance, but Sam could also present other digital cards Sam has in Sam's digital wallet such as a digital card that proves Sam is a resident of a specific territory or to prove Sam's current address. Once Sam authorizes the permission request with Sam's biometric such as a fingerprint scan, Example Insurance now receives the digital card and is able to verify that it was indeed issued to Sam by ABC University, and it is indeed Sam who is presenting this digital card to Example. Once Example Insurance completes the verification, it can now offer a discount to Sam! Sam can now view that Sam's digital wallet app has a receipt for this card, indicating that this card was presented to Example Insurance on a given date and for a specified purpose with Example's terms and conditions. Some implementations may further enable Sam to *revoke* the access Example Insurance has to view Sam's digital card. This revocation action may generate another *receipt* that clearly indicates the date and time Sam revoked Example's access to Sam's digital card. Once again, Sam can accomplish all this from Sam's digital wallet app on Sam's mobile phone, and all the digital cards that Sam collects over time and Sam's associated receipts are under Sam's control.

4. Sam can collect many such digital cards in Sam's digital wallet and at some point may even need to present multiple cards, such as in the case if Sam wants to attend an advanced enterprise architecture training academy, both proving Sam is a ABC

University alumni as well as a certified enterprise architect. The academy can then instantly verify both credentials presented and enable Sam to access Sam's advanced training material.

It is important to clarify that Sam sends a *verifiable presentation* to Example Insurance. The verifiable presentation contains a nested artifact which is the *verifiable credential* Sam has received from ABC University. In this manner, Example Insurance that is acting as the verifier, can verify the following two critical elements:

- Based on the digital signature of the *verifiable credential*, Example Insurance verifies that the verifiable credential is authentic and was indeed issued by ABC University to Sam
- Based on the digital signature of the *verifiable presentation*, Example Insurance verifies that it is indeed Sam who is performing this credential presentation

After Example insurance has verified the above, it is able to confidently present Sam with Sam's vehicle insurance discount.

Decentralized Identity Technical Implementation

The following sequence is a technical explanation of the same scenario presented above. It outlines the steps that must be taken to setup the decentralized identity experience as well as the verifiable credential issuance and presentation flows. However, this scenario assumes that the decentralized public key infrastructure (dPKI) has already been setup and will not be detailed here.

Setup

1. ABC University represents the issuer. A generates a decentralized identifier (DID) tied to a public/private key pair and registers their DID on the dPKI. The private key is stored by the ABC University IT team in a Key Vault or Hardware Security Module. The corresponding public key is published to a decentralized ledger such as a blockchain so that anyone can find it.

2. ABC University IT publishes a DID document that associates its DID to the registered public Domain Name System (DNS) domain, such as A.edu. This represents a domain linkage verifiable credential. ABC University IT can host this file on their website which both proves ownership of the domain and the specific DID. The verifier (such as Example Insurance) can use this DID document to confirm the DID ownership for ABC University and ensure that the verifiable credential it receives is indeed issued by ABC University and not by some other issuer claiming to be ABC University.

3. ABC University IT develop a contract that describes the requirements for the issuance of the verifiable credential. For example, ABC University IT can specify which attestations should be self-issued directly by the user, and which other verifiable credentials, if any, the individual must first provide. In our scenario, the IT team has mandated that the

student authenticate with a federated IdP that supports the OpenID Connect protocol, so that it will be able to receive a security token and extract claims from it, such as first name, last name, and student number. The issuer will then be able to map it to attributes it will issue in the verifiable credential. Importantly, ABC University will indicate the schema(s) to which the verifiable credential will conform, so that other verifiers around the world will be able to consume the content of the verifiable credential those verifiers receive.

4. Finally, ABC University IT administrators can setup and customize the branding of the soon-to-be-issued verifiable credential cards such as card color, logos, icons, images, and helpful text. The administrators can customize the helpful text strings via metadata that will appear as part of the cards based on the attestations issued with the card for credential data. This will help design the look and feel of verifiable credential alumni cards issued by ABC University, and ensure the issued digital cards reflect the brand of the university. In the future, these graphical elements should be standardized so that students enjoy a consistent digital card visual rendering experience regardless of which vendor develops the user agent or digital agent the student chooses to use.

Verifiable Credential Issuance

1. The credential issuance request flow begins when Sam scans a QR code using Sam’s mobile phone. The purpose of the issuance request is for Sam’s user agent to retrieve the requirements for credential issuance as dictated by the issuer and to display the appropriate UX to the user via the user agent. As such, the QR code is displayed on the ABC University website and scanning the QR code opens Sam's digital wallet mobile app and triggers an issuance request retrieval operation from the user agent to ABC University. Once the user agent receives the issuance request from ABC University, it begins the flow to issue the credential. The issuance request is digitally signed by ABC University and the user agent can verify the authenticity of such a request. The issuance request includes a reference to the contract that describes how the user agent should render the UX and what information Sam needs to provide in order to be given a verifiable alumni credential.

2. After the user agent verifies that the request is genuine, it renders the UX to Sam. Because of the specific requirement that A has for issuing digital alumni cards in our scenario, Sam needs to sign in with Sam’s existing ABC University account, which, in turn, will issue a security token to the user agent with claims such as Sam's first name and last name, degree, and graduation date. (Note that during setup above, the issuer can be configured to accept security tokens from any trusted and compliant OpenID Connect identity provider and the user agent will use this identity provider during the issuance process.) Therefore, when the individual presses ‘Login to ABC University’ on the user agent, the user agent can redirect the individual to authenticate with the IdP, and it is there the individual can perform standard authentication tasks such as entering their username and password, performing Multi-Factor Authentication (MFA), accepting

terms of service, or even paying for their credential. All this activity occurs on the client side via the user agent (e.g., a mobile app). When the user agent finally receives the security token from the IdP, it can pass it along to the issuer which can then extract claims from it, as mentioned above, and inject these as attributes into the resulting verifiable credential, potentially enriching the claims with information obtained from other sources. As well, after the individual authenticates with the IdP, the user agent can display additional input fields that the individual is free to self-select. After the individual has provided all the required information, the user agent can verify that it has all the necessary issuer requirements fulfilled, and it can go ahead and ask if Sam would like to accept the card.

3. In our scenario, when Sam accepts the card, Sam is asked to use a biometric gesture such as a fingerprint scan. This action generates a private/public key pair for Sam’s DID whereby the private key is stored on the mobile phone in a secure enclave, and the public key is published to a distributed ledger.

4. Finally, the issuer receives all the required information alongside Sam’s DID and issues the digital card to Sam who then receives the verifiable credential, which is a JSON Web Token (JWT) following the W3C standard for verifiable credentials. The JWT includes both the DID of the subject, Sam, and the DID of the issuer, ABC University, as well as the type of the credential, and any attestations such as first name, last name, major, and graduation date. It also contains a way to find out the credential's revocation status in case the credential is later revoked by the issuer - ABC University. This verifiable credential is digitally signed by the issuer's DID.

5. Once the user agent validates the verifiable credential received from ABC University, it inserts this digital card into Sam's digital wallet as a card Sam can now present to other organizations such as Example Insurance.

Verifiable Credential Presentation

1. When Sam visits the Example Insurance website on their mobile phone to receive a discount on their vehicle insurance, Sam presses the ‘Verify Credentials’ button on the Example website (which is a deep link) or simply scans a QR code generated by Example via their mobile phone. This generates a presentation/verification request for Sam to verify Sam’s ABC University alumni status. The request describes the type of card(s) that Sam should present to Example Insurance, such as Sam’s digital alumni card from ABC University, and this request is digitally signed by the verifier's DID, which in our case, is Example Insurance. The presentation request can also include Example's terms of service.

2. After the signature of the request is verified by the user agent, Sam is presented with a UI on the user agent indicating that Example Insurance is requesting permission to see

Sam’s ABC University alumni card with a reason as to why Example needs to see it (such as for Sam to be able to receive their discount).

3. After Sam approves the request with a biometric gesture, such as with a fingerprint scan on the mobile phone, the verification response, which is essentially a presentation of a credential response (also known as a verifiable presentation), is sent to Example Insurance. The response is signed by Sam's private key and includes the verifiable credential issued by ABC University to Sam nested inside the JWT payload.

4. Example Insurance attempts to match the person performing the presentation of the credential with the subject of the nested verifiable credential to ensure that it is indeed Sam who is presenting it to Example Insurance, and not anybody else. Therefore, the DID of Sam is present in both the outer JWT payload since Sam is performing the presentation of the credential, as well as inside the nested JWT payload as the subject of the verifiable credential issued by ABC University. Once Example Insurance confirms that the DID in the presentation matches the subject of the issued credential, Sam is both authenticated to the Example Insurance website and authorized to claim Sam’s discount! This is much better than simply possessing a username and password, since, in this mechanism, Example Insurance knows that the person presenting this credential is the same person to whom the card was issued. With a username and password, someone else can use it to impersonate you. In this architecture, however, this is significantly harder to do. Someone else will need to take control of Sam's private key stored on Sam’s phone's secure enclave to be able to accomplish this malevolent task.

5. At last, Example Insurance can extract the data it requires from the verifiable credential such as Sam's first name, last name, major, graduation date, and go ahead and present Sam with Sam’s vehicle insurance discount!

6. The credential verification flow completes when Sam stores a signed receipt by Example Insurance that will be associated with the card in Sam’s wallet. Sam now has a single place where Sam can view all the websites where Sam has presented Sam’s alumni card over time. In our scenario, the receipt includes information about Example Insurance, the reason Example needed to receive the card, Example's terms and conditions, and the date the receipt was generated. This signed receipt is associated with the card in Sam's digital wallet and will always be under Sam's possession.

7. Some implementations may further enable Sam to go ahead and decide to revoke Example's access to Sam’s ABC University digital alumni card. Example should thus implement the necessary revocation measures to ensure it complies with Sam's request. The verifier should then cease to use the data from the card Sam presented to it. Sam can later prove that Sam issued a revocation request if such a need arises, and this can help with General Data Protection Regulation (GDPR) compliance.

Scenario Summary

In our simple use-case above, the issuer of a verifiable credential was ABC University, but in other contexts, the issuer can be an employer, a government agency, a device, a daemon process, or even the individual. Likewise, a verifier can also be any of the previously mentioned actors. The decentralized identity ecosystem is very broad and the standards allow for opportunities to unlock a more flexible, secure, and privacy-preserving way to perform digital interactions in a myriad of contexts.

The components presented in the flow above are based on open standards. The verifiable credentials issuance and presentation flows depend on the foundational specification of the *W3C Verifiable Credentials Standard*, and the decentralized system, such as blockchains and ledgers, are based on *W3C Decentralized Identifiers* work. The purpose of the decentralized ledger technology is to support a decentralized public key infrastructure (dPKI). The dPKI anchors DIDs and their public keys and thus enables ownership of DIDs to be validated without relying on only a few privileged identity providers or certification authorities.

The Decentralized Identity Foundation is leading the effort on decentralized identity, but more work remains to fully define the space.[iii] For example, the decentralized identity community is discussing how to enable better privacy preservation by empowering Sam to present Sam's age in a privacy-preserving way without unnecessarily disclosing Sam's exact date of birth to the verifier. Another area under discussion is how to empower Sam with performing self-owned key recovery in case Sam loses or damages Sam's phone, so that Sam can more easily retrieve all Sam's previously acquired digital cards back onto a different device or onto a different user agent in a more seamless manner.

Decentralized Identity Limitations

While decentralized identity has the potential to improve an individual's productivity and digitize existing business processes for governments and corporations, it does have known limitations and areas where further research or investigation would be required. A decentralized identity ecosystem can only be successful when it achieves critical mass adoption by governments, businesses, and individuals. When Apple released the first iPhone, it ushered in a new and immediate change in the user experience the moment the purchaser took possession of their new device. In contrast, an individual may not gain much benefit in obtaining a verifiable credential from an issuer unless they can then use that verifiable credential with many verifiers. A digital passport, for example, is only useful to a citizen if it can be used at most airports and border crossings around the world. Organizations may hesitate to be issuers or verifiers of verifiable credentials unless there is already a healthy ecosystem in place, but that ecosystem cannot develop unless there are entities willing to issue and verify these new credentials.

Decentralized identity is a digital identity. Without the necessary technology to hold a digital wallet, such as on a mobile phone or some sort of computing device, it will be very difficult for the promise of digital identity to be realized by all individuals around the world. If an individual loses their device or decides to share their device with others without proper precautions, it can become a challenge to recover their data onto a different device or to prove who performed a specific interaction. Asking the average person to understand this and to safeguard their private key material remains a significant challenge to decentralized key management.

In most decentralized identity use-cases, the developers assume all parties involved have access to the Internet. That may not be the case. Other scenarios that take the individual away from Internet access leave open the question of how verifiable credentials can be verified in such scenarios. Verifying verifiable credentials requires looking up information on the dPKI, or at the very least, checking if a credential that is being presented has been revoked, and that requires network connectivity. In purely disconnected offline environments this poses a challenge, and a potential hurdle to decentralized identity adoption in specific contexts and situations.

The promise of decentralized identity is to empower individuals to own and control their digital identity and personal data. However, if a person provides a verifiable credential containing personal data to the service provider, the service provider is able to copy this data to its own databases for marketing purposes or to be able to continue providing services to the user. The individual can attempt to revoke access that the service provider has to the verifiable credential but there is no guarantee that the service provider will honor such a request and delete all the data it has stored about the user. This would be a very challenging problem to solve via strictly technological measures and would most likely require legal and policy frameworks in place to ensure everyone's personal data is protected, to ensure audit records are kept, and to establish a documented process for dispute management and resolution.

Final Words

Decentralized identity can enable entirely new business opportunities and empower citizens to be more in control of their identity and personal data. Today, IT administrators need to perform cryptographic key exchange ceremonies to establish trust between two organizational entities. This does not scale when doing business with dozens or perhaps hundreds of other vendors in a more ad-hoc manner. Today, when a bank issues a credit card to a customer, that customer can use that credit card to make purchases with almost any merchant worldwide. In such a scenario, it is not feasible to expect every merchant to exchange cryptographic keys a-priori with every possible bank that issues credit cards. A decentralized identity ecosystem can enable a similar concept to credit card associations by introducing governance authorities and frameworks for many different trust communities in a wide array of industry verticals. As a result, merchants, or other verifiers, can avoid setting up multiple trust federations – they can simply ask the issuer to present additional proofs proving that the issuer is indeed a member of a specific governance authority with which the verifier already has an established trust relationship.

One of the major hurdles for adopting blockchain today in enterprise scenarios is the lack of a decentralized identity infrastructure. After all, it’s not very logical to have a decentralized blockchain network if all the identities on it are still relying on centrally controlled accounts. Furthermore, in a decentralized identity ecosystem, consumers will be more easily able to track which websites they visit online and with whom they transact. You will know which businesses have your personal data, and you will be able to revoke access to it should you so desire. Instead of sharing paper documents or physical cards, you will be able to share digital documents and digital cards in a fully digital, privacy-preserving, and auditable manner. For organizations, this may reduce GDPR-related risk since personal data will be stored in the identity hub under the individual’s control, while the organization will only have access to specific data as granted by the user. Furthermore, the individual may have the opportunity to revoke access to their data, and this may simplify the GDPR compliance for an organization as well as streamline such requests for the individual. As well, GDPR compliance may be eased for an organization as it will be able to possess cryptographic proof as evidence that the individual has indeed provided them with specific data.

As discussed, the digital wallet contains a digital agent app with which the user interacts. Such digital or user agents are mostly based on open source software. The individual can download a user agent from a commercial corporation, or perhaps even a government entity. An individual may even develop their own user agent from existing open source software. Conceptually, an individual must trust the user agent and it should be under the individual’s control.
While it is extremely challenging to attempt to predict how the decentralized identity landscape will evolve given its nascent state, current trends are indicating government interest to ease the burden on citizens and businesses via government-issued digital IDs. Tailwinds from the unprecedented global COVID-19 pandemic are urging government institutions to streamline citizen and business access to government-provided services. As well, increasingly stringent regulatory compliance requirements and further demand by users for better user experience and increased convenience may further drive demand for digital identity in the form of verifiable credential exchange. Finally, verifiable credentials may prove very useful in situations where the same credential must be presented both online in digital transactions as well as in offline in-person interactions, since this can result in increased business efficiencies for the enterprise and a more consistent and simplified user experience.

Conclusion

Decentralized identity is a conceptual shift from the way the identity and access management community has been approaching identity in the past, yet it is able to co-exist with the account-based identity model that has existed for decades. Decentralized identity can add a lot of value to transactions that require high assurance of trust to make authorization decisions. If an individual continues to authenticate with a website using a traditional “account”, it does not preclude the individual from having to present verifiable credentials in order to, say, transfer large sums of money to another individual or organization. This offers the possibility to unlock a myriad of new opportunities for digital commerce and enable consumers, employees, and citizens around the world to transact on the web in a more secure, safe, and privacy-preserving manner. It may pave the path for a digital wallet with digital cards, like the way we all use a physical wallet and physical cards today. Verifiable credentials are easy to reason over because many of them will simply be digital representations of the physical cards we already carry in our wallets every day.

We are still at the early days of decentralized identity. It is not a technology that a single company can simply release to the market. It requires both standards as well as collaboration between the private and public sector to have a healthy ecosystem of *issuers*, *holders*, and *verifiers*. When we finally reach critical mass adoption, digital experiences may look and feel much different from the experiences of today. Decentralized identity is an exciting development in the identity space, and it has the potential to offer more trustworthy digital experiences and unlock more value for everyone.

Change Log

Date	Change
2020-10-30	V1 published
2022-02-28, 2023-03-31	V2 Editorial changes only (changed example business names to non-Microsoft specific ones; changed A University to ABC University)

Author Bio

Leo Sorokin has over 10+ years of experience in various solution architecture and enterprise architecture roles with large organizations in the financial, manufacturing, and software industries. He is currently a Cloud Solutions Architect at Microsoft helping the largest Canadian organizations adopt cloud technology. Leo has extensive experience with identity, service-oriented architecture, application integration, cloud-native application and hybrid-cloud architecture, as well as security software architecture. Leo is also TOGAF® 9 Certified, a Microsoft Certified Azure Solutions Architect and holds a Computer Science degree from York University. Leo has also taught technology related courses in several educational institutions.

[i] "Verifiable Credentials Data Model 1.0," W3C Recommendation, World Wide Web Consortium (W3C), 19 November 2019, https://www.w3.org/TR/vc-data-model/.

[ii] “Decentralized Identifiers (DIDs) v1.0,” W3C Working Draft, World Wide Web Consortium (W3C), 27 October 2020, https://www.w3.org/TR/did-core/.

[iii] Decentralized Identity Foundation (DIF), [Online]. Available: https://identity.foundation/.

Non-Human Entities

Non-Human Account Management (v4)

By Graham Williamson, André Koot, Gloria Lee

To comment on this article, please visit our GitHub repository and submit an issue .

Table of Contents

Abstract

Non-human accounts are often the "Achilles' heel" of a robust IAM environment. While IAM professionals concern themselves with managing identities, authentication, RBAC, ABAC, governance, and auditing of user accounts, other IT staff are deploying devices and services that are given access to protected resources via hard-wired accounts, exposed services, and APIs.

The management of non-human account control should be consistent with user-based account management, and controls placed on user account access to high-assurance applications should also be applied to non-human accounts.

There is no single solution for dealing with non-human accounts. Some IAM professionals suggest all accounts should be managed via the same processes and same infrastructure to ensure consistent policy deployment. This consistency, they argue, should ensure that non-human accounts are not 'left-out' when IAM deployments occur. Others consider this impractical and recommend that purpose-specific processes be deployed for non-human accounts. But regardless of the mechanism(s) used to manage non-human accounts, ensuring that they are managed is paramount. Otherwise, non-human accounts will continue to be a cybersecurity attack vector favored by hackers for gaining access to corporate facilities.

Introduction

A non-human account is usually associated with a service or device rather than a human user. An example is a machine-to-machine service, such as a backup routine that runs during non-business hours to create an offline copy of production data. In this instance, the account permissions should be restricted, i.e., they should not have standard user access nor general Administrator privileges.

Devices such as sensors that provide data to be monitored are sometimes deployed with access to an account so that they can write to a database. Again, such an account should have limited privileges.

Fortunately, the use of such accounts is diminishing as the use of APIs becomes more sophisticated, providing better security and eliminating the practice of hardcoding usernames and passwords in connection routines.

While IAM professionals typically focus on user accounts, these non-human accounts represent a potential attack vector for organizations. These accounts should be considered when formulating policies for access to computer systems.

A comparison between the characteristics of these accounts is shown below:

	Person Identity	*Non-human Identity*
Usage	Multi-faceted, must accommodate multiple access requirements to many applications or protected resources	Purpose-specific, with a single requirement for each deployment
Lifecycle	Created during the 'joiner' process, modified when 'moves' occur, continually monitored for compliance, disabled, and then deleted according to the 'leaver' process.[i]	Created on deployment of the device/service, deleted on termination.
Access control	Dynamic – continual risk-assessed authentication matched to the assurance level requirements of the requested application or protected resource. MFA is used for authentication elevation.	Static – determined at the time of account creation. No MFA requirement.
Access endpoints	Users typically access computer services from smartphones, PCs, and laptops on an interactive basis.	Endpoints are typically devices or device controllers. They can also be computer applications, service routines, or Internet bots.

Table 1 - Account type characteristics

There are two broad categories of non-human accounts that IAM practitioners should differentiate:

- Machine-to-machine accounts used by devices or services to perform a specific function; these 'server' accounts should be monitored and alarm on any incident that is an anomaly to the expected operation.
- Accounts that have access to system functions but are not assigned to a specific individual; these 'system' accounts include administrator accounts with elevated privileges.

Terminology

- Bot – sometimes called an Internet bot, short for 'robot' but referring to a software routine that performs automated tasks over the Internet, a web robot referring to an autonomous network application, or simply a 'bot' referring to an automated, typically repetitive, task used for a specific purpose.
- Identity – defining attributes for a human user that may vary across domains, e.g., a user's digital identity will have a different definition in a work environment as opposed to the user's bank. A device identifier is sometimes referred to as its identity.
- CIA Triad - the fundamental Information security concepts of risk classification of resources from the perspectives of Confidentiality, Integrity, and Availability.
- Non-human/person account – any account not used by a person, including accounts used for devices, services, and servers.

- Server account – an account established with access rights to a specific server operation; this includes service accounts used by a computer application to access another application or service or an account used for a device connection. Note: these accounts are username accounts typically secure via a password.
- System account – a generic term for a privileged account that has extensive permissions that enable system configuration changes.

Non-human Access Control

A significant concern for the IAM practitioner is how to manage access control to and from devices, particularly with services not used interactively by humans. This includes bots that are increasingly being used for automated processes.

IoT Devices

IoT devices can be either a sensor or an actuator. In some cases, sensors provide a continuous stream of data that is displayed in real-time or discrete readings that are written to a database for periodic analysis. Actuators are devices typically used to control a process, turning something on or off. They may be used to open or close a valve by pulsing a servo motor a sufficient number of times until the desired aperture is reached. In many cases, devices are remotely located and connected via a controller to the supervisory system located in a central location. It is noted that IoT devices are becoming increasingly sophisticated with control capabilities and communication facilities built-in. This eliminates the need for a username/password account as IoT devices typically communicate to an API with encryption and digital signing functionality.

In a typical IoT configuration, there are three zones:

1. IoT devices (sensors & actuators). Managing access to and from devices should be governed by a policy that imposes requirements for encryption of the communications channel, such as DNP3, MQTT, and/or digital signature technology (e.g., PKI), to suit the required security level. In low-security environments, static passwords might be used that remain in service until the equipment is decommissioned. In higher-sensitive applications, the security credentials (passwords, certificates, etc.) will be periodically rotated. The selected security requirement must match the capability of the devices, but technical limitations often constrain IoT devices. “Terminology for Constrained-Node Networks" (RFC 7228) nominates three classes of devices:[ii]
    a. Class 0 – no capacity to support configurable authentication.
    b. Class 1 – limited capacity for key management, token support, etc.
    c. Class 2 – fully configurable and able to support dynamic authentication mechanisms.
2. The Controller (to which the devices are connected). If sensor device data is aggregated by a device controller that maps each sensor or actuator to its control logic, providing access control to actuators and protection on writing collected data to a database is required (see Service Accounts, below).

3. Human-Machine interface application (HMI) such as a controller app or a SCADA app monitoring or controlling the IoT devices. In some cases, sensors will write data directly to a database that is read by another application, such as a SCADA app or similar human-machine interface (HMI). Access to these applications will be by humans and should be managed via the IDM environment.

Historically IoT environments have been managed by a team responsible for operational technology (OT) and have had little to do with the information technology (IT) environment within an organization. The specialist nature of IoT technology has justified this organizational structure, and it is often corporate policy to isolate OT from potential compromise via the IT environment. But the requirement for isolation is diminishing as security technology improves. Integrating IoT systems with the IAM environment will improve access control capabilities and provide better corporate governance over operational technology deployments for most industrial applications.

If allowed by regulatory controls, best practice is to integrate the OT environment with the IT IAM environment. This enables the OT to set system entitlements via the IAM system and for OT staff to use their corporate credentials for authorization, potentially via a Privileged Access Management system.

There is increasing concern regarding the provenance of IoT devices and tracking devices throughout the supply chain to ensure no modifications have been made that could potentially deploy 'back-door' access.[iii] The IAM practitioner may wish to ensure corporate policy defines the certification processes to be employed for IoT devices and ensure that compliance with software supply chain policy is in place. This is increasingly important in regulated industries.

Just as important as securing the device itself is protecting the IoT device data. In many cases, databases with IoT devices are not adequately secured. A risk management approach should be employed to determine the adequacy of protection; building environment device data might be low risk but plant production data that is not adequately protected from industrial espionage might be considered critical. The IAM professional should ensure appropriate access controls are placed on industrial data stores. It is good practice to assign a data controller role to an industrial database.

Vulnerability Mitigation

There is no 'correct answer' when it comes to deciding the involvement of IAM practitioners in the management of IoT devices. At one end of the spectrum is the use case whereby all IoT deployments and management are the domain of OT personnel. In this case, the IAM involvement will be restricted to the human accounts that access the OT systems. Group management of entitlements to accounts that can configure IoT systems will heighten the level of security.

At the midpoint of the spectrum, components of the IoT configuration and operation will fall under IAM services. The IAM provisioning workflow will route configuration requests and potentially password rotation requests, to the responsible person. The IoT devices will participate in both attestation reporting to the responsible manager and compliance management with integration to the Security Operations Center (SOC) and possibly the Security Information and Event Management (SIEM) system.

At the other end of the spectrum, the provisioning of devices is included in the identity management infrastructure. IoT devices are treated the same way as individuals, applying a 'digital identity' to devices. Their entitlements can be set via the normal account provisioning workflows, and their access control can use the same protocols. Most modern API systems, including gateways, use OAuth 2.0 for machine-to-machine communications, while Open ID Connect can be appropriate for IoT device controller authentication.[iv]

Service Accounts

There is a wide variety of service accounts. They are typically used in processes that are periodically run on an automated basis, e.g., via a UNIX cron job or Windows Task Scheduler. Auditors often overlook the accounts used by these processes because they are not accessed by users interactively. Since users do not log into them, they are typically basic, single-purpose accounts with restricted privileges.

Examples include:

- An account used to perform a nightly backup of data
- An account providing access to the HVAC system for monitoring purposes
- An account used for replication of data between directory instances.

The term 'batch account' is sometimes used for a service account. These often refer to one or more utility operations that run periodically during non-production hours to perform a system function. Multiple batch processes may use a single batch account.

Vulnerability Mitigation

Service accounts are a significant source of concern for many organizations because they are often established with a static password that, if not encrypted, can be read by any system administrator. If their access rights are not tightly scoped, these accounts can then be used interactively by a malicious actor and possibly used for lateral movement to other servers in the organization's network. If corporate data loss protection extends to service accounts, tools such as authentication monitoring for anomalies can guard against such vulnerabilities. User behavior analysis tools baseline the normal activity on an account; any deviation from this will generate an alert to the event monitoring system. Alternatively, static service accounts can be migrated to APIs that typically impose a strict security and monitoring regime.

Note: the term 'service account' is sometimes used to describe an account accessed periodically by a service person, e.g., an HVAC technician. Such accounts are user accounts and should be addressed in a company's IAM strategy. They are not addressed in this document.

Bots

The term 'bot' has come from the Robotic Process Automation (RPA) sector that had its genesis in plant automation, where software routines are deployed for repetitive processes.[v] Bots are now used for everything from website crawlers to retrieve usage information to denial-of-service malware. Increasingly they are being used by organizations to automate repetitive tasks such as retrieval of building information management data or consolidating customer transaction data. In these cases, access by bots will be restricted to a specific purpose.

Bots typically use the Internet to access remote services or resources. A publicly available website should apply mechanisms to limit bot activity and avoid malicious access. These mechanisms might include applying screen-scraper controls, human verification checks, and DDOS protection. A common form of malicious activity is 'credential stuffing,' whereby a hacker alters login credentials to take control of a session.

Organizations need to prepare for the external use of bots. Bots will exhibit different characteristics compared to 'normal' non-human access to a process or service. For the IAM practitioner, user behavior analysis can be used to identify access anomalies.
A process for reviewing the use of bots should be established, testing their functionality prior to deployment and analyzing their usage patterns. Monitoring is a continuous task since malicious corruption of bots is a constant concern.

Vulnerability Mitigation
For the corporate application of bot technology, the IAM practitioner's task is to ensure that appropriate controls on credentials are observed and that PKI signatures and encryption are used as appropriate. Only sanctioned activities should be allowed.

For instance, a bot accessing website data will typically authenticate via HTTPS using an assigned session token. It is a good practice to expire session tokens periodically. The length of time a token should be valid should depend on the sensitivity of the service or resources being accessed.

Client Devices

Traditionally identities are people; they have identifiers stored in an identity datastore and then used to authenticate users to protected resources. It is increasingly necessary to also track the endpoint devices that users employ to access corporate resources, such as laptops, tablets, or smartphones. To track those devices, an object is created in the organization's directory or other data stores that record the detail for each device. This data allows us to grant access to a resource based on the device being used to access it.

There are several benefits to registering client devices:

- It can provide a second factor during a human authentication event, thus reducing the risk score associated with the authentication.
- It can be used to customize the presentation and improve the user experience by passing the details of a user's device to an application.
- It can enable unattended device authentication to support scheduled events such as device updates or data retrieval.
- It can remove a vulnerability and improve governance options when client device objects from the data store are disabled or removed when the time period from the LastLogonTimestamp has been exceeded.

Whether your environment is on-premise, hybrid-cloud, or multi-cloud, managing the client device identity lifecycle is key to reducing the organization's attack surface and maintaining compliance per corporate policy.

Vulnerability Mitigation
With the ubiquity of client devices these days, managing client devices can improve an organization's cybersecurity profile. For instance, a smartphone can be a valuable device for multi-factor authentication (MFA). It can provide a 'possession' factor, e.g., the user is using their registered mobile phone. It can also be used to provide a biometric check for an 'inherence' factor.

Some organizations use a Mobile Device Management (MDM) tool to manage client devices. MDM facilitates the tracking and management of devices and will typically include a self-service module to allow users to register and deregister their devices as new devices are acquired or old devices are lost or retired.

Selecting and deploying the appropriate solution for managing client device 'identities' is a core capability in enabling non-human access control.

System Account Access Control

System accounts give humans access to physical or virtual systems or servers and grant entitlements to privileged system functionality. While not strictly non-human accounts, system accounts are included here because they have no single individual to which they are assigned. System accounts typically refer to administration accounts that are established when a system/server is commissioned. Since this type of account is not directly associated with a single person, they are generally not managed via an organization's joiner–mover-leaver HR processes. IAM practitioners must concern themselves with the management of these accounts.

Admin or Root Account

The admin or root account of Windows and Linux or Unix servers is a highly privileged account with access to system-level operations on the respective platform:

- It is authorized at the highest level.
- It has access to every file and process running on a platform.
- It has permissions to configure the system operation and thereby influence the behavior of the platform.
- Logs from a system will typically display commands that have been run and responses that have been viewed
- Operational use of the account should be continuously monitored.

Note: virtualization and hypervisor platforms (VMware, Citrix, Xen) and container platforms (Docker, Openshift, DCOS, Kubernetes) have administrative accounts that provide an attack vector if not properly managed.

Superuser Account

The term Superuser applies to a business information system or application account that has elevated privileges over standard user accounts. It is generated as part of the system commissioning process when the system is deployed. The Superuser account has permission to modify a configuration, making it a mission-critical account in an information system.

Server Account

Accounts for middleware processes like DBMSs, ESBs, or other ICT components that run in the Windows or Linux operating system environments, are sometimes called server accounts. These are privileged accounts in an application such as a DBMS to give administrative access to a resource owner.

Consumer Devices

There is increasing concern regarding the vulnerability of consumer devices that have connections to the Internet. Recent incidents include:

- Privacy violations by devices that have audio or video capture capabilities and that are sending sensitive data back to a monitoring agency.
- Security incidents such as DDoS attacks as common, published administrative passwords are used, giving hackers access to consumer devices that are then used to conduct Distributed Denial of Service (DDoS) attacks.

Most jurisdictions are now requiring products to adhere to an appropriate set of standards that typically include:[vi]

- Ending the use of default passwords. All devices are shipped with a unique password that is not resettable to a common default setting.
- Enabling support for software updates. Devices are shipped with firmware that can be readily updated in the event that a vulnerability is detected.
- Supporting the secure storage of credentials. All credentials should be stored securely with encryption protection and/or a trusted storage mechanism.
- Shipping with a more secure default configuration. Attack surfaces are minimized by closing unused ports, restricting exposed services to only the functionally necessary, and running software with the lowest level of privileges necessary for the system operation,
- Restricting the storage of Personal Identifiable Information (PII). PII is never stored on the device, as per requirements of privacy regulation in the target geographies.

System Account Characteristics

Since system accounts are not assigned to a single identity, they cannot be wholly managed by an IAM solution, e.g., when the person with administrative privileges leaves an organization, it is not appropriate for such an account to be deleted. A common practice is to provide access to privileged accounts via a managed group so that all users in the group are granted access to the account. But management outside the IAM environment is still required. Good practices include:

- Using a configuration management database in which the server/service is registered as an attribute of the identity it belongs to.
- Assign an account owner to be accountable for the use of the account, typically the owner of the system/service that it belongs to. If no system owner has been defined, a responsible person in the IT department should be the accountable party.
- Interactive accounts should only be used for infrastructural changes or calamities. Admin privileges should be granted via a user's account, e.g., via membership in the appropriate Admin group.
- Passwords for Admin/root accounts must be closely managed. They can be secured via a manual procedure, a password vault, or a Privileged Account Management system.

Vulnerability Mitigation
IAM practitioners should assist in the protection of access to all system accounts. In a UNIX environment, this might be via the removal of the 'etc/passwd' file and the use of SUDO for privilege escalation. In a Microsoft Windows environment, a privileged access management (PAM) system is a common solution. In this case, system passwords are made specifically complex and rotated as appropriate. Access to such an account is via a PAM system, which restricts access to specific individuals with the appropriate entitlements and logs all access events.

If a PAM is not used, Windows supports the time-limited elevation of account privileges, with notification to management. Manual intervention that ensures appropriate use and management of system and server accounts is also good practice, as is including server accounts in corporate audits. This level of management will require corporate policy to be established for server accounts which will heighten the visibility of account management practices.

Increasingly, applications are being deployed on cloud services requiring an access control environment that suits each deployment. This type of deployment might mean configuring a resource manager to protect master account privileges or setting policies that ensure applications do not use the master account for database access.

The Future

The ubiquity of IoT devices will become more prevalent. Devices will span both the corporate and the consumer world, and integrating IoT devices and dataflows will be a new corporate risk. Automation will increasingly be deployed with Machine Learning and Artificial Intelligence, adding to the complexity of the access control environment. Integration with the IAM environment via the use of API gateways, database gateways, service meshes, and Policy-Based Access Control solutions should be considered.

Increasingly APIs are being used for machine-to-machine (M2M) communication. APIs provide the ability to apply consistent security controls on a communication channel and also to monitor it for management purposes. Companies adopting a gateway approach have the ability to provide consistency across M2M communications which is virtually impossible if each service instance is deployed individually.

As the adoption of cloud services continues to accelerate, the use of microservices and containerization will become prevalent. The IAM practitioner should ensure that the appropriate information security solutions are put in place to protect communications between services that communicate identity data.

The use of bots will also continue to accelerate; deployment of behavioral analytics and gateway technology should be considered. The US Department of Homeland Security[vii] advises the following:

- Nefarious bot developers will target new IoT devices for vulnerabilities as they are released to the market and will compete with each other to deploy malware.
- Bot code size will get smaller and more sophisticated to avoid detection and frustrate defenses.
- Botnets will be extended and better monetized, likely through interfaces to social media platforms.
- Botnet operators will operate increasingly globally, taking advantage of regional vulnerabilities. Attacks from foreign nation-state operators will increase.

Access control for non-human entities is a critical competence for risk-averse organizations. It is increasingly important to make sure devices and bots adequately identify themselves, move to APIs with consistent security and monitoring controls, and deploy data-loss prevention technologies such as behavioral analysis tools.

Conclusion

All too often, IAM practitioners are sequestered from non-human account management and only focus on the provisioning and access control associated with user accounts. This is unfortunate because it fragments the host organization's risk management approach to cybersecurity and frustrates the governance task. At the very least, the IAM practitioner should ask the appropriate questions as to how IoT devices are being secured, how server accounts are being managed, and what defenses are in place to thwart malicious bots. It is preferable that the IAM and InfoSec teams within an organization work together to ensure the consistent application of cybersecurity controls that are aligned with corporate policy.

Author Bios

Graham Williamson

Graham Williamson is an IAM consultant working with commercial and government organizations for over 20 years with expertise in identity management and access control, enterprise architecture and service-oriented architecture, electronic commerce, and public key infrastructure, as well as ICT strategy development and project management. Graham has undertaken major projects for commercial organizations such as Cathay Pacific in Hong Kong and Sensis in Melbourne, academic institutions in Australia such as Monash University and Griffith University, and government agencies such as Queensland Government CIO's office and the Northern Territory Government in Australia and the Ministry of Home Affairs in Singapore.

Graham holds an electrical engineering degree from the University of Toronto and a Master of Business Administration from Bond University. As a member of the IDPro Body of Knowledge Committee, he looks forward to helping create the definitive body of knowledge for the IAM sector.

André Koot

André Koot is IAM Strategist and Chief Customer Success Officer at Sonic Bee. His IAM experience comes from a financial accounting and auditing background. This background in anti-fraud detection and prevention business processes led to research in the area of authorization principles.

Gloria Lee

Gloria Lee is a Senior Program Manager in the Azure AD Engineering team at Microsoft. As part of the customer experience team for Identity and Network Access, her role is driving customer success in the Azure Identity division. Gloria is focused on helping customers increase security posture with the deployment of Azure Active Directory, Azure hybrid cloud-based solutions to provide identity management.

Prior to joining Microsoft, Gloria was a seasoned engineer/architect with 18+ years of experience in the areas of Identity, security, deployment of Microsoft O365 services as well as messaging and collaboration. She had previously spoken at various events such as Microsoft Identity Driven Airlift Conference for partners, GrayHat 2020, and Texas Security Summit. Outside of technology, she enjoys spending time with her kids/family and travel bargain hunting.

Change Log

Date	Change
2020-10-30	V1 published
2021-04-19	Author affiliation change
2022-02-28	Added a section on client devices; added Gloria Lee as an author
2023-04-10	Various changes to improve the clarity of the article such as the addition of device vs. service information

[i] Cameron, Andrew and Olaf Grewe, “An Overview of the Digital Identity Lifecycle,” IDPro Body of Knowledge, 30 October 2020, https://bok.idpro.org/article/id/31/.
[ii] Bormann, C., Ersue, M., and A. Keranen, "Terminology for Constrained-Node Networks", RFC 7228, DOI 10.17487/RFC7228, May 2014, <https://www.rfc-editor.org/info/rfc7228>.
[iii] Hashemi, Soheil, and Mani Zarei. “Internet of Things Backdoors: Resource Management Issues, Security Challenges, and Detection Methods.” Transactions on Emerging Telecommunications Technologies. Wiley, October 12, 2020. https://doi.org/10.1002/ett.4142.
[iv] See section ‘Mobile & API Innovation Gave Us OAuth & Delegated Authorization Frameworks’ in Dingle, Pamela, “Introduction to Identity - Part 2: Access Management,” IDPro Body of Knowledge, 17 June 2020, https://bok.idpro.org/article/id/45/.
[v] “What is a bot in RPA?,” n.d., https://www.nice.com/guide/rpa/what-is-a-bot-in-rpa/.
[vi] For example, see Fernandez, Angel, "New IoT security regulations: what you need to know,” Allot blog, 30 January 2020, https://www.allot.com/blog/new-iot-security-regulations-what-you-need-to-know/#.
[vii] Botnet Roadmap Status Update, Department of Commerce and Homeland Security, July 2020, https://www.commerce.gov/sites/default/files/2020-07/Botnet%20Road%20Map%20Status%20Update.pdf.

Project Management

Introduction to Project Management for IAM Projects (v3)

By Graham Williamson, Corey Scholefield

Please visit our GitHub repository to submit an issue and comment on this article.

Table of Contents

Abstract

This article serves as an introduction to the practice of project management for an IAM project, describing basic project management terminology and practices. Given the number of systems an IAM project generally impacts, excellent project management is essential for the stakeholders involved.

Importance of Project Management

IAM practitioners may be familiar with the scenario of an IAM project proceeding under the control of an IT systems group without a formal project manager. While this method of deploying a new product or service may be considered an expedient way to get a system installed or updated, it is likely to cost the organization more money in the long term. An IAM service is connected to many critical systems within an organization. Making changes to that service without considering the possible impact on the various connected systems, managing the required resources, or keeping all stakeholders advised of the effort will almost certainly result in a substandard deployment.

Project management has a cost: typically between 5-10% of a project's total expenditure, but it represents the best return compared to any other investment an organization is likely to be afforded.

Terminology

- Project - a time-limited activity to achieve a defined outcome(s)
- Project Charter - documented authority for the project manager to proceed with a project; it will usually include a succinct statement of the project's purpose.
- Schedule -- a document that defines the activity and resources required to achieve the planned deliverable(s) and outcome(s)
- Gantt Chart - a popular schedule format that displays both activity and timeframes in a single chart
- Project Plan - a document that describes a project; it will usually include a scope statement, schedule, resource plan, communications plan, and quality plan
- Task - lowest-level of defined activity; multiple tasks will typically be grouped into stages or project phases
- Agile project management - a framework that uses a continuous, iterative process to deliver a defined piece of functionality, typically a component of a product or service. Scrum is a popular framework.[i]

Readers interested in pursuing information on project management should review the Project Management Institute (PMI) Framework and the PMI Body of Knowledge for further information.[ii]

Characteristics of a Project Manager

In the IT sector, a project manager often has little authority over staff or project stakeholders. They are expected to bring a project in on time and within budget with minimal assistance from upper management and minimal visibility within the organization. In reality, a project manager needs sufficient authority and resources to adequately monitor and manage the project. They also need regular communications with a steering committee consisting of representatives from upper management with the necessary authority to assign resources and remove roadblocks.

Two prime characteristics are essential to a project manager:

Predictability	Management doesn't like surprises. Therefore, a project manager should determine and report on a project's duration and related costs to a defined degree of confidence.
Flexibility	Gone are the days when a project manager slavishly follows an approved Gantt chart to the detriment of anyone who wants a change. These days, IT projects will typically undergo several baseline changes during execution to accommodate scope changes, dependencies on other projects, and changes in resource availability.

Project managers require competence in the five components of project management:

- Planning
- Organizing
- Resourcing
- Directing
- Controlling

PMI Framework

By definition, projects are time-bound; they must have a start and a finish. A major upgrade might be project work if it requires significant resources and coordination of stakeholders. Operational or regular maintenance work is never a project and does not require the skills of a project manager.

Projects are not unexpected; something will instigate the need for a project. Before a project starts, there will be some preparatory work to define the concept and scope for the project.

Between the commencement and completion of a project, there are discrete stages that comprise the project work. It is not until after project completion that the deliverable will enter an operational status and become business as usual.

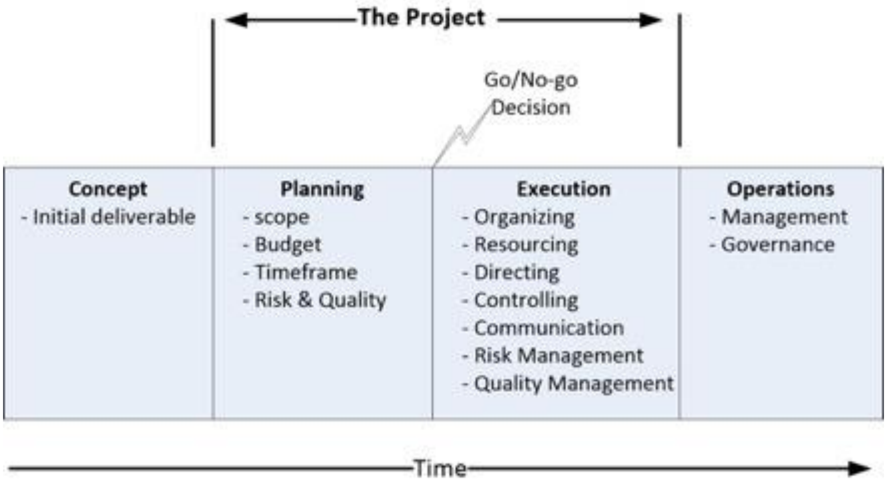


Figure 1 - The Project Lifecycle

Concept

Projects come out of a need. In the IAM world, examples of such a need include reducing costs and improving security by using identity information more effectively for onboarding and offboarding staff or a need for an enterprise LDAP directory upgrade. Such projects are typically initiated by an IT resource rather than a business resource, though a line-of-business resource might also initiate a project (e.g., to move an application from an on-premises environment to the cloud to save capital expenditure budget). The project sponsor will communicate the requirement, set the project charter, and evaluate the required activity's cost and duration. The sponsor will typically fund this stage and then engage a project manager to complete the planning stage.

Planning Stage

Once the approval to proceed has been received, the project manager will engage with the stakeholders to define the project scope. Having a clear scope around an IAM project is critical to its success. Since IAM touches virtually every application in a company and is at the core of cybersecurity protection processes, the scope can quickly expand beyond the original intent and budget if not managed carefully. As participation by one or more representatives of each relevant application is required, the project scope will also define the stakeholders. For instance, if the Finance Administration application is to be integrated with the IAM system, a representative from the Finance Department must be engaged as a stakeholder in the project.

A project initially focused on deploying an identity management package to provision staff—for example, establishing Active Directory records and email accounts—might see a request for including provisioning into corporate applications. Or someone in Corporate Governance may request additional functionality such as periodic attestation reporting and re-certification. The project manager must ensure that the appropriate stakeholders are engaged and respond to requests for input. The project manager does not decide whether requested functionality should be included; that decision is made via a Steering Committee or project sponsor and reviewed when the full scope of the project is defined.

Once the scope has been determined, the project manager will engage subject matter experts to quantify the work required and construct the project budget and schedule. The planning stage will develop a project plan that will include:

Schedule	The schedule will define the timeframe and resources required for the project to calculate the cost. A schedule is typically expressed via a Gantt chart in classic project management. A high-level Gantt chart is also helpful for Agile project management.
Stakeholder Analysis	The project manager will construct a list of project stakeholders. This list will typically include the sponsor, finance manager, human resources (HR) manager, system owners, and representatives from the IT groups that will be engaged in the project.
Resource Plan	A basic tenet of project management is that the desired resources are never available; they are typically fully engaged in other activities. The project manager must negotiate with the appropriate stakeholders to get the desired resources assigned and alter the project schedule accordingly.
Communications Plan	The project communications plan defines the "who" and the "how" for a project manager to report on project progress. The project team will likely have a file folder, wiki, or SharePoint site for the project. The project manager will regularly email a project report to the Stakeholders and send meeting agendas and status summaries to the steering committee before the project review meetings. The project plan should include a communications register that logs all communications with the stakeholders and within the project team.
Quality Management	A mechanism to ensure adequate quality in project deliverables should be defined. This mechanism should include management reviews of project documentation and properly constructed test and release procedures.
Risk Management	A project manager constructs a risk register that identifies the anticipated risks, quantifies them in terms of probability and impact, and includes appropriate risk mitigation activities.

At the end of the planning stage, there should be a good understanding of the project activities, timeframe, and cost. Typically, the project cost and duration will be known within a 10% margin.

In terms of time and money, this understanding of project costs allows the organization to make an informed decision as to whether they want to dedicate the necessary resources to a project. A decision not to proceed with a project is as successful an outcome for a project manager as a decision to proceed. It means that the organization has been spared the expenditure of resourcing a project that might otherwise have proceeded, only to be prematurely terminated when the costs blow out, resulting in associated sunk costs.

Deployment Stage

The project deployment stage will vary depending on how the project is managed: via a classic (waterfall) mechanism or an Agile project management approach.

Classic

In classic project management, the project manager will manage all project activities according to a detailed schedule that shows all the individual tasks, assigned resources, and duration. They will also schedule a regular project team meeting to review the project's progress against the schedule and note any impediment to be escalated to the steering committee for resolution.

At the Steering Committee meeting, the committee will formally accept the project deliverables, approve the phases, and resolve any issues or roadblocks that the project manager identifies. Again, the project manager cannot extend the project scope, schedule, or budget without Steering Committee approval.

The components of a classically managed project are:

Team meetings	The project team should hold regular progress-review meetings (weekly or bi-weekly). These meetings allow everyone to mark progress against the Gantt chart and determine what issues, if any, must be escalated to the steering committee.
Steering committee	The project manager will periodically present to the steering committee to review progress on the project schedule and act on any issues the team has identified. The project status report should include the progress made since the last meeting, any issues to be resolved by the steering committee, and the planned activities for the next period.
Phase transitions	The project schedule (Gantt chart) will show the project's phases. At the end of each phase, the steering committee will review the deliverables for that phase and determine whether a phase transition can be approved.
Deliverable acceptance	Each project deliverable should be formally accepted. This acceptance will typically involve the appropriate stakeholder(s)

	who must agree that the deliverable has been produced to an adequate quality level.
Project closure	A project should always include a proper project closure procedure. This procedure will typically involve a formal project review that will document the activities that went well and any learnings from the project. "Those who don't learn from history are doomed to repeat it."

Agile

Many organizations have now adopted Agile project management for smaller projects that don’t warrant the cost of a classic project management approach. These projects are typically an activity that exceeds the capacity of the normal operations staff. For instance, a significant upgrade, or migration to cloud services, will likely require a system outage or out-of-hours cutover activity. The execution of such project work must be managed. Agile project management ensures appropriate stakeholders are engaged and issues are addressed promptly, without waiting for steering committee approval.

Agile methodology divides a project into ‘scrums’ that are then further divided into ‘sprints.’ Each activity comprising a sprint will be put up in a tile on the project wall and will be moved from the "To Do" activity list to "In Progress" and then to "Done." A review meeting, sometimes called a "stand-up," will be held every few days to review current activity and document any issues or roadblocks.

Project wall	The essence of Agile project management is visibility. The Project Wall provides a physical or virtual place where the project team can view the completed, current, and waiting tasks and resource assignments.
Sprints & Scrum	These terms are used differently depending upon the context. Scrum is a framework that uses an iterative process to deliver a defined piece of functionality. It could be a product, service, or a new piece of functionality for an existing product (e.g., deploy a DBMS connector to the IAM environment). A sprint usually describes a scrum component, a time-limited activity that contributes to a scrum deliverable (e.g., 30 days for developing the reporting module).
Deliverable acceptance	One area that can suffer when using an Agile project management approach is reviewing and accepting deliverables. Acceptance testing will verify that the requirements established for a viable product have been achieved and are demonstrable. A sprint team sometimes advises on completing a piece of work and moves to the next without formal acceptance of the deliverable. A mechanism to record the acceptance of a module or deliverable is needed.

Project closure	A team meeting can be dedicated to the requisite project review in a classically managed project. It is sometimes difficult to manage the project closure in an Agile project, in which many participants have contributed to the outcome. In either project management model, a mechanism is required for all participants to agree that a project has been completed and that the resources used can be reassigned.

PMO Issues

In large organizations with a Project Management Office (PMO), an IAM project must follow corporate procedures. Typically, a PMO will have defined gating factors, or ‘gates,’ through which all projects must pass. For instance, there will normally be a project approval gate in which the appropriate managers will review the project plan and indicate their approval. There will usually be a gate in the form of a budget review to approve the assignment of resources. Similarly, there might be a gate in the form of an architecture review to approve the solution architecture. Finally, the governance outcomes should be reviewed as a necessary gate for the project. The PMO should orchestrate all these activities.

One of the benefits of a PMO is the visibility it gives to projects within an organization. This visibility is beneficial to the IAM team; it allows them to ensure any projects with an identity component are properly identified and accommodated in the appropriate work program. For instance, if an authentication gateway is being installed, any application undergoing development should be modified to use the gateway rather than maintaining LDAP lookups. Without a PMO, it is sometimes difficult for the IAM team to impact projects.

A PMO provides the opportunity to educate project managers on identity issues and to insert IAM requirements into IT projects within an organization. A project manager will use the PMO framework to:

1. manage the project through the project gates;
2. communicate the project's progress to the organization's management;
3. gain acceptance within the organization that the project goals were achieved within the approved budget and schedule.

IAM Projects

It’s often said that a good project manager can keep a project on track regardless of the topic. While this may be true, if a project manager for an IAM project is not competent in the subject, they will be disadvantaged. It is recommended that they engage a project lead who is familiar with the components of an IAM environment and understands the competency of the skills-base within the organization. If an organization cannot complete a project with in-house resources, the project manager will need to engage contractors to work on the project.

Example Project

Let’s assume the project is commenced to replace the existing IAM processes used to onboard new staff members or contractors with a new system purchased from an IAM solution vendor. The sections below work through the different project management stages for such a project.

Planning

The single most important element to define for an IAM project is the project scope. The IAM environment touches so many operational components and processes within an organization; the PM's role means they must clearly communicate to all stakeholders the full scope of the project. To properly determine the scope of an IAM project requires the PM to understand the nature of the IAM solution and its impact on other systems in the organization. The Addendum suggests some questions that should be asked in the planning phase of an IAM project.

The PM is responsible for ensuring the scope of the project is clear. Too many IAM projects proceed with misunderstandings regarding the project scope. The IAM project lead, for example, might think the project is to implement a provisioning module, whereas the application owner might think the goal is to provide better authentication functionality. The auditor, in turn, might want improved governance. Reaching a common agreement on the scope will focus all stakeholders on the extent of the project.

The following items are often inside the scope of a project of this nature:

- configuring and deploying the IAM tool
- integrating with the email system
- integrating with the system(s) that provide enterprise resource planning (ERP) functions (i.e., the computer systems that support the organization's operations)

The HR and financial management systems, however, are out of scope of this example project. While tight integration with HR could improve both the HR and IAM systems—the HR system potentially able to increase its span of control, and the IAM system benefitting from tight integration with HR for better provisioning of staff entitlements (e.g., training status, project membership, and employment status of staff)—the HR department is often reticent to make any changes to their onboarding and offboarding procedures. Evidence of a well-managed project may alleviate this fear.

The Finance department also has challenges that discourage them from agreeing to anything that will impact their systems. They typically maintain a fine-grained authentication capability within the financial management system and often distrust any external entity’s capability to do this. Externalizing access control to the IAM system will typically be less expensive and improve security, but working with Finance will require its own focused effort.

In scope will be the applications that will rely on the IAM system. The PM must communicate with each system's owners and determine what data attributes are required for users accessing each system. For example, the email system will need to know a user's first and last names and, likely, middle initial, to construct their digital identifier correctly. It might also need to know their department or group memberships. Ideally, email systems should participate in a company's single-sign-on (SSO) solution, i.e., users will be authenticated as part of the SSO solution used in the organization.

The computer applications that provide operational functionality to users should also use the organization's SSO solution. In the real world, such applications might include a production machine, a process control system, an asset control system, a learning management system, a health monitoring facility, a vehicle registration application, and so on. Any computer system that must be protected via an access control mechanism that ensures users only get access to the facilities to which they are entitled should be integrated into the organization's SSO solution. The project manager for an IAM project must ensure the requirements for these applications are canvassed at the commencement of the project.

Organizing

The success of an IAM project depends on how well it is organized. This dependency relates to how well the PM utilizes the hierarchy within the organization. Often, the execution of an IAM project is left to the people in the IAM unit within the company. This is poor practice because the IAM unit has an operational role in maintaining the IAM environment; an IAM project, however, is a time-limited initiative that will stretch the ability of the IAM unit and divert resources from their task of managing the IAM environment. While personnel with IAM experience should be involved in the deployment project, if they are seconded from the IAM unit, they should be backfilled with other personnel while they are engaged in the IAM project.

The following activities are recommended for the successful 'organizing' of an IAM project:

- Establish a steering committee – this should include the project sponsor, appropriately high-level personnel in the IT department, HR, Finance, Manufacturing, Sales & Marketing, and any other business unit directly impacted by the project. A steering committee will periodically review the project's progress and resolve any issues raised by the PM.
- Hold appropriate committee reviews – the PM must be aware of all gating factors and committees that must review the project's progress. These will include the PMO's gating (phase exit) meetings, governance reviews to ensure audit compliance, enterprise architecture committees ensuring that IAM systems comply with supported technology platforms, and finance reviews ensuring budget support for the project.

- Document a communications register – this should list to whom and via what mechanism the PM will send their project progress reports. It should include the frequency (e.g., bi-weekly), the mechanism (e.g., email, website, or other notification tools), and the media (e.g., Word document, MS Project file, etc.).
- Verify the support of a Quality Assurance (QA) program – responsible for the quality of project deliverables (such as the documents, milestones, or other deliverables). This program is particularly important to establish the accuracy (both in format and content) of the data files supporting the test plan. Identity data should be suitably anonymized for test purposes and must be restorable for regression testing.
- Create a risk register - The project team should compile a risk register that identifies the risks to the project's ability to meet its schedule, cost, and quality constraints. Each risk should be assessed for probability and impact. An IAM project should not proceed with any risk evaluated as 'high.'

Resourcing

It's a project management maxim that the preferred resources are never available. Good staff are very busy and cannot be easily seconded to a project. In an IAM project, it is essential that personnel with detailed knowledge of the company's identity management systems and policies be involved. The PM must be able to negotiate the availability of critical personnel and modify the project schedule accordingly.

As noted above, the project's budget must accommodate backfilling personnel seconded to the project. If it is necessary to 'buy in' resources, the steering committee will typically decide on the final resourcing plan and may choose to use contractors for the maintenance activity and assign experienced IAM staff to the IAM deployment project. Since the PM of an IAM project typically has no functional authority within the organization, they must use the steering committee to get the right resources assigned to the project at the right time.

A perennial problem for an IAM project is how to build IAM staff competence in a new IAM tool being acquired. The options include:

- Send selected staff from the IAM unit for training prior to the deployment activities
  It is unrealistic to expect, even experienced, IAM staff to develop competence in the new package without hands-on experience.
- Engage the vendor to do the deployment with IAM staff observing.
  This engagement is the most realistic option because it puts some onus on the vendor to 'make it work' and ensure technology transfer to the IAM staff.
- Engage the vendor for a turn-key project with the IAM unit engaged to undertake acceptance testing on the transition to operational status.
  This engagement is not ideal since, without the IAM team's active involvement, the IAM solution's successful integration into the organization's operations will be difficult.

Directing

The Directing element of an IAM project will vary greatly depending on whether a classical or an Agile project management methodology is followed.

Classic

The Gannt chart becomes the main tool for directing the project. The PM will ensure tasks are commenced on time and progress to plan by conducting a weekly or biweekly review of the schedule in periodic team meetings. Team members will report on the progress to plan for each task to which they are assigned. For tasks behind schedule or expecting to encounter problems, the PM will attempt to put a contingency in place. If a slip occurs, the PM must go back to the steering committee with a recommended strategy and seek approval or additional direction (for example, the direction to accept the slip and modify the Gantt chart or the direction to invest the resources necessary to restore the original schedule). If the steering committee approves the change, the project schedule can be re-baselined.

Agile

The PM will establish regular 'stand-up' meetings, typically several times a week, at which each 'sprint' is reviewed and tasks moved on the Project Wall from 'waiting' to 'current' to 'completed.' Each scheduled task will be discussed, and any impediments to completing a 'sprint' will be noted by the PM and addressed with appropriate management. For instance, transition to production might occur during non-business hours requiring coordination with multiple business units. The PM must ensure agreement, and appropriate resourcing, from involved parties.

The PM will raise unresolved issues with the appropriate managers.

Controlling

Control is probably the PM function that is most often performed poorly in IAM projects.

Control is a function of project management that provides feedback to the PM regarding the likelihood that the project will meet its schedule and budget constraints. PMs will typically assume that if they have planned well, organized the communication and quality assurance, adequately resourced their project, and properly directed the project tasks, nothing can go wrong. But the stories are legion that IAM projects have overrun because they impact so many functions within an organization. Managing this impact is where control comes in. Given that you cannot manage something if you cannot measure it, monitoring progress to plan is at the core of the control function. A tried-and-true tool in the PM's toolkit is Earned Value Analysis (EVA). EVA involves calculating the budgeted cost of work scheduled (BCWS), the budgeted cost of work performed (BCWP), and the actual cost of work performed (ACWP). These calculations will compare the percent completion against the budget spent and quickly identify a project experiencing overspend or over-budget issues.

As an example, a project's progress might be depicted as follows:

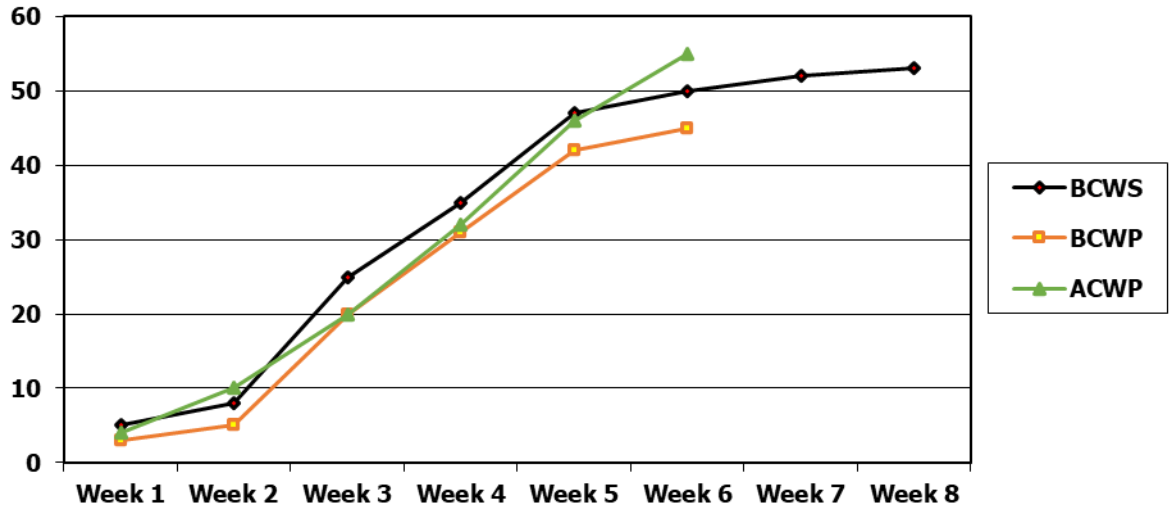


Figure 2 - Sample Budget Cost Schedule

The BCWS shows the project's schedule. It's a two-month project with a budget of $53,000 (Y axis in thousands of dollars). The current spend is $55,000 with two weeks to go. The budgeted cost of the work performed to date is $45,000. So the EVA clearly shows the project is behind on its deliverables and is currently $10,000 overspent on its budget.

Another tool is calculating a project's performance indices using quick ratios to gauge the probability of an on-time and in-budget project completion. Common indices are:

- Cost Variance: CV = BCWP - ACWP
- Scheduled Variance: SV = BCWP - BCWS
- Cost Performance Index: CPI = BCWP/ACWP
- Schedule Performance Index: SPI = BCWP/BCWS
- Critical Ratio: CR = CPI * SPI

A third useful tool is the S curve, which tracks the resource burn rate to ensure the project expenditure reduces appropriately, particularly at the end of a project. In the example above, the Actual Cost curve is not adhering to the 'S' shown in the scheduled work curve. Management of the resource burn rate is important for IAM projects since additional tasks, such as system documentation, are often not properly accommodated in the project schedule at project inception. These should not be added to the scope of the project. Instead, they should be completed as part of standard operations (i.e., outside the project).

Organizational Variances

When managing an identity project, it is worth understanding the type of organization for which the project is being undertaken.

Public Sector

When managing a project for a government department, there will often be organizational structures that make it difficult. In a project to deploy an enterprise-level solution for a large government agency with multiple departments, there were two major obstacles. Firstly, only five of the departments in the agency agreed to participate in the project; the two largest departments declined to be involved. Secondly, the agency engaged an internal technology unit to deploy all their IT projects, making it difficult to engage with the end-users directly.

The first obstacle resulted in a meeting at which the benefits of the solution were explained, and the department personnel in attendance agreed to take a 'watching brief' as to how the project developed. They agreed to proceed with the solution once they observed a successful deployment. The second obstacle was overcome by inserting a 'workshop' task in the schedule that required participation by knowledgeable persons from the involved departments. The workshop was very successful, with participants demanding ongoing involvement in the project.

Part of the stakeholder analysis for public sector projects is to understand the motivation of the sponsor and other involved public servants; their motivation it's not always to benefit the agency for which they work; it is sometimes to advance their career.

Private Sector

There is a danger to commercial projects that the scope will be too narrow. Because projects in the private sector are typically cost-constrained, there will be a reluctance to engage widely to build a comprehensive list of stakeholders that will ensure wide benefit across the company. When identifying items that will extend the scope, the project manager will often be told to place them in 'phase 2'.

While this might make commercial sense, the project manager should ensure the Steering Committee understands the ramifications of not extending the project to include the requirements of a wider stakeholder cohort. It is far better to include requirements in the scope of the initial project than to be forced to extend the project once work has started. The scope must be determined before the schedule is developed and before the execution of the project commences. Adding new requirements during the execution phase will require the project to be re-baselined, which will involve more work and should be avoided.

Academia

Managing an identity project in the academic sector can be quite complicated. The administrative staff are divided into two large cohorts: administrative officers who keep the university or school operating and academic staff with diverse identity management requirements. If the institution is involved in research, there will be a further requirement to participate in cross-institutional identity federations to access documents from remote locations.

Then there's the student cohort which consists of undergraduates, graduates, higher-degree by research students, and staff enrolled as students. Alumni comprise an additional cohort that may need to be considered.

When determining the scope, the project manager must agree on the user base to be accommodated. It is also noted that academics typically engage in a wide and diverse range of applications that might need to be accommodated by the IAM infrastructure.

Conclusion

Project management methodology should be applied to all IAM projects, even small ones. Project management ensures that a structured process is applied to the activity and that the impact of the activity on affected business units will be considered and, if necessary, included in the planning. Failure to manage an IAM activity as a project will raise the likelihood of mistakes being made and additional costs being incurred.[iii]

Author Bios

Graham Williamson

Graham Williamson is an IAM consultant working with commercial and government organizations for over 20 years with expertise in identity management and access control, enterprise architecture and service-oriented architecture, electronic commerce, and public key infrastructure, as well as ICT strategy development and project management. Graham has undertaken major projects for commercial organizations such as Cathay Pacific in Hong Kong and Sensis in Melbourne, academic institutions in Australia such as Monash University and Griffith University, and government agencies such as the Queensland Government CIO's office and the Northern Territory Government in Australia and the Ministry of Home Affairs in Singapore.

Corey Scholefield

Corey is currently a Sr. Technical Product Manager with Workday, supporting operations engineering service delivery for Workday's cloud-ERP suite. Corey has a background in public-sector identity management, having spent over 15 years working in higher education, with positions at both University of Victoria and BCNET in British Columbia, Canada.

At BCNET, Corey led a federated-identity service bureau that supported regional adoption of eduroam and SAML capabilities under the umbrella of the Canadian Access Federation. At the University of Victoria, Corey's team established an identity-management program that supported campus-wide access-management needs. Corey has deployed many IDAM technologies, including OpenLDAP, CAS SSO, Sun IDM, Shibboleth IDP, and SailPoint IdentityIQ.

Change Log

Date	Change
2021-06-21	Editorial updates; substantive revisions to IAM Projects section
2022-09-30	Substantive revisions to Planning Stage, Agile; new section on Organizational Variances

Addendum: Questions for an IAM Project Manager to ask

Identity Management

- How are user accounts created when a new staff member joins the organization? Are employees and contract staff provisioned differently?
- How are user attributes collected/determined?
- What is the business process surrounding end-users being granted entitlements to access given applications? Is user self-service supported? Is there an approval workflow to gather authorization for establishing user entitlements?
- Is there a different process for privileged accounts (e.g., accounts with admin privileges)?
- What repositories of identity information exist in the organization (e.g., LDAP directories, Databases, Active Directory), and what interfaces to the identity management environment are needed (e.g., SCIM import, REST API, Webservices Gateway; CSV import)?
- What is the business process for disabling an account and eventually deleting it?

Access Control

- What authentication mechanisms are supported (e.g., local application database, corporate LDAP directory, Active Directory, RADIUS)?
- Are multiple assurance levels supported (e.g., assurance elevation for sensitive resources)?
- Is MFA supported (e.g., U2F, DUO, push authenticators)?
- Is SSO supported? Is it only for web apps, or are other applications supported as well?
- How are SaaS apps supported (e.g., periodic synchronization of identity data, SAML)?
- How are user entitlements within an application managed (e.g., internally within the app, via an attribute passed in an HTTP header message, SAML assertion, Active Directory group membership)?
- How are application administrator rights managed (e.g., manually, via approval workflow)?

Governance

- What governance processes (e.g., re-certification/attestation reporting) are required? What audit processes must be supported?
- What governance interfaces are required to collect user account information from corporate applications (e.g., REST API, SCIM, Webservice gateway, service-bus messaging, CSV export)?

[i] Scrum Alliance, “Your Quick Guide to All Things Scrum,” accessed 29 June 2021, https://www.scrumalliance.org/about-scrum/overview.

[ii] Project Management Institute website, https://www.pmi.org/, accessed 29 June 2021.

[iii] Project Management Institute, "PMBOK® Guide and Standards - Practice Standards & Framework," accessed 29 June 2021, https://www.pmi.org/pmbok-guide-standards/framework.

Operational Considerations

Managing Identity in Customer Service Operations

By Arynn Crow, Senior Technical Program Manager, AWS Identity
and Jp Rowan, Staff Solutions Architect, Auth0

To comment on this article, please visit our GitHub repository and submit an issue.

Table of Contents

Abstract

This article will establish recommendations for best practices when managing the identities of your end-users in a customer service environment, considering the risks of both external and malicious insider threats. The following recommendations are built from the authors' experiences and observations, and the recommendations included should be considered a starting point to inspire discussion. More rigorous study is necessary to further refine guidelines for this subject.

Introduction

Even in today's highly automated world, there are many jobs that still just need a human. For many organizations, customer service is one of those jobs; when your end users have problems and have exhausted their ability to self-serve, they will turn to your customer service (CS) operations team for support with any number of the services or features you offer. Your CS team is on the frontline and feels your users' pain points more acutely than any other department. Their job, and your users' expectations of them, is to resolve any problem quickly and easily. Therein lies the tension and a core problem for the security-minded identity professional: how do we deliver on our promises of good experience and convenience to our customers while upholding our responsibility to protect their identities?

The cross-section of customer service, IAM, and security is an area that has received comparatively little attention across industry publications and working groups. It is essential to get identity management in CS right due to the consequences for your users and organization for getting it wrong.[i]

Terminology/Glossary

Account Recovery - The process of updating a user's credentials within a scenario where the user cannot validate those credentials
Account Takeover - Account takeover is a form of identity theft and fraud, where a malicious third party successfully gains access to a user's account credentials.[ii]
Agent (also "Customer Service Agent") - The person responsible for communicating with and solving problems on behalf of customers or end-users.
Channel - The communication avenue between you and your end-user, or your agent and their customer. This could be phone, chat, social media, or others.
Credentials - Any attribute or shared secret that can be used to authenticate a user.
Fractured Identity - A case where a single end-user has multiple disparate digital identities.
Impersonation - A scenario where a user is able to perform actions as though they are a known user other than themself.

Knowledge-Based Authentication (KBA) - A method of authentication that uses information known by both the end-user and the authentication service but is not necessarily a secret.
Personal Data - Personal data are any information which are related to an identified or identifiable natural person.[iii]
Social engineering - Social engineering is a method of manipulating people so they give up confidential information, such as passwords or bank information, or grant access to their computer to secretly install malicious software.[iv]
Step-up Authentication - A method to increase the level of assurance (or confidence) the system has regarding a user's authentication by issuing one or more additional authentication challenges, usually using factors different from the one(s) used to establish the initial authenticated session. The need for increasing the level of assurance is typically driven by the risk associated with the sensitive resource the user is attempting to access.[v]
Threat Modeling - Threat modeling is an analysis technique used to help identify threats, attacks, vulnerabilities, and countermeasures that could impact an application or process.[vi]
Username - An identifier unique to the authentication service used in conjunction with a shared secret to authenticate a user.

Why is this different from the rest of my IAM stack?

At first blush, it may be hard to see where customer service – a very operational function of the organization – fits into your otherwise very technical IAM strategy. In fact, CS operations are a critical part of your **IAM strategy,** not only because they represent your organization to customers during important moments ("Why can't I log in to your service my multimillion-dollar business relies on?"), but also because CS operational processes create rich attack vectors for motivated social engineers. We rely on CS agents ("agents" hereafter) to help our customers when they can't help themselves; to ensure their success in this endeavor, we entrust agents with access to private customer data and elevated privileges, from account creation to recovery. Your IAM system could be built with the most secure, sophisticated technology available, but your organization will be perpetually vulnerable unless your CS operational touchpoints are also hardened.

The number of touchpoints between your identity services and your customer service will vary by your type of organization and by the maturity of your organization for automating self-service functions for sensitive account functions. The most common use cases include:

- **General inquiries –** Typically low-risk support requests that do not require modifying account data or divulging personally-identifying information. These requests could include order status updates, troubleshooting, checking balances, etc.

- **Transactional support –** Requests to execute changes on behalf of the account holder, such as making a payment, placing or canceling orders, modifying subscriptions, or adding addresses
- **Account creation and onboarding –** Establishing information about a new administrator or user during account setup, or adding additional delegated users to a "base" account in a nested account schema.
- **Account recovery and state changes** – Highly sensitive requests to restore account access to an end-user, terminate an account, or transfer account ownership to another user
- **Compliance-related requests** - Data Subject Access Requests (GDPR), data deletion requests, Right to Know (CCPA), or similar requests that fall into the scope of a data privacy framework. These operations are sensitive because they deal with potentially large volumes of private customer data, which can result in additional penalties for mishandling.

These use cases likely feel similar to those you must consider elsewhere within your IAM systems, so what makes CS operations different? Your end-users, especially customers, have high expectations about the availability of customer service; the communication channels agents use to interact with customers extend beyond your application stack. Complicating matters further is the reality that the tools you deploy to authenticate and authorize end-users in your web or application environment may be unavailable or impractical to you in a CS environment. Agents often operate across a blend of phone calls, online chats, ticketing, in-person kiosks, social media, and embedded in third-party applications like WeChat. These diverse conditions challenge the application of consistent security rituals like authentication, even if they've been implemented on your online login portal. Organizations looking to preserve both their customer experience and security must weigh the risks of executing functions on the customer's behalf against their relative certainty of a given actor's identity; ideally, this decision-making process should be formalized in an internal framework to ensure decisions are applied consistently and can be inspected.

Establishing Assurance

Key to formalizing a framework for consistency, establishing levels of assurance for the available authentication methods will provide a baseline to determine what types of transactions should be permitted. The concept of assurance levels will likely have already been established as part of the rest of your existing access control policies, but in this case, these levels should be adapted to align with the channels and constraints of your CS interactions. It is likely your customers will have multiple interactions with customer support, and you may track those collective interactions as a "case" or something similar. For the purpose of establishing an assurance level, we will need to look more granularly at the individual interaction, which we will refer to as a "session."

For each session with an agent, the transactions that your agent is allowed to perform should be predicated on the current assurance level. Assurance level will depend on the communication channel or other circumstances but can be increased in a way that your agents are enabled to assist your users without introducing unnecessary risk.

Considering the challenges and constraints that your users will face in a session, it may be necessary to introduce authentication methods that otherwise would not be used for authenticating into your applications or for other self-service workflows.

In comparison to your application stack, it may seem abstract to refer to the process your users are going through in a CS interaction as authentication. In reality, the same primitives can be applied in these scenarios. Designing your authentication methods will help to assess the current assurance level while also reconciling the unique conditions that come into play in a CS session.

Authenticating Through an Application

Your agents and users might communicate through a support portal, contact form, or similar channel directly integrated within your application stack. If so, users will ideally be authenticating through the same service they would for any other application. If that is the case, then authentication and your associated assurance framework should map directly with the actions your agents are allowed to perform.

Authenticating With an Agent

Alternatively, customers may need to rely on an external channel to communicate, and therefore the burden of authentication may fall on your agents. In this case, the goal remains the same, to establish proof the user is who they claim to be.

Notably in the CS experience, some interactions might be very low risk, and it may be acceptable to complete the transaction with an assurance level that would not be acceptable for application access. In contrast, higher-risk operations warrant higher-fidelity authentication methods, or even Step-Up authentication, which requires progressively greater assurance relative to the requested action.[vii] Furthermore, some authentication methods used within a CS interaction may be completely unique to your application stack.

There are many options when it comes to authenticating a user in a customer service interaction. A common theme with all of these authentication methods is the need to create an association with the user's digital identity ahead of time. Establishing a high assurance level in your customer service sessions requires options tailored to the channels that you communicate through. Those options are only available if you establish the channel or method prior to the session. Creating a secure customer service interaction requires planning and implementation that starts much higher in your IAM stack.

Knowledge-Based Authentication

Knowledge-Based Authentication or (KBA) is possibly the most common, but also the least secure, second-factor mechanism to authenticate your users. KBA involves authenticating a user by asking a set of questions that your user would know the answer to. Common KBA questions include user credentials such as email or username, Social Security Number, date of birth, mother's maiden name, but can be custom or something more specific to the user's interaction with your product. The challenge with knowledge-based questions is that it is particularly difficult to ask a question your user both knows the answer to and that no one else would know the answer to. KBA is hard to store and validate in a secure manner. Unlike a password that can be stored and validated using a one-way hash, KBA answers are typically stored in plain text, which also make them particularly susceptible to being exposed to nefarious actors.

KBA is generally a weak form of authentication, which has been discouraged by the National Institute of Standards and Technology (NIST) in other environments.[viii] It should be understood that having knowledge of a user does not prove they are the account owner or should be entitled to make changes to an account. As an example, children in a household will likely have information about their parents' email addresses, physical addresses, and birth dates, but should not be entitled to access or change information on utility provider accounts their parents own. This information is also sometimes readily accessible online or is a major target for theft (e.g., social security numbers) and is available in criminal databases. Even exposing this information to your agents for the purposes of verification creates a vulnerability. When deciding if KBA is appropriate for some operations in your organization, you should consider the likelihood that the information has already been compromised.

PIN Authentication

PINs and passcodes, like passwords, fall into the category of a memorized secret.[ix]   The intent is to provide similar verification to a password but in a format that can easily conform to a constrained communication channel. PINs can easily be entered into a phone; passcodes can be communicated verbally.

Although PINs can be stored as a one-way hash, due to limited variability in characters they are more easily decrypted. Additionally, if they are provided directly to an agent, PIN authentication suffers similar shortcomings of KBA. As such, whether or not PINs are stronger than methods like KBA is highly dependent on how they are deployed.

Social Authentication

A less obvious option, but valid when leveraging an external application, is to leverage proof of access to the communication channel. In cases where the support channel leverages a social messaging platform (Twitter, WhatsApp, Facebook, Slack), it is possible to access the tool as a form of authentication.

An important step here is that the association of the existing user account and social provider needs to be made ahead of the customer service interaction in order to consider it a valid authentication method. While a viable option, it is important to consider all the common risks associated with social authentication. Social identity providers do not always verify ownership of email or phone number; they can be created at-will by an end-user and are susceptible to attack outside of the controls of your organization.

Registered Communication Channel (SMS, Phone, and Email) Authentication
Sending a one-time passcode to your user via SMS, email, or phone call is another common method of authentication used to validate a CS session. The code sent to the user is only valid for a single use and should be time-bound; if exposed to your agent or through a man-in-the-middle attack, it does not carry the same risk of being replayed like a memorized secret (KBA, PIN, passwords, etc.).

The use of SMS authentication does suffer some weaknesses. An attacker could gain possession of the user's phone or perform a SIM swap attack.[x] Furthermore, requesting a user to communicate a one-time passcode to an agent normalizes the behavior which could be used as part of a phishing attack. Despite these flaws, SMS continues to be a popular option due to ease of customer use and widespread adoption in application authentication.

Voice Biometrics

Biometrics are increasingly common authenticators across the web, appreciated for their convenience and improved security over methods like password-based authentication. Naturally, organizations have begun deploying these methods in their customer service environments as well, such as by deploying voice biometrics in phone channels to provide low-friction authentication. The appeal of tackling the hard problem of sufficient assurance in customer service with something convenient and secure like biometrics is obvious, but a few challenges you need to consider are:

- There will be different regulatory requirements in each country you operate (or, as with the US, even different regions within the same country may have different requirements). The perception your end users have about biometric ethics may impact the way you collect, store, and apply biometric data. User privacy is paramount.

- If you do not already offer or plan to offer biometric sign-in on your web platform, you're faced with the prospect of building or buying a system only for your customer service channel. In that scenario, you will need to campaign to get your customers to register a biometric specifically for contacting customer service (which they likely hope to never need); alternatively, some organizations "passively" enroll callers into voice authentication.

Biometric implementation in customer service is a complex topic that will require the cooperation of your security, software engineering, and legal teams to ensure you're implementing the correct authenticator for your organization's needs and adhering to all compliance requirements.

Device Authentication

In cases where your users have installed an application on their device, it might be possible to leverage that device as a form of authentication. The most common way is to have the agent trigger a push notification to be sent to the user's device. The service the agent used to trigger the notification then waits for a response back from the device to notify the agent the message has been accepted. This method provides a particularly high level of assurance since it leverages an existing session with your application and proof of possession of the device.

Account Recovery

We are distinguishing account recovery from routine authentication to underscore the increased sensitivity and need for special diligence. Your first goal with account recovery should be for your users to not need it often, as a result of proactive account and security hygiene. Your second goal should be to avoid relying on manual recovery for your customers, such as intervention with customer service, because it is a high-risk operation. Many organizations find that they cannot achieve both objectives and enable their customer service representatives to assist with break-glass account recovery measures when end-users have forgotten or lost access to all their means of authentication, such as by modifying the user's email or password. There are a few common methods of verifying identity when users need assistance recovering their accounts:

- Use of established authenticators previously associated with the account. These methods are strong but may be of limited utility by use case.

- Use of KBA. Even in low-risk use cases, KBA is weak and should be avoided; if this is not possible, bias towards challenge questions that are more extensive than your low-level authentication questions, cannot be obtained online (such as order history questions known only to you and your customer), and cannot be easily phished from your frontline operations.

- Use of real-world identity documents, such as driver's license and utility bills. If you didn't collect these documents from your user previously for comparison, these should be used in combination with another method to ensure the person who provides you with a document is the correct account owner.

Ultimately, account recovery is a high-risk operation; your users may contact you because they've lost access to any authenticators they could use to self-recover, which means you will be faced with the choice of accepting that your user will be unable to recover their account, or accepting a degradation in your overall security posture. If you maintain a high bar for creating and logging into your accounts, but a weak one for recovering them, this information could proliferate online and be used exploitatively. Always notify your customers about changes to account identifiers and credentials, and give them the option to report, approve, or revert changes initiated with low assurance.

Your organization will need to decide its tolerance for risk in account recovery - or if any risk is acceptable at all - versus its user experience, which may vary depending on what types of accounts you manage. As an example, high-value, high-risk sectors, like large Business to Business accounts, may warrant different processing than retail consumer or public library accounts; there may even be cases where it is appropriate to delegate part of this function to your legal team for more intensive identity verification than your operations will be able to execute.

More on account recovery is available in the IDPro Body of Knowledge article, "Account Recovery."[xi]

Controls

Understanding that your CS operations teams may have access to elevated data and privileges, it is important to have controls in place to prevent misuse (intentional or otherwise) and identify problems quickly. These controls should be considered for all areas within your organization, but there may be additional complexities in organizations with large CS environments.

Permissions controls

In fast-changing environments where seconds matter, operations management will be keen to ensure there is as little downtime for their employees as possible and that the agents have sufficient privileges necessary to perform their jobs. The decision to aim for immediate issue resolution at first contact by assigning extended privileges to the agent is a recipe for overprovisioning. Over time, with insufficient baselining and auditing procedures, this effect can snowball; employees will continue to collect privileges as the demands of their job evolve. Over time (especially in large, complex organizations), the governance conventions of the resources and policies gating those resources shift, leading to role, policy, or attribute explosion, depending on your governance system. This also leads to overprovisioning and, worse, an inability to effectively audit potentially over-provisioned users, as the administrator may not understand what privileges should be removed.

A full analysis of different access control governance models is beyond the scope of this article; other resources, like the IDPro Body of Knowledge "Introduction to Access Control" and "Policy-Based Access Control" offer a more detailed overview of the advantages and disadvantages of Policy-Based Access Control, Role-Based Access Control, and Attribute-Based Access Control.[xii] While the fundamentals of access control do not change for your operations team, depending on the size of your organization, the scale and complexity might; you may find that your operations access needs more drastic and frequent change than sales, engineering, management, et cetera. Finally, it is imperative that your team or IAM resource administrators have mechanisms for auditing privilege use against your organization's policies to ensure your controls are working as intended and preventing misuse.

Risks/Consequences

Administrators of IAM operational functions will, by nature of the job, encounter a number of unique scenarios and edge cases within their organizations beyond what can be fully cataloged in this article. Operations environments can be fast-paced and quick to change, adapting to support the organization as it evolves; nevertheless, it is critical to remain diligent. The channels through which users interact with customer support are desirable attack vectors. Bringing a human into the equation creates the opportunity for exploitation that your application stack would otherwise not be vulnerable to.

The coming paragraphs acknowledge the most common risks, known anti-patterns, and suggested best practices as a reference. This list should not be considered definitive; it is a good starting point to avoid common pitfalls.

Social Engineering

The industry is increasingly acknowledging the significance of the threat posed by social engineers; a 2020 Verizon Data Breach Investigation found phishing and other forms of social engineering were involved in 22% of attacks.[xiii] Customer service agents are especially vulnerable because they are your direct line to the public, they're entrusted with sensitive privileges necessary to resolve tough customer problems, and they likely have a vested, performance-driven interest in making your customer happy. Unmitigated, this can be a severe risk for your organization.

Do:

- **Provide access only to resources that are required to perform the job**. This mitigates damage in case your agent is targeted in a social engineering attack.
- **Routinely educate personnel on the most common types of phishing and engineering attacks.** Ensure they know how to recognize and escalate suspected attacks. Phishing attacks are constantly evolving and becoming more sophisticated; continuous monitoring and updating on current trends is an important part of agent education
- **Establish regular audits of your resources and access rights** to ensure you are continuing to enforce least privilege even as job functions change over time.
- Establish a thorough catalog of the resources your organization maintains and an understanding of their relative sensitivity; require progressively higher-fidelity proofs to gain access to more sensitive resources for both employees and end-users, such as management chain approvals, additional identification, or other checks as appropriate.

Do Not:

- **Use information that is easily accessible to the public** - online or offline - as part of your account authentication or recovery processes

Account Takeover

In a customer service interaction, account takeover is made possible by allowing an attacker to modify a victim's credentials from something the victim knows and has access to, to something the attacker knows and has access to. Credential changes are the catalyst to a chain of events that can result in a valid user losing all access to their account and instead place full control in the attacker's hands. This is the worst case and common result of poor controls within a customer service stack.

It is important to note that credentials can be more than just a password. If a phone number or email address can be used as a channel for account recovery, they too should be considered a credential.

Do:

- Leverage existing authentication methods to establish a secure session with users in customer service interactions. Whenever possible, use existing authentication workflows to establish a legitimate session with your users.
- **Align your session assurance levels with those applied to your applications.** Only when the assurance level matches the requirements for a specific transaction should it also be allowed in a customer service interaction.
- **Leverage existing self-service channels for account recovery when possible.** All self-service account recovery channels should have been threat modeled with the

design of your IAM stack and therefore would not require additional vetting for the purpose of customer service interactions
- **Notify the end-user in the case that a credential has been updated in a customer service interaction.** A message should be sent to all possible (prior) validated communication channels to notify an end-user when a credential has been updated by a Customer Service Agent.
- **Establish controls that allow for changes made by a Customer Service Agent to be reversed by the end-user.** In addition to notification, end-users should have the option to escalate or reverse credential changes enacted against their accounts that they did not authorize.

Do Not:

- **Allow users to update credentials in Customer Service interactions** unless you can satisfy the level of authentication required for these high-risk operations as required by your risk framework

Impersonation

Within an application, impersonation occurs when actions are taken on behalf of a user, without being initiated by that user, are unidentifiable as such. Because customer service agents will often need to perform actions on behalf of other users or possibly replicate another user's experience, it is quite possible that the tools provided to the Customer Service Agents might result in enabling impersonation.

Typically, impersonation occurs because it is simply easier to have a customer service agent login on behalf of the user they are assisting than it is to build out the necessary tooling for them to perform their job securely. Once operationalized, tools and workflows that rely on impersonation create opportunities for users to be harmed without notice and are an enticing target for attackers that wish to wreak havoc without a trace.

Do:

- **Build tools that allow Customer Service Agents to manage end-user data outside of the core application.** Separating the customer service use cases from your core applications makes it easier to audit the actions taken by your agents and helps to avoid scenarios where impersonation might accidentally occur
- **Require end-users confirmation before Customer Support Agents can perform actions on their behalf.** Establishing consent workflows helps build trust with your users and helps to ensure that elevated actions taken by agents are scoped to specific user interactions.

Do not:

- **Allow Customer Service Agents to login as an end-user.** Any scenario where a customer service agent is acting on behalf of an end-user or needs to replicate the end-user experience must be auditable as such. All actions taken by the agent should be recorded in the system of record as such.

Fractured Identity

Fractured identity occurs when a user is unintentionally associated with multiple accounts. In the case of customer service interactions, this typically occurs when agents establish a new user identity for an existing known user or when a user identity created in a customer service interaction cannot be reconciled with their digital identity. Creating multiple digital identities for an end-user results in a poor end-user experience and can typically result in more overhead expenses wasted to reconcile the fracture.

Do:

- **Create tooling to search for user accounts by fuzzy terms and multiple indexes.** Fractured identities are often introduced when friction is introduced into an agent's workflow and identifying the existing account is more effort than the agent feels is worth the effort. Tooling to find the appropriate user accounts should be implemented with diligence to ensure it aligns with the necessary privacy controls avoiding overexposure of customer data.
- **Create tooling that allows a user to link disparate accounts.** If you have circumstances where fractured account identities might be common, creating self-service tooling to link or merge accounts will save time and minimize frustration for both your agents and customers.

Do Not:

- **Make credentials immutable.** Users will always have justifiable cause to want to update their email, phone number, or username.

Unnecessary Friction

The most secure application is one that doesn't exist. In that vein, it is easy to dismiss the user experience, and therefore any friction incurred by implementing rigorous security controls, as a cost of doing business securely. However, the tradeoff isn't that simple. Bad security experiences have potential risk and financial implications; users who can find workarounds for aggravating security controls will use them. Inefficient processes can also impact your bottom line: every second that your agents spend on the phone or chat attempting to identify a customer is money spent. Review your processes to ensure there are no duplicated steps and verify that there are pathways for customers to authenticate via the same convenient factors they would employ in your web environment (such as hardware authenticators and biometrics).

Do:

- **Match the level of assurance to the risk of the operation.** It may be more appropriate for more onerous authentication processes to start with a basic level of assurance and use step-up authentication later on if necessary. Deciding which process to use might require you to work closely with your operations teams to categorize different types of actions and assign appropriate authentication methods.

- **Go for stronger proofs instead of layers of weaker proofs.** Delegate as many authentication procedures as possible to something the customer **has** or something the customer **is,** as opposed to knowledge-based authentication, for both security and user experience.

Do Not:

- Pile on authentication layers if they aren't necessary to achieve an appropriate level of assurance for the support your customer needs.

Conclusion

Some concepts from this article may be new to you or instead may offer new ways of looking at and addressing age-old problems for Identity. Because there are likely as many facets to your operations as there are to your business or organization, measures to address their challenges securely won't be one-size-fits-all. It is important to establish a strong partnership between your operations and security teams to solve problems collaboratively. Drawing from the use cases and best practices within this article, as well as other resources within, you will be well-equipped to start these conversations within your organization and begin building or improving a strategy to meet your user needs while protecting their data.

[i] As demonstrated by the 2020 Twitter security incident, in which numerous high-profile accounts were compromised, support tooling is a low-complexity vector for high-impact attacks. See: "An Update on Our Security Incident" Twitter, July 2020. https://blog.twitter.com/en_us/topics/company/2020/an-update-on-our-security-incident.html

[ii] "Terminology in the IDPro Body of Knowledge," IDPro Body of Knowledge, accessed 17 April 2021, https://bok.idpro.org/article/id/41/.

[iii] Ibid.

[iv] Ibid.

[v] Ibid.

[vi] Ibid.

[vii] Kaushik, Nishant, "Designing MFA for Humans," IDPro Body of Knowledge, 30 October 2020, https://bok.idpro.org/article/id/49/.

[viii] "NIST 800-63b FAQ". January 2020. https://csrc.nist.gov/publications/detail/sp/800-63b/final

[ix] Paul Grassi (NIST), Elaine Newton (NIST), James Fenton (Altmode Networks), Ray Perlner (NIST), Andrew Regenscheid (NIST), William Burr (Dakota Consulting), Justin Richer (Bespoke Engineering), Naomi Lefkovitz (NIST), Jamie Danker (DHS), Yee-Yin Choong (NIST), Kristen Greene (NIST), Mary Theofanos (NIST). "Digital Identity Guidelines: Authentication and Lifecycle Management," Section

5.1.1.1, National Institute of Standards and Technology Special Publication 800-63B, June 2017. https://csrc.nist.gov/publications/detail/sp/800-63b/final

[x] Barrett, Brian, "How to Protect Yourself Against a SIM Swap Attack," Wired, 19 August 2018, https://www.wired.com/story/sim-swap-attack-defend-phone/.

[xi] Saxe, Dean, "Account Recovery," IDPro Body of Knowledge, 17 April 2021, https://bok.idpro.org/article/id/64/.

[xii] Koot, André, "Introduction to Access Control," IDPro Body of Knowledge, 17 June 2020, https://bok.idpro.org/article/id/42/ and McKee, Mary, "Policy-Based Access Control," IDPro Body of Knowledge, 19 April 2021, https://bok.idpro.org/article/id/61/.

[xiii] "2020 Verizon Data Breach Incident Report" P 7. Gabriel Bassett, C. David Hylender, Philippe Langlois, Alexandre Pinto, and Suzanne Widup. https://enterprise.verizon.com/resources/reports/dbir/2020/introduction/

Identity and Access Management Workforce Planning

Put identity at the center of your cybersecurity workforce.

By Kenneth M. Myers

Table of Contents

Abstract

This article offers a practical approach to help identity and access management (IAM) practitioners and managers understand how to advise organization leadership on identity and access management workforce planning. While workforce planning is usually a Human Resources (HR) task, the IAM practitioner, their hiring managers, and their HR teams should know the tasks, knowledge, and skills expected across the IAM industry. By capturing the tasks, knowledge, and skills across the various identity and access management service areas, this competency model is tailorable to fit most organizations' needs to include any sector-specific training. Using the U.S. Federal Government's IAM frameworks as a working example, this article seeks to help mature the identity and access management profession and create a more consistent experience across organizations for identity and access management practitioners.

Keywords: identity and access management, cybersecurity, workforce planning, competency model, enterprise architecture, competency model, work roles

Introduction

Identity and Access Management (IAM) is a challenging profession. An identity process is usually the first interaction a new employee or customer experiences with an organization and is often not smooth. These interactions may include:

1. Filling out a job application multiple times for identity verification.
2. Creating a username and password at almost every website for authentication.
3. Making requests across multiple help desks asking for user access and sometimes waiting days or weeks for approval.

Identity and access management are fundamental to digital transactions. When non-identity professionals are responsible for everyday identity tasks, organizations may see misconfigurations, suboptimal user experience, or potential data leakage. Most importantly, organizations put themselves at an increased risk due to a lack of a holistic view of user access and security across the organization. To clarify job responsibilities and required skills, organizations should use a cybersecurity workforce framework for workforce planning.

- A workforce framework is **a set of tasks, knowledge, and skills (TKS)** for someone to be effective in their job.
- Workforce planning ensures an organization **has the right talent** to execute business and technical objectives.

While workforce planning and a workforce framework are primary tasks of human resources personnel, IAM practitioners need to be active participants in providing the TKS required for a workforce framework in order for a workforce planning effort to be successful. A workforce framework can also be an effective tool to allow practitioners to identify skill and knowledge gaps. A workforce framework consists of multiple parts.

1. Competency – A method to assess someone. A competency is comprised of TKS statements.
2. Task – an activity toward an achievement.
3. Knowledge – A retrievable set of concepts within memory. Multiple statements may be required to complete a task.
4. Skill – The capacity to perform an observable action. There is a many-to-1 or 1-to-many relationship between skill statements and task accomplishment.
5. Work Role – A consistent method to describe the competencies and TKS needed to perform a responsible work area.

It's worth noting a few clarifying points.

1. A competency model is a set of TKS needed for effective job performance. A competency model is part of a workforce framework.

2. In terms of workforce planning, a maturity model is a method to measure capabilities to a specific seniority or optimization level.
3. A work role is not the same as a job title. A job title is usually organizationally set, while a work role is a consistent way to describe a type of work. A title may be specific to an organization, but a work role should be consistent across organizations.

A maturity model can incorporate a competency model to outline a collection of TKS per level of seniority, from entry-level to senior-level.

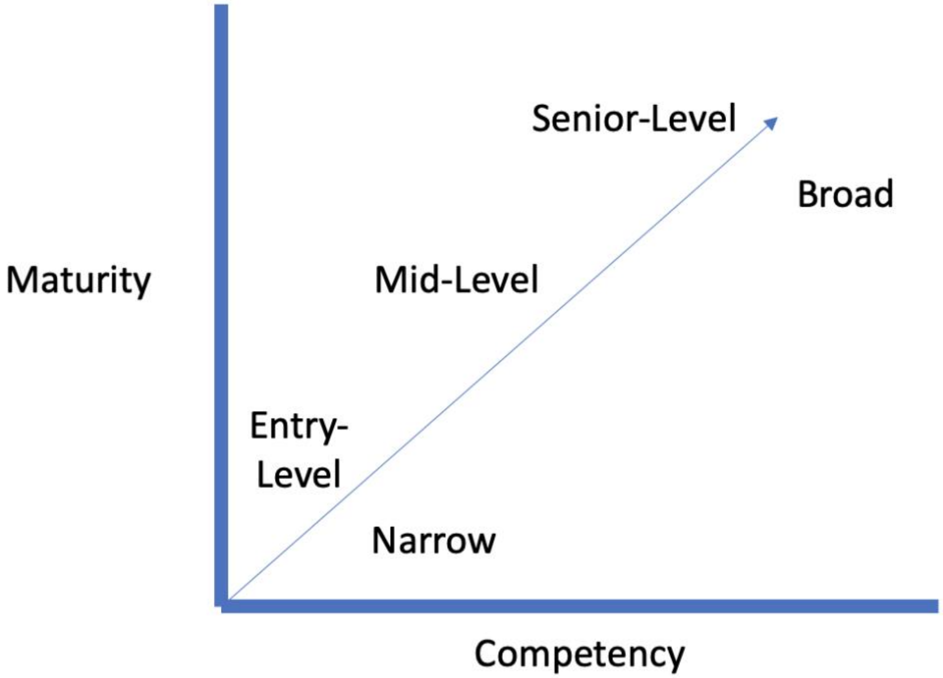


Figure 1. Example of the interrelation of a competency model and maturity model

This article offers a practical approach to help identity and access management practitioners and managers understand how to advise organization leadership on identity and access management workforce planning. The next section outlines why the IAM profession needs its own workforce planning and competency model.

Terminology

- **Access Management** – Use identity information to provide access control to protected resources such as computer systems, databases, or physical spaces.
- **Attributes** – Key/value pairs relevant to the digital identity (username, first name, last name, etc.).
- **Authenticator** – The means used to confirm the identity of a user, processor, or device, such as a username and password, a one-time pin, or a smart card.
- **Binding** – Associating an authenticator with an identity.

- **Competency Model** – A collection of tasks, knowledge, and skills (TKS) needed for effective job performance. A competency model is part of a workforce framework.
- **Credential** - A credential allows for the authentication of an entity by binding an identity to an authenticator.
- **Credential Management** – How to issue, manage, and revoke authenticators bound to identities. Credential Management roughly corresponds to the IDPro term for Credential Services; we use the term Credential Management here to correlate to the Federal Identity, Credential, and Access Management (FICAM) initiative's terms.[i]
- **Identity and Access Management** – The discipline that enables the right individuals to access the right resources at the right times for the right reasons.[ii]
- **Identity and Access Management Workforce Planning** – Activities involved in ensuring an enterprise identity and access management team are staffed with the right talent to execute business and technical objectives.
- **Identity Management** – A set of policies, procedures, technology, and other resources for maintaining identity information.
- **Identity, Credential, and Access Management** – Programs, processes, technologies, and personnel used to create trusted digital identity representations of individuals and non-person entities, bind those identities to credentials that may serve as a proxy in access transactions, and leverage the credentials to provide authorized access to an organization's resources.[iii]
- **Workforce Framework** – An outline of the job categories, work roles, and competency models needed to execute workforce planning.
- **Workforce Planning** – Activities that ensure an organization has the right talent to execute business and technical objectives.

Acronyms

- CISM - Certified Information Security Manager
- FICAM – Federal Identity, Credential, and Access Management
- IAM – Identity and Access Management
- ICAM – Identity, Credential, and Access Management
- MFA – Multi-factor authentication
- NICE – National Initiative for Cybersecurity Education
- NIST – National Institute of Standards and Technology
- TKS – Tasks, Knowledge, and Skills

Problem Statement

While various research and frameworks exist on general cyber workforce planning, there is a lack of specific information for IAM workforce planning. The U.S. Federal Government has many publicly available documents that help see the evolution of cybersecurity workforce planning in large organizations with diverse cybersecurity workforce and enterprise architecture. The Office of Personnel Management, the head human resources organization for the U.S. Federal Government, identified identity management as a technical cybersecurity competency and references the National Institute of Science and Technology (NIST) National Initiative of Cybersecurity Education Framework (NICE) as the primary source for identifying and defining cybersecurity roles.[iv] However, the NIST NICE Framework does not include specific IAM roles.[v]

Outside of the U.S. government, various frameworks may be adapted for general use. Additionally, there are a variety of vendor-specific training materials available, including:

- Mastering Identity and Access Management with Microsoft Azure[vi]
- Identity, Authentication, and Access Management in OpenStack[vii]
- Oracle Identity and Access Management[viii]
- Securing the Perimeter (using Gluu)[ix]

This focus on vendor-specific training is one potential reason why there appears to be a growth in knowledge around specific products versus a focus on the underlying standards and technologies that enable IAM. The 2021 IDPro Skills, Programs, and Diversity Survey also highlighted this finding in the context of the Dunning-Kruger Effect.

- The survey noted that 16% of respondents are interested in vendor-neutral training leading to certification. The IDPro addressed this need with the new Certified Identity Professional vendor-neutral certification.
- The survey noted a Dunning Kruger effect to describe why someone proficient in a particular vendor product could create a belief that they are experts in IAM overall.

Major cybersecurity certifications include Identification and Authentication or Identity and Access Management as a knowledge domain and include overviews on access, authentication, and authorization principles. While important, including IAM as a sub-topic in the field of cybersecurity is insufficient to help IAM practitioners learn what they need to know to work effectively in their roles. The next section outlines why IAM practitioners should be involved in workforce planning.

Why is IAM Workforce Planning Necessary?

This paper asserts that organizations need IAM workforce planning to ensure they hire and train their IAM staff and decrease potential IAM-related attack vectors. Without knowledge and training, IAM processes may be implemented by individuals with only a basic understanding of IAM best practices, resulting in regularly exploited attack vectors. For example, the top two exploit actions in the 2021 Verizon Data Breach Investigation Report included phishing and stolen credentials.[x] One of the primary mechanisms to reduce the successful use of phishing and stolen credentials is to implement multi-factor authentication (MFA). Using MFA is a known best practice among IAM professionals, but is it known to software developers or system administrators? We can help address this competency gap by creating and growing a professional IAM workforce through workforce planning and a competency model.

Using the same example from above, implementing MFA is the top mitigation technique, but not all MFA is the same.[xi] An untrained professional may recommend a non-phishing-resistant option that is more robust than just a username and password. A more experienced professional may additionally suggest a combination of phishing-resistant and non-phishing options with the risk and cost of each approach. The next section outlines how IAM practitioners can get involved in workforce planning.

Define Your IAM Team

The Federal Identity, Credential, and Access Management (FICAM) architecture is a U.S. government reference architecture designed for federal agencies.[xii] (See Figure 2 for a depiction of the FICAM architecture.) This paper takes the U.S. Federal ICAM architecture as a starting point for IAM workforce planning, including building a competency model. A workforce framework and competency model are guidelines, usually managed by your human resources office but developed by practitioners.

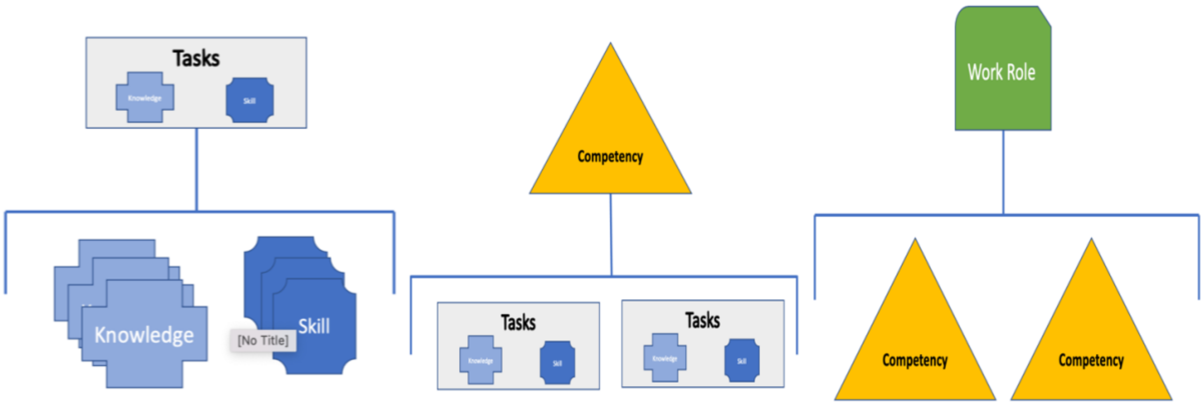


Figure 2. Knowledge and Skill combine to encompass a task. Multiple tasks encompass a competency. Multiple competencies define a work role.

Even though the FICAM architecture was developed for the U.S. Government, many of the capabilities and services are common for all organizations in that all organizations should manage identities, credentials, and access. Organizations can adopt and adapt this approach to their specific identity reference architecture as well.

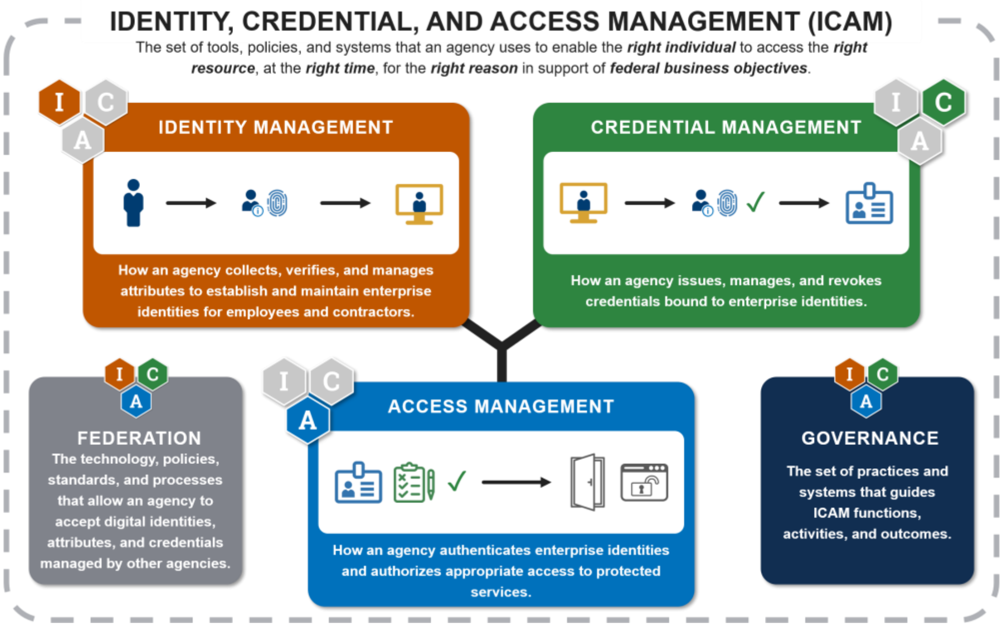

IDENTITY, CREDENTIAL, AND ACCESS MANAGEMENT (ICAM)
The set of tools, policies, and systems that an agency uses to enable the right individual to access the right resource, at the right time, for the right reason in support of federal business objectives.
I C A
IDENTITY MANAGEMENT
How an agency collects, verifies, and manages attributes to establish and maintain enterprise identities for employees and contractors.
I C A
CREDENTIAL MANAGEMENT
How an agency issues, manages, and revokes credentials bound to enterprise identities.
I C A
FEDERATION
The technology, policies, standards, and processes that allow an agency to accept digital identities, attributes, and credentials managed by other agencies.
I C A
ACCESS MANAGEMENT
How an agency authenticates enterprise identities and authorizes appropriate access to protected services.
I C A
GOVERNANCE
The set of practices and systems that guides ICAM functions, activities, and outcomes.

Figure 3. FICAM Architecture

The FICAM architecture defines five domain areas:

1. identity management
2. credential management
3. access management
4. (programmatic) governance
5. federation

After defining your IAM architecture, the next step is to use the NIST NICE Framework to convert the FICAM architecture capabilities into TKS. The NIST NICE Framework uses a simple formula to develop easy-to-read and understandable statements.

- Task – an activity toward an achievement.
- Knowledge – A retrievable set of concepts within memory. Multiple statements may be required to complete a task.
- Skill – The capacity to perform an observable action. There is a many-to-1 or 1-to-many relationship between skill statements and task accomplishment.

Table 1 contains an example of an ICAM Competency Model from the Identity Governance Framework.[xiii] This ICAM Competency Model is only an example and can be modified or altered to fit your organization's needs. One distinct difference between the FICAM architecture and other IAM architectures is including identity proofing as part of the identity management service. In an enterprise scenario, identity proofing may be a human resources task as part of employee onboarding or a third-party business application task in customer onboarding. The FICAM Architecture has primarily focused on workforce identity use cases, and additional research is necessary to add customer or non-person TKS.

	Identity Management	**Credential Management**	**Access Management**
Task	1. Perform identity proofing activities 2. Develop an identity directory maintenance plan 3. Review identity information for currency and accuracy 4. Install, update, and maintain identity directory services 5. Conduct role and group modeling 6. Create and automate workflows for provisioning, entitlements management, and identity records management	1. Enroll users in a credentialing process 2. Bind an authenticator to an identity 3. Perform Credential lifecycle management actions such as activate, renew, reset, suspend, revoke, renew, or terminate 4. Issue Public Key Infrastructure (PKI) and other types of credentials	1. Configure and manage single sign-on services 2. Configure directory and agent integration with Single Sign-On 3. Identify methods and integrate applications with Single Sign-On 4. Operate and Manage policy decisions and enforcement points 5. Configure applications
Knowledge	1. Knowledge of identity lifecycle management 2. Knowledge of identity proofing methods, strengths, and weaknesses 3. Knowledge of identity directory technology and services 4. Knowledge of identity aggregation techniques 5. Knowledge of privacy laws and impact on	1. Knowledge of credential lifecycle management 2. Knowledge of authenticator types, strengths, and weaknesses 3. Knowledge of authenticator binding techniques	1. Knowledge of authorization models 2. Knowledge of network and cloud authentication techniques 3. Knowledge of access policy lifecycle management 4. Knowledge of privilege access management 5. Knowledge of network routing

	identity data collection and maintenance 6. Knowledge of entitlements management and workflows		
Skill	1. Skill in identifying an identity proofing process to an identity assurance level 2. Skill in configuring and maintaining an identity directory service 3. Skill diagnosing directory connection issues 4. Skill in performing identity lifecycle management 5. Skill in preparing and executing access reviews and recertifications 6. Skill in managing entitlements	1. Skill in identifying an authenticator to an authenticator assurance level 2. Skill in binding authenticators to directory records across various authenticators 3. Skill in performing credential lifecycle management	1. Skill in determining an appropriate authorization model based on the use case 2. Skill in implementing authentication techniques across multiple environments 3. Skill in managing access requirements using a policy decision and enforcement point 4. Skill in implementing and managing privileged access management tools

Table 1. An IAM Competency Model aligned with the FICAM Architecture

An organization can now define the roles necessary to perform the tasks with a competency model. The list below describes the most common organizational roles to operate an enterprise identity infrastructure. Smaller organizations may have fewer roles performing more tasks, while larger organizations have more roles performing more fine-grained tasks. The following table provides an example of how an identity task differs between a large organization of multiple operating divisions and a small organization of fewer operating divisions. For example:

Task	**Large Organization with Multiple Operating Divisions**	**Small Organization of Two or Fewer Operation Divisions.**
Perform identity proofing activities	All identity proofing is outsourced to a 3rd party with a system administrator configuring a federation with the 3rd party.	Human resource personnel typically perform workforce identity proofing. For business applications, may perform custom-coded knowledge-based questions to 3rd party.

Issue authenticators and other types of credentials	Multiple administrators for each type of credential. It may involve a dedicated PKI team.	A small number of administrators perform this task for all credentials.
Configure directory and agent integration with Single Sign-On	It may involve multiple teams and administrators depending on where a directory location and which office owns it (e.g., cloud, enterprise, or application)	It may involve one team or administrator.
Provision accounts to endpoint services and applications	Integrated solution with human resources and endpoints to keep attributes and entitlements synced.	Various system administrators perform manual tasks.

Table 2. Sample IAM tasks based on organization size

The next section includes suggested NIST NICE work roles and an example evolution of an IAM team.

Evolve Your IAM Team

IAM-specific TKS now exist to define an overall IAM competency. This IAM competency can now be added to NIST NICE-defined work roles. The seven key roles, modeled after the NIST NICE Framework, within most IAM programs include:

1. Program Manager – A managerial role responsible for leading, coordinating, communicating, and integrating the program's efforts. This role is accountable for the program's overall success, ensuring alignment with critical agency priorities. A program manager is the overall responsible party for the enterprise identity program. Depending on your organizational naming structure, this role may also be called a director, branch chief, or associate vice president. This person should report directly to an executive to ensure proper corporate support.
2. System Administrator – A purely operational role that installs, configures, troubleshoots, and maintains server configurations (hardware and software) to ensure confidentiality, integrity, and availability. A system administrator usually manages accounts, firewalls, and patches. They are responsible for access control, credential management, and account creation and administration, and their role may be shared with other departments outside of IAM. Their actual job title may likely align with specific vendors ("Vendor Name" Administrator) or a function (Directory Administer).

3. Software Developer – Generally either a system design or system operations role, this role is responsible for developing and writing new (or modifying existing) computer applications, software, or specialized utility programs following software assurance best practices. Most likely, software developers may code a login page or federation assertion for broader software development tasks.
4. Network Specialist – A purely operational role that plans, implements, and operates network services/systems, including hardware and virtual environments. A network specialist may double as a system administrator or be responsible for establishing and maintaining network authentication and authorization services. This specialist is often shared with other departments outside of IAM.
5. Enterprise Architect – Primarily a system design role that is responsible for developing and maintaining business, systems, and information processes to support enterprise mission needs. This includes developing information technology (IT) rules and requirements that describe baseline and target architectures. An identity enterprise architect may double as a security architect, or their work role is labeled a security architect.
6. System Security Analyst – Often either a system design or system operations role responsible for analyzing and developing the integration, testing, operations, and maintenance of systems security. An analyst can be a technical or non-technical role that collaborates with application owners and other enterprise teams to translate business requirements into IAM workflows and processes. Sample tasks may include role mining, access requirements, attribute mapping, and similar IAM tasks.
7. System Testing and Evaluation Specialist – Often either a system design or system operations role responsible for planning, preparing, and executing systems tests to evaluate results against specifications and requirements and analyze/report test results. They develop and execute software and IAM process testing before being implemented in a production environment. This role may have a title of QA or Tester.

An organization should have the ICAM team report to an executive steering or governance body to help integrate digital identity processes into overall enterprise risk management.

Evolution of Team Development

Most organizations follow a similar group development pattern that aligns with Tuckman's group development stages: Forming, Storming, Norming, Performing, and Adjourning.[xiv] This paper looks at the first three stages on the way to a well-performing IAM team.

Forming

In the forming stage, an organization learns about the opportunities and challenges of not having a dedicated IAM function. The organization agrees on creating a dedicated position as the start of a broader IAM function. Most organizations find that they need a central person to track or liaison across the various identity functions within an organization. This decision is usually precipitated by corporate events such as an audit finding, a cyber incident, or a new security shift. This role typically aligns with a **Program Manager** and may report to either a CIO or CISO or one level below an executive position. The primary function of the program manager in this stage is to identify, track, and report on high-risk identity processes and recommend methods to mitigate risk. They may not have a dedicated team or responsibility at this stage.

Storming

In storming, IAM responsibility is being established with broader organization acceptance. Leadership supports help gain operating division acceptance of some loss of IAM control for the greater good of organizational efficiency and potential cost savings. At this stage, the **Program Manager** has gained increased responsibility and can create a primary identity team of existing **System Administrators** or **Software Developers** depending on the organization's enterprise architecture. These administrators may specialize in a single product or a specific technology, such as directories or authentication. Centralizing the responsibility and team may coincide with a shift in the technology approach. The Program Manager may identify additional positions, such as an Identity Architect, otherwise known as an Enterprise Architect, to develop rules and requirements for the desired identity infrastructure target state. Smaller organizations can utilize senior system administrators as an architect because they are most familiar with the systems, vendors, and organization's mission to propose a target state. Larger organizations may choose an Architect removed from the day-to-day technical challenges to focus on longer-term planning.

Norming

In the norming stage, the IAM function is established with a dedicated team and established lines of responsibility. At this stage, the team is working productively together. The Program Manager may identify a need to expand organizational collaboration to an extended set of corporate members, including physical security, legal, privacy, human resources, information technology, and compliance offices. This comprehensive set of members may create a governance body or steering committee to help plan target state or organizational support to increase the return on investment of identity systems. For example:

- Collaborate with human resources to support remote identity proofing.
- Collaborate with physical security to integrate physical access control decisions with enterprise access management tools.
- Collaborate with the compliance office to automate compliance reporting.

An organization may go into the performing stage or circle around based on organizational needs and direction. Identity is a critical component of enabling efficient business processes but also an area of organizational risk. Program managers may need to adapt to new initiatives such as cloud services migration or zero trust architecture.

Conclusion

Organizations need an IAM workforce framework to ensure they hire and train their identity workforce. The most prevalent cybersecurity attack vectors are identity-based. This article introduced an IAM workplace planning model based on TKS aligned with a large organization's IAM enterprise architecture. It further aligned those tasks with how a typical organization identifies and staffs an IAM workforce. An organization can use the competency model to define consistent IAM roles across organizations or tailor them to fit their needs.

Author Bio

Kenneth Myers is a doctoral candidate with Marymount University and an Information Security IT Specialist with the U.S. General Services Administration. Reach him at kmm57090@marymount.edu or https://idmken.github.io.

Additional Reading

OMB. (2019). *Enabling Mission Delivery through Improved Identity, Credential, and Access* Management. Retrieved from OMB: https://www.whitehouse.gov/wp-content/uploads/2019/05/M-19-17.pdf.

Sharma, A., Sharma, S., and Dave, M. (2015). Identity and Access Management - A Comprehensive Study. 2015 International Conference on Green Computing and Internet of Things, 1, 1481-1485. https://doi-ieeecomputersociety-org.proxymu.wrlc.org/10.1109/ICGCIoT.2015.7380701.

Schneider, F. B., & Mulligan, D. K. (2011). A Doctrinal Thesis. *IEEE Security & Privacy Magazine, 9(4)*, 3–4. Retrieved from https://doi.org/10.1109/msp.2011.76.

NIST. (2021c). *Glossary - Phishing*. Retrieved from Computer Resource Security Center: https://csrc.nist.gov/glossary/term/phishing

NSA. (2020). *Detecting Abuse of Authentication Mechanisms.* Retrieved from Cybersecurity Advisory: https://media.defense.gov/2020/Dec/17/2002554125/-1/-1/0/AUTHENTICATION_MECHANISMS_CSA_U_OO_198854_20.PDF.

Reiner, S. (2020, 12 29). *Golden SAML revisited: The solorigate connection*. Retrieved from CyberArk Blog: https://www.cyberark.com/resources/threat-research-blog/golden-saml-revisited-the-solorigate-connection

Tan, Y., Li, W., Yin, J., & and Deng, Y. (2020). A universal decentralized authentication and authorization protocol based on Blockchain. *2020 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, (pp. 7-14).

Li, W., & Mitchell, C. J. (2020). User Access Privacy in OAuth 2.0 and OpenID Connect. *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, (pp. 664–672).

ACM, IEEE, AIS, S., & IFIP. (2017). *Cybersecurity Curricular Guideline.* Retrieved from CSEC 2017: https://cybered.hosting.acm.org/wp/.

ISC2. (2021). *CISSP – the world's premier cybersecurity certification*. Retrieved from ISC2: https://www.isc2.org/Certifications/CISSP.

CompTIA. (2021). *Security+ (plus) certification*. Retrieved from CompTIA IT certifications: https://www.comptia.org/certifications/security.

University of Bristol . (2020). *CyBOK Version 1.0.* Retrieved from Cybersecurity Body of Knowlege: https://www.cybok.org/knowledgebase/.

IDPro's Body of Knowledge. (2022). Retrieved from IDPro: https://idpro.org/body-of-knowledge/.

Rose, S., Borchert, O., Mitchell, S., & & Connelly, S. (2020). *Zero trust architecture.* Retrieved from NIST Special Publication: https://doi.org/10.6028/nist.sp.800-207.

NIST. (2018, April 16). *Framework for Improving Critical Infrastructure Cybersecurity.* Retrieved from NIST Cybersecurity Framework: https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf.

DHS. (2018). *PRIVMGMT: The First Step Toward CDM Phase 2 Capabilities.* Retrieved from Continuous Diagnostic and Mitigation (CDM): https://us-cert.cisa.gov/sites/default/files/cdm_files/FNR_CPM_OTH_NovWebinarSlides.pdf.

Kim, K., Smith, J., Yang, T. A., & Kim, D. J. (2018). An Exploratory Analysis on Cybersecurity Ecosystem Utilizing the NICE Framework. *2018 National Cyber Summit (NCS)*. Published. https://doi.org/10.1109/ncs.2018.00006.

Cabaj, K., Domingos, D., Kotulski, Z., & Respício, A. (2018). Cybersecurity education: Evolution of the discipline and analysis of master programs. *Computers & Security*, 24-35.

Bicak, A., Liu, M., & Murphy, D. (2015). Cybersecurity Curriculum Development: Introducing Specialties in a Graduate Program. *Information Systems Education Journal (ISEDJ)*, 99-110.

Hoag, J. (2013). Evolution of a cybersecurity curriculum. *Proceedings of the 2013 on InfoSecCD '13 Information Security Curriculum Development Conference - InfoSecCD '13*, (pp. 94-99).

Ran, F. X., & Sanders, J. (2020). Instruction quality or working condition? The effects of Part-Time faculty on student academic outcomes in community college introductory courses. *AERA Open*.

Furnell, S. (2020). The cybersecurity workforce and skills. *Computers and Security*, 100.

Gordon, A. (2016). The Hybrid Cloud Security Professional. *IEEE Cloud Computing, 3*(1), 82-86.

CIISec. (2019). *CIISec Roles Framework, Version 0.3.* Retrieved from Charted Institute of Information Security: https://www.ciisec.org/CIISEC/Resources/Capability_Methodology/Roles_Framework/CIISEC/Resources/Roles_Framework.aspx.

NIST. (2021a). *Glossary - Credential*. Retrieved from Computer Security Resource Center: https://csrc.nist.gov/glossary/term/credential.

Endnotes

[i] Flanagan (Editor), H., (2021) "Terminology in the IDPro Body of Knowledge", *IDPro Body of Knowledge* 1(8). doi: https://doi.org/10.55621/idpro.41.

[ii] Gartner. (2021). *Gartner Glossary*. Retrieved from Gartner: https://www.gartner.com/en/information-technology/glossary/identity-and-access-management-iam.

[iii] NIST. (2021b). *Glossary - ICAM*. Retrieved from Computer Security Resource Center: https://csrc.nist.gov/glossary/term/Identity_Credential_and_Access_Management

[iv] OPM. (2015). *Guidance for Identifying, Addressing and Reporting Cybersecurity Work Roles of Critical Need.* Retrieved from CHCOC: https://chcoc.gov/sites/default/files/Attachment%20to%20Memo%20-%20Guidance%20for%20Identifying%20Addressing%20Reporting%20Cyb.._.pdf.

[v] Petersen, R., Santos, D., Smith, M., Wetzel, K., & Witte, G. (2020, November). *National Initiative for Cybersecurity Education (NICE) Cybersecurity Workforce Framework.* Retrieved from Computer Security Resource Center: https://csrc.nist.gov/publications/detail/sp/800-181/rev-1/final

[vi] Nickel, J. (February 2019). Mastering Identity and Access Management with Microsoft Azure. Packt Publishing.

[vii] Martinelli, S., Nash, H, and Topol, B. (December 2015). Identity, Authentication, and Access Management in OpenStack. O'Reilly Media.

[viii] Ramey, K. (December 2016). Pro Oracle Identity and Access Management Suite. Apress.

[ix] Schwartz, M. and Machulak, M. (December 2018). Securing the Perimeter. Apress

[x] Verizon Enterprise. (2021). *2021 Data Breach Investigations Report.* Retrieved from Verizon Enterprise: https://enterprise.verizon.com/resources/reports/dbir/2021/

[xi] Grassi, P., Garcia, M., & Fenton, J. (2017). *800-63-3; Digital Identity Guidelines.* Retrieved from NIST Special Publication: https://doi.org/10.6028/NIST.SP.800-63-3

[xii] GSA. (2020). *Federal ICAM Architecture*. Retrieved from FICAM Playbooks: https://playbooks.idmanagement.gov/

[xiii] GSA. (2021). Identity Governance Framework: https://playbooks.idmanagement.gov/docs/playbook-identity-governance-framework.pdf

[xiv] Stein, J. (n.d.). *Using the Stages of Team Development*. Retrieved from MIT Human Resources: https://hr.mit.edu/learning-topics/teams/articles/stages-development