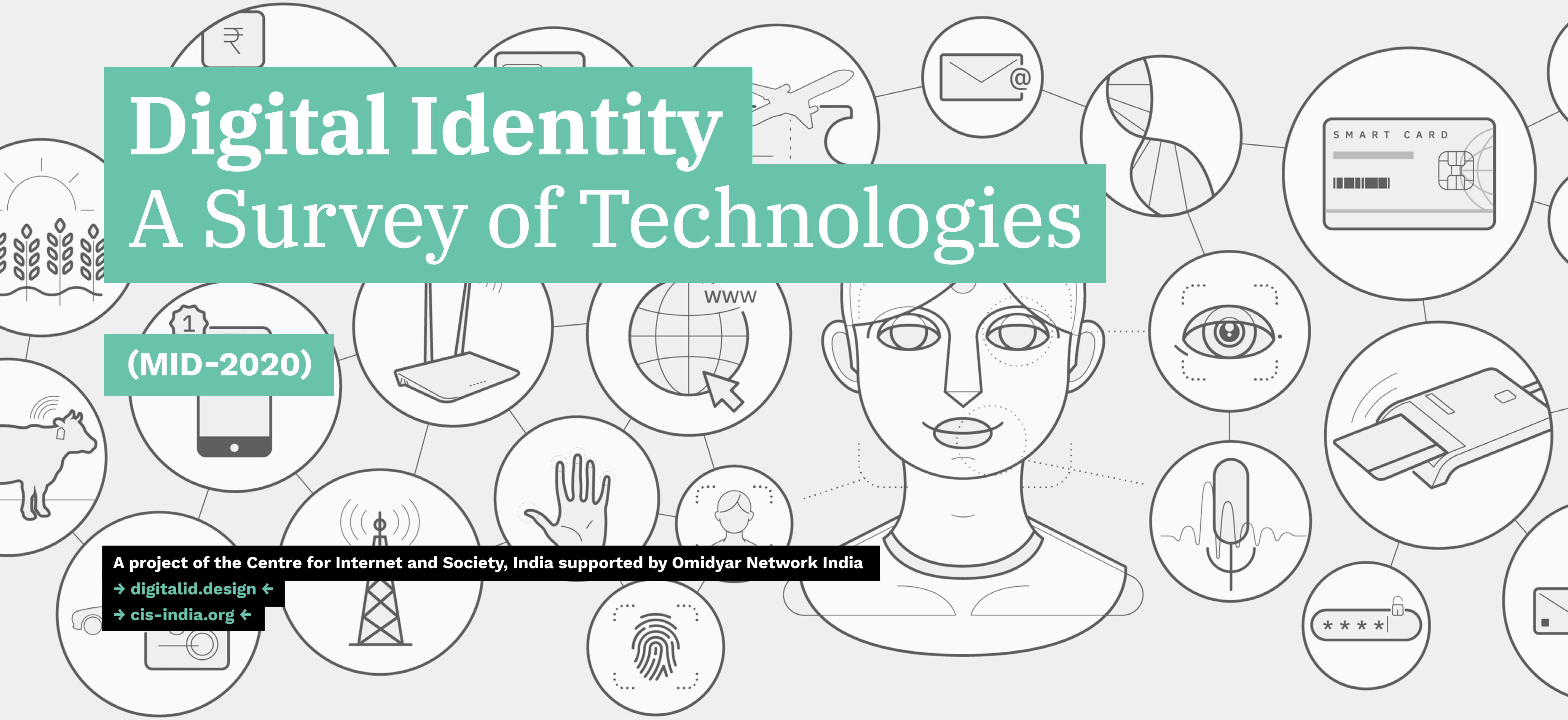
Digital Identity
A Survey of Technologies

(MID-2020)



A project of the Centre for Internet and Society, India supported by Omidyar Network India
→ digitalid.design ←
→ cis-india.org ←

CONCEPTUALIZATION
Amber Sinha and Divyank Katira

RESEARCH & WRITING
Divyank Katira

EDITING
Anubha Sinha and Akash Sheshadri

RESEARCH ASSISTANCE
Jahnavi Selvakumar and Swati Gautam

DESIGN
Pooja Saxena

TYPESETTING & COVER ILLUSTRATION
Akash Sheshadri

THE CENTRE FOR internet & society

INTRODUCTION

There is a growing trend, in both public and private sectors, to apply digital technologies to the processes of identification of individuals, authentication of their established identity, and allowing authorisation on their behalf, in the context of governance and provisioning goods and services. Through the use of software and hardware, these digital technologies attempt to create virtual models of the complex, nuanced notions of identity in our analog real world, and allow for individuals and organisations to transact based on them. For narrowly defined tasks that are within the scope of these virtual representations, technology-assisted processes offer a higher level of efficiency and accuracy than solely human-operated ones.

However, these advancements also bring with them unique risks. Individuals whose identities are not adequately represented by their virtual models may face exclusion, and technical faults may impede access to essential services, particularly when the use of technology is made mandatory. Digital processes also generate, distribute, gather, and store more information about individuals than their analog counterparts. This brings risks related to the security of personal information handled by these systems and the privacy of the individuals participating in them.

In this survey, we evaluate technologies used to compose digital identity processes and systems with an emphasis on preventing or reducing the associated risks. It is divided into four parts:

1. **Identification and Authentication Factors** are the different kinds of information about individuals that are used by digital identity systems to identify them and subsequently authenticate their identity;

2. **Identity Artifacts** are physical or digital credentials assigned to individuals;

3. **Digital Identity Workflows** are common configurations adopted by digital ID systems in relation to how identification, authentication, and authorisation are carried out within them; and

4. **Digital Identity Standards** which have emerged from public and private sector entities working on these systems.

These are the key considerations in designing technology-assisted processes for functional or foundational digital identity systems.

EVALUATION CRITERIA

We look at the core technological choices made in a digital identity system and evaluate them on the basis of the following criteria, where relevant:

SECURITY

This criterion relates to how well the technology can resist malicious use, ensure the confidentiality and integrity of data and general availability of the system.

PRIVACY

This criterion relates to an individual's agency to withhold and selectively reveal information about themself, and their right to understand and control how their data and metadata is used.

ACCURACY

This criterion evaluates how well the given technology can identify an individual under ideal conditions.

SCALABILITY

This criterion considers technical barriers to widespread adoption including data and bandwidth scalability, and computational resources required for operation.

MATURITY

This criterion questions how long a given technology has been in use, what standards exist and how well implementations can interoperate.

COST

This criterion evaluates the cost-effectiveness of the technology.

We assign ratings to technologies which are relative to similar technologies in the same category or sub-category, where applicable.

IDENTIFICATION AND AUTHENTICATION FACTORS

Identification and Authentication Factors are pieces of information about an individual that are used to identify them and subsequently authenticate their identity.

They can be categorised in the following five categories:

A. Biometric Factors

Biometric identification involves an enrolment process during which biometric information about the individual is collected. To ensure uniqueness, the information may be compared with all previously collected records of other individuals in the database (called deduplication). Authentication involves scanning of biometric information and matching it against the previously collected record of an individual. There are two kinds of biometric factors:

1. **Inherence factors** rely on the use of physical attributes of the individual.
    a. Fingerprint Scanning
    b. Facial Recognition
    c. Iris Scan
    d. DNA Profiling
2. **Behavioural factors** rely on measurements of behavioural aspects of an individual.
    a. Voice Recognition
    b. Gait Recognition

SECURITY ▼ LOW

Biometric factors form weak authentication mechanisms. They are immutable and, in most cases, publicly visible. This makes them impossible to change in case of a data breach, and susceptible to forgery. As such, their utility is limited to use as secondary factors for offline authentication, such as in end-user devices like smartphones, security keys, and smartcards, where biometric data is stored and matched on-device. Creating databases of biometric information for online authentication is ill-advised and not sustainable, as every breach diminishes the utility of using biometric factors for authentication as a whole.

PRIVACY ▼ LOW

Biometrics pose threats to privacy, particularly factors like facial and gait recognition — which can be deployed without an individual's consent or knowledge. Their immutable nature makes it hard to place meaningful limits on how they will be used on the long-term.

ACCURACY ▲ HIGH

Depending on the technologies chosen, a combination of biometric factors can be a fairly accurate way of identifying individuals. However, no biometric technology is fully accurate and relying on them as a sole means of identification leads to exclusion.

SCALABILITY

Identification is a more computationally expensive process as it involves comparing a collected biometric factor with all previously collected records in a database (called 1:N matching). Whereas authentication entails a simple comparison with a previously collected record (called 1:1 matching).

MATURITY

Variable depending on the biometric technology in use.

COST

Requires dedicated hardware sensors for biometric collection and specialised software for identification and authentication. They are more cost-effective for use in online authentication in national ID systems, where sensors only need to be installed at points of collection and service delivery, however, such use is not recommended due to security issues.

B. Self-asserted Factors

These factors are independently verified by the identifying entity. They do not require verification by a third party authority but rely on factors that the individuals being identified can verify themselves. The features of self-asserted factor technologies vary on the basis of technological and information design option in use.

1. Email Verification
2. SIM Verification
3. Web-tracking

C. Third-Party Attested Factors

Third-Party Attested Factors are issued by a third-party that is different from the identifying entity and indicates a trust boundary between the two. The features of third-party attested factors vary on the basis of technological and information design option in use.

1. Document Verification
2. Credit Agency Reference Data

D. Possession Factors

In case of possession factors, the identity credential assigned during Identification and used during Authentication is an object that is supposed to be in the possession and control of the Individual. The features of possession factors may vary on the basis of technological and information design option in use.

1. SMS-based One-time-password (OTP)
2. App-based OTP
3. Public Key Infrastructure (PKI)

E. Knowledge Factors

In the case of knowledge factors, authentication is carried by testing for information that the individual is expected to know. The features of knowledge factors may vary on the basis of technological and information design option in use.

1. Password/PIN
2. Secret Questions

Fingerprint Scanning [0] [1] [2]

Minute physiological features present on the surface of the skin such as endings and bifurcations of friction ridges and the distance between them are used to identify individuals.

There are four types of fingerprint scanners:

Optical Scanners use visible light to effectively take a photograph of the finger. These are known to be low-cost.

Capacitive Scanners pass small amounts of electrical current to the finger to detect patterns on the surface. These are most commonly found in smartphones.

Ultrasonic Scanners bounce ultrasonic sound waves off the finger to generate a three-dimensional representation of it. These are considered to be highly accurate and do not require contact.

Thermal Scanners detect variations in temperature in fingertip ridges and valleys. Requires finger to be moved over the sensor to operate.

USAGE

Fingerprint scanners have been used for identification and authentication in national ID schemes, border and physical access control, and crime forensics.

They are found embedded in consumer electronics (mobile phones, personal computers, security keys, and smartcards) where they are used for offline authentication.

ACCURACY

Friction ridges on fingers can be damaged or obscured. They can be worn off with age or for people from certain professions. They can be inaccessible due to disease or injury [3] and, in some cases, completely absent. Otherwise, these systems are fairly accurate; the false positive rate is low.

SECURITY ▼ LOW

Fingerprint scanning is susceptible to forgery. Fingerprints can be copied from photographs or from surfaces the individual comes in contact with. These copies can be used to create artifacts to mimic fingerprints.

Many or all fingerprints are collected and used to mitigate these issues.

SCALABILITY ▲ HIGH

Widely deployed in consumer electronics and large scale national digital identity schemes such as Aadhaar.
No known scalability challenges exist.

MATURITY ◁▷ MEDIUM

Widely deployed in consumer electronics and national digital identity schemes such as Aadhaar. Considered relatively mature.

COST ▽ LOW TO MEDIUM

Even though dedicated hardware is required, the cost of fingerprint sensors has been falling due to widespread consumer adoption.

[0] *https://www.ncsc.gov.uk/collection/biometrics/fingerprint*
[1] *https://www.bayometric.com/biometric-devices-cost/*
[2] *https://www.androidauthority.com/how-fingerprint-scanners-work-670934*
[3] *https://www.hindawi.com/journals/bmri/2012/626148*

Facial Recognition [0] [1] [2]

Facial features can be used to recognize individuals in inputs (photographs, videos, or real-time feeds) through the use of both visible light and infrared waves.

The current state-of-the-art of this technology relies on machine learning techniques such as deep learning. Broadly, in these methods, inputs are treated as a matrix of pixels and passed through a series of statistical units termed “artificial neurons”. These units output numerical weights based on factors such as pixel colour and density. Many layers of such interconnected units are provided with several labelled inputs which form a heuristic through which the weights are adjusted. This is called the training phase and through this, the units collectively “learn” to recognize features such as edges and shapes corresponding to given labels. When a new input is given, the units will output a high weight when they encounter features that were present in the training data. These weights are normalized and used to produce labels with a confidence number (for e.g. this input is 95% a zebra, and 60% a horse).

These systems also typically include a liveness check such as blinking or a series of head movements to ensure the subject being identified is a real person.

USAGE

Automated facial recognition has seen use in public surveillance systems, border control and law enforcement for identification. It is also used in consumer electronics for authentication of users.

ACCURACY

Affected by lighting conditions, obscured faces. Relies highly on quality of the training data provided. A study has shown that such systems are prone to algorithmic bias. Gender, skin colour, and ethnicity can lead to a significant drop in accuracy [3][4].

SECURITY ◁▷ MEDIUM

Forgery is possible through more sophisticated techniques such as 3D-printed face masks [5].

SCALABILITY ◁▷ MEDIUM

Identification requires large amounts of training data. Computational requirements are relatively high.

MATURITY ▼ LOW

The application of automated facial recognition is relatively recent and has not seen wide scale deployment outside of consumer electronics.

COST ▲ HIGH

Requires dedicated hardware for both capture (cameras) and processing of data (GPUs). Large amounts of training data required for identification.

[0] *https://www.ncsc.gov.uk/collection/biometrics/face*
[1] *https://machinelearning.apple.com/2017/11/16/face-detection.html*
[2] *https://www.thalesgroup.com/en/markets/digital-identity-and-security/government/biometrics/facial-recognition*
[3] *https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html*
[4] *http://proceedings.mlr.press/v81/buolamwini18a.html*
[5] *https://thehackernews.com/2017/11/iphone-face-id-unlock-hack.html*

Iris Scan

An iris scanner captures images of either one or both human irises with high resolution to compare and match it with the existing iris pattern of an individual saved in the database.

The biometric iris scanner operates based on two processes:

Enrollment 2 digital photographs (one using visible light and the other using infrared light) are taken and analyzed by a computer which identifies around 240 unique features (5 times more points of comparison used by fingerprint systems) [0].

These features are then converted into a simple 512 digit number called an IrisCode which is stored with details such as name and are stored in a database. An automatic process which takes a maximum of two minutes.

Verification The eye is photographed again from which the IrisCode is extracted and compared with the existing entries in the database.

Contactless, fast and renowned for accuracy. Can also operate at long distances. Different from retinal scanning.

USAGE

Iris scans are used in public ID schemes, and border and physical access control for both identification and authentication purposes. They are also found in some consumer electronics for authentication.

ACCURACY

At least ten times more accurate than fingerprinting [1]. Since the Iris is naturally protected by the cornea, it remains unchanged for several decades. Changes have occurred with surgeries such as cataract. Affected by eye diseases [2]. Less accurate for children between ages 1-4. Accuracy of scanners can be affected by lighting.

SECURITY ▼ LOW

Forgery is possible through a high resolution image of an iris. Scanners have also been tricked by images generated from compromised digital codes of stored irises.

SCALABILITY ▲ HIGH

Widely deployed at large scales. No known scalability challenges exist.

MATURITY ◁▷ MEDIUM

Has been a reliable security measure since 2001 (Amsterdam Airport Schiphol). Widely deployed in large scale national identity schemes like Aadhaar and is considered relatively mature.

COST ▽ LOW TO MEDIUM

Dedicated hardware required but cost decreases with increase in use.

[0] *https://www.explainthatstuff.com/how-iris-scans-work.html*

[1] *https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/100825/feasibility_study031111_v2.pdf*

[2] *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2659699/*

DNA Profiling [0]

DNA profiling is the process of uniquely identifying an individual through their genetic characteristics.

DNA profiling begins with a sample of an individual's DNA (typically called a "reference sample"). A reference sample is then analyzed to create the individual's DNA profile using one of a number of techniques, which include PCR (Polymerase Chain Reaction - means to create billions of exact copies of a specific region of the genome) and STR (Short Tandem Repeat markers - the frequency of short repeated sequences of DNA is observed) analysis. The DNA profile is then compared against another sample to determine whether there is a genetic match [1].

Rapid DNA Analysis is a 2 hour identification process developed by the FBI which consists of a fully automated (hands free) process of developing a DNA profile from a reference sample buccal (cheek) swab without human intervention [2].

USAGE

DNA profiling is used in crime forensics for identification purposes. It is also used in genealogy and medical research.

ACCURACY

Only fails in the case of monozygotic twins. False positive rate is extremely low [3].

SECURITY ◁▷ MEDIUM

Despite best efforts, DNA profiles will always be observed in the laboratory environment and there is danger of cross contamination of samples. DNA samples can be inadvertently left on surfaces.

SCALABILITY ▼ LOW

Applied in niche use cases and has not seen large scale use. Scalability of these techniques is unproven.

MATURITY ◁▷ MEDIUM

It has been applied widely in forensics with evolving techniques, recent improvements in Rapid DNA Analysis show promise in greater improvements and adaptability.

COST △ MEDIUM TO HIGH

Differs with the technique used. AFLP costs low [4].

[0] *https://www.sciencedirect.com/topics/neuroscience/dna-profiling*
[1] *https://www.biometricupdate.com/201309/explainer-dna-and-dna-profiling*
[2] *https://www.fbi.gov/services/laboratory/biometric-analysis/codis/rapid-dna*
[3] *https://strbase.nist.gov/pub_pres/Vallone_DOC_talk_March12_2015.pdf*
[4] *https://www.ncbi.nlm.nih.gov/probe/docs/techaflp*

Voice Recognition [0] [1]

Voice recognition software records an individual's voice as the input, converts the analog signal into a digital signal, and extracts a voiceprint for identification and authentication.

It differs from speech recognition as it focuses on the features of the user's voice as opposed the identifying the words spoken.

Voice recognition is the process by which a computer identifies spoken words and splits them into text dependent (keywords and phrases) and text independent (measures minutiae of the voice) [2]. The voiceprint obtained includes more than 100 unique physical and behavioural characteristics of each individual, such as length of the vocal tract, nasal passage, pitch, accent and so on [3].

Voiceprints can be measured passively as an individual speaks naturally in conversation, or actively, if they are made to speak a passphrase. According to a survey by Unisys, voice recognition is the highest ranked biometric measure preferred by users due to its convenience and familiarity [4].

USAGE

Voice recognition is used in consumer electronics, call centres, telephones and internet transactions, IVR-based based systems for both identification and authentication.

ACCURACY

Affected by poor-quality voice samples (variability in a speaker's voice due to illness, mood), background noise, changes in the sample collection technology (digital vs. analog, upgrades to circuits and microphones, etc).

SECURITY ▼ LOW

Vulnerable to attacks that use cloned voice samples which can be obtained via sham phones calls and covertly captured recordings or publicly available voice samples.

SCALABILITY ▲ HIGH

Has seen deployment in voice-based applications. No known scalability challenges exist.

MATURITY ◁▷ MEDIUM

Used for authentication in call centres, telephones and internet transactions, etc. Steady updates in combining with speech recognition in consumer electronics to improve accuracy.

COST ▽ LOW TO MEDIUM

Requires dedicated hardware and software.

[0] *https://www.business.att.com/learn/tech-advice/how-secure-is-voice-recognition-technology.html*

[1] *https://www.plumvoice.com/resources/blog/voice-biometrics/*

[2] *https://www.researchgate.net/publication/289299616_Biometric_Voice_Recognition_in_Security_System*

[3] *https://www.theguardian.com/money/2018/sep/22/voice-recognition-is-it-really-as-secure-as-it-sounds*

[4] *https://www.securityinformed.com/insights/co-3108-ga.4100.html*

Gait Recognition [0] [1]

Gait recognition is based on the coordinated cyclic motions that result in human locomotion (walking, running, climbing stairs). Features are automatically extracted and are used to identify and authenticate an individual.

It consists of a preprocessing phase which includes background subtraction and body silhouette extraction, eventually identifying the Degree of Freedom points, which generally corresponds to body joints, in order to track an individual's gait [2]. The number of pixels in the foreground reaches a maximum when the two legs are farthest apart (full stride stance) and drop to a minimum when the legs overlap (heels together stance) for each silhouette. A counter is run to calculate the number of images that form a gait cycle [3].

It is non-contact, non-invasive biometric identification which is hard to imitate and can be used without an individual's consent. It also offers potential for recognition at a distance or at low-resolution. There are two main types :

1. An analysis of video samples of a subject's walk and the trajectories of joints and angles. A mathematical model of the motion is created, and is subsequently compared against other samples in order to determine identity.
2. A radar system, which records the gait cycle that the various body parts of the subject creates, which is then compared with other samples [4].

USAGE

Outside of some initial deployment in public surveillance systems for identification, gait recognition software has not seen much use yet.

ACCURACY

Confounding factors such as terrain, injury, footwear, fatigue, athletic training, personal idiosyncrasies, etc. High false positive rates with large databases.

SECURITY ▲ HIGH

Difficult to mimic or forge.

SCALABILITY ▽ LOW TO MEDIUM

Mostly works only in controlled scenarios. Different technologies used for data acquisition in different conditions (thermal/infrared). Synchronization in the case of multiple sources is still an issue and requires additional computational costs.

MATURITY ▼ LOW

There have been significant improvements to improve the technology (algorithm analysis), however, due to restricted applications which require controlled scenarios, it has not seen wide scale deployment.

COST ▲ HIGH

Hardware cost is high due to different technologies used (different types of cameras, radars, sensors). Training data required is very large.

[0] *https://www.intechopen.com/books/motion-tracking-and-gesture-recognition/gait-recognition*
[1] *https://apnews.com/bf75dd1c26c947b7826d270a16e2658a*
[2] *https://www.sciencedirect.com/topics/engineering/gait-recognition*
[3] *https://www.biometricupdate.com/201311/explainer-gait-recognition*
[4] *https://pages.cpsc.ucalgary.ca/~boyd/papers/biometric-summerschool.pdf*

Email Verification

An individual is sent a verification email which contains a link or a code. Demonstrating access to these is considered proof that the individual has access to the email address provided.

USAGE

An individual is identified by their email address. This is a standard practice across the web (e-commerce, blogs, forums).

ACCURACY ▼ LOW

Not a strong identification factor as email addresses can be anonymous.

PRIVACY ▲ HIGH

Maintains privacy as email addresses can be created anonymously.

SECURITY ◁▷ MEDIUM

Depends on security practices of email provider. Susceptible to phishing attacks, poor password habits, lack of multi-factor authentication.

SCALABILITY ▲ HIGH

Proven at internet scale.

MATURITY ▲ HIGH

Has been in use for a long time for a variety of purposes.

COST ▼ LOW

The cost of sending an email is negligible. Assumes individuals being identified possess email capable devices.

SIM Verification [0]

Possession of a SIM card is asserted through a phone call or SMS that conveys a verification code. Demonstrating access to the code is considered proof.

USAGE

An individual is identified by their phone number. Such identification is used in both public and private sectors.

See One-time Passwords (OTPs) for authentication.

ACCURACY ◁▷ MEDIUM

Stronger identification factor than email as the issuance of SIM cards is regulated by some form of KYC requirements in 140 countries [1].

It is susceptible to the use of virtual numbers issued by online services that allow anonymous registration.

PRIVACY ◁▷ MEDIUM

Less private than email as phone numbers are typically linked to an individual's identity.

SECURITY ▼ LOW

Depends on the security practices of the cell service provider. SIM cards can be shared and transferred between people, and are vulnerable to sim-swapping attacks.

SCALABILITY ▲ HIGH

Proven at internet scale, and in large scale digital ID systems such as Aadhaar.

MATURITY ◁▷ MEDIUM

Has gained popularity with the usage of multiple-factor authentication where it typically serves as a second factor.

COST ▼ LOW

The cost of sending an SMS is negligible. Assumes individuals being identified possess cell phones.

[0] *https://dl.acm.org/doi/10.1145/3025453.3025961*

[1] *https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2018/02/Access-to-Mobile-Services-and-Proof-of-Identity.pdf*

Web Tracking [0] [1] [2] [3]

A unique, persistent identifier is created or derived and used to track individuals across the internet for behavioural analytics and targeted advertising.

The common methods are:

Tracking Cookies Cookies are a feature in web browsers that allow websites to store information on a computer. When a website different from the one visited does this, it is called a third-party cookie. This allows for third-parties to embed content (such as a like or share button) on other sites. This is used to store a persistent identifier to identify an individual across the internet.

Browser and Device Fingerprinting Minute variances in an individual's web browser such as plugins and fonts installed, and device characteristics such as screen size and processor capabilities are used in concert to uniquely identify an individual.

USAGE

This is a standard and wide-spread identification method used by the online advertising industry.

ACCURACY ▲ HIGH

The unique identifier provides a highly accurate method for identifying individuals with high levels of assurance.

PRIVACY ▼ LOW

This form of identification is not usually linked to personally identifiable information (de-identified) but is used to collect highly sensitive data about an individual which can potentially be re-identified or combined with other identification mechanisms.

SECURITY ◁▷ MEDIUM

The identities derived through tracking cookies and device fingerprinting cannot be easily forged.

SCALABILITY ▲ HIGH

Proven at internet scale.

MATURITY ◁▷ MEDIUM

Forms the basis for many business models on the internet. It has been in use and evolving for a long time.

COST ◁▷ MEDIUM

Prices are set by a market that is monopolized by a few companies that are large enough to have presence on many internet properties for this method to be effective.

[0] *https://privacy.net/stop-cookies-tracking*
[1] *https://panopticlick.eff.org/about*
[2] *https://samy.pl/evercookie*
[3] *https://www.eff.org/deeplinks/2018/06/gdpr-and-browser-fingerprinting-how-it-changes-game-sneakiest-web-trackers*

Document Verification

Physical identity documents such as passports, driver's licenses, voter IDs, birth certificates, or other ID cards that are issued by a third-party, typically a government, are used for both identification and authentication.

Physical documents include security features such as holograms, microprinting, ghost images, tactile impressions, and rainbow printing to prevent forgery [0]. Verification techniques include:

Machine-readable Zone (MRZ) This is a standardized format [1] [2] in which text is formatted in a specific way to facilitate optical character recognition (OCR). It is widely used in passports and cards [3].

Computer Vision Machine learning techniques are used to validate documents. Commercial products exist that claim high accuracy (98.7%) in detecting forged documents and support for documents from 195 countries [4].

Manual Verification The information present on the document as well as the security features are manually checked, sometimes with the aid of a magnifying glass or UV light.

USAGE

Document verification is used for border control, licensing, and personal identification in both public and private sectors.

ACCURACY ◁▷ MEDIUM

Machine-based techniques are more accurate than manual verification. For high assurance requirements, the document can even be cross-checked with the issuing authority.

PRIVACY ◁▷ MEDIUM

Physical documents typically have more information printed on them than is necessary for identification or authentication. They are typically issued for a single or narrow set of purposes and the tangible nature of the credential makes it clear what information is being shared.

SECURITY ◁▷ MEDIUM

With sufficient security features present on the document, it can be hard to forge. Physical documents may reveal more information than necessary to the relying party.

SCALABILITY ▲ HIGH

Physical identification documents are widely used, their scanning and verification do not present any scalability challenges.

MATURITY ▲ HIGH

Been in use for a long time. Machine learning based computer vision techniques are relatively recent.

COST ▽ LOW TO MEDIUM

Machine-based validation requires dedicated hardware and software.

[0] *http://documents.worldbank.org/curated/en/199411519691370495/Technology-Landscape-for-Digital-Identification.pdf*

[1] ICAO Document 9303 *https://www.icao.int/publications/pages/publication.aspx?docnum=9303*

[2] ISO/IEC 7501-1:2008 *https://www.iso.org/standard/45562.html*

[3] *https://en.wikipedia.org/wiki/Machine-readable_passport*

[4] *https://onfido.com/document-verification*

Credit Agency Reference Data [0] [1]

Credit bureaus (also called credit agencies or credit information companies) are commercial entities that purchase, aggregate, and sell information related to individuals' borrowing and spending habits.

This information is provided in two forms:

A numeric **credit score** quantitatively derived from a credit history.

A **credit report** which includes more detailed credit information and PII of an individual.

These are primarily used to gauge the creditworthiness of an individual. This information is provided to lenders, landlords, employers, and other entities deemed appropriate by credit bureaus and underlying regulation [2].

Credit bureaus also provide identity verification services [3] [4] based on the information they collect about individuals. Experian, Equifax, and Transunion are key players in this industry globally.

USAGE

Individuals are identified based on financial records that credit agencies maintain on them.

ACCURACY ▲ HIGH

The credit data used for identification is considered fairly accurate as it is provided by the financial institutions an individual is associated with.

PRIVACY ▼ LOW

Financial information from various third-parties is aggregated, often under uninformed, coercive notions of consent.

SECURITY ◁▷ MEDIUM

Centralised collections of highly sensitive personal information held by credit agencies are prone and have been subject to massive breaches and other forms of fraudulent access by unauthorized entities [5] [6] leading to identity theft.

SCALABILITY ▲ HIGH

These techniques have been deployed at nation-scale and no known scalability challenges exist.

MATURITY ◁▷ MEDIUM

Not widely used in countries with underbanked populations. Separate credit agencies maintain different scores with disparate data sources that do not interoperate.

COST ◁▷ MEDIUM

There are costs associated with trading and aggregating the personal information used for this identification technique.

[0] *https://en.wikipedia.org/wiki/Credit_bureau*
[1] *https://www.bankbazaar.com/credit-score.html*
[2] *https://www.equifax.com/personal/education/credit/report/who-is-allowed-to-access-your-credit-report/*
[3] *https://www.transunion.com/solution/id-verification*
[4] *https://www.experian.in/prove-id*
[5] *https://epic.org/privacy/data-breach/equifax/*
[6] *https://www.tripwire.com/state-of-security/security-data-protection/4-credit-bureau-data-breaches-predate-2017-equifax-hack/*

SMS-based One-Time-Password (OTP)

An SMS one-time password (OTP) is a dynamically generated numeric or alphanumeric string of characters that is valid for only one login session or transaction.
It is generated by the authenticating server and delivered to the individual via text message.

USAGE

SMS-based OTPs are used for authentication in public and private sector services, typically as a secondary factor.

PRIVACY ▼ LOW

Phone numbers, that are typically linked to an individual, are revealed to the authenticating party.

SECURITY ▼ LOW

Susceptible to wireless interception, mobile phone trojans and, SIM swap attacks. Several attacks against GSM and 3G networks have shown that confidentiality for SMS messages cannot necessarily be provided [0] [1].

SCALABILITY ▲ HIGH

Proven for internet scale deployments.

MATURITY ◁▷ MEDIUM

Has gained popularity with the usage of multiple-factor authentication where it typically serves as a second factor.

COST ▼ LOW

Assumes individuals being authenticated possess cell phones.

[1] *https://www.ijert.org/sms-based-one-time-password-vulnerabilities-and-safeguarding-otp-over-network*

[2] Collin Mulliner, Ravishankar Borgaonkar, Patrick Stewin, and Jean- Pierre Seifert, SMS-Based One-Time Passwords: Attacks and Defense, Springer-Verlag Berlin Heidelberg, LNCS 796, pp.150-159, 2013.

App-based One-Time-Password (OTP) [0] [1] [2]

An OTP is generated on an end-user device (smartphone application, key fob, etc.) independently of the server performing authentication i.e. the device may remain offline.

This is done by exchanging a secret key with the server beforehand during registration (in the form of a QR code, for example). This secret key is used to generate a random OTP through two methods that have been standardised [3] [4] at the Internet Engineering Task Force (IETF):

Time-based OTP (TOTP) A variable OTP is generated using the current time and the pre-shared secret key. This method requires time synchronisation with the server.

HMAC-based OTP (HOTP) A incremental counter is used as the variable input instead of time along with the pre-shared secret key. The counter may go out of sync with the server in this method.

Workarounds are available for the synchronisation issues and this method works quite accurately.

USAGE

App-based OTPs are used for authentication in web services, typically as a secondary factor.

PRIVACY ▲ HIGH

Allows an individual to be authenticated anonymously.

SECURITY ▲ HIGH

More secure than SMS Based OTPs.

SCALABILITY ▲ HIGH

Proven for internet scale deployments.

MATURITY ▲ HIGH

Well known. Open standards and implementations exist.

COST ▼ LOW

Slightly higher than SMS-based OTP which does not require a smartphone that can support applications.

[0] *https://github.com/google/google-authenticator/wiki*

[1] *https://www.freecodecamp.org/news/how-time-based-one-time-passwords-work-and-why-you-should-use-them-in-your-app-fdd2b9ed43c3/*

[2] *https://www.protectimus.com/blog/10-most-popular-2fa-apps-on-google-play*

[3] *https://tools.ietf.org/html/rfc6238*

[4] *https://tools.ietf.org/html/rfc4226*

Public Key Infrastructure (PKI)

Public Key Infrastructure is a fundamental cryptographic building block for digital credentials. It forms the basis for authentication flows where demonstrating possession of a private key is used to authenticate individuals. It is also used to produce and verify digital signatures.

It is an asymmetric encryption scheme, which means that different keys are used for encryption and decryption. There is a private key, which is kept secret and is used for digital signing and decryption, and a corresponding public key that is used for signature verification and encryption.

For example, Bob can use Alice's public key to encrypt a message meant for Alice, which she can decrypt with her private key. And Alice can digitally sign a message or claim with her private key which Bob can verify using her public key.

USAGE

Not directly used, but forms the basis for authentication in many digital ID schemes and web-based services. It is also the underlying technology for digital signatures.

PRIVACY ▲ HIGH

Allows an individual to be authenticated anonymously.

SECURITY ▲ HIGH

When sufficient key sizes and stable ciphers are chosen, this method can be highly secure. Key management (secure storage, rotation, revocation) is considered to be a challenge in this method.

SCALABILITY ▲ HIGH

Proven at internet scale — it forms the basis of the Transport Layer Security (TLS) protocol which is used to authenticate owners of websites. It is not an efficient choice for encrypting large amounts of data.

MATURITY ▲ HIGH

It has been in use for decades and mature standards and open-source implementations exist.

COST ▼ LOW

Compatible with low computational and storage requirements. Can be embedded in a smartcard/ security key.

Password/PIN

A unique string or number is chosen by an individual, knowledge of which is used to authenticate their identity.

USAGE

Passwords are used as a standard authentication mechanism on the internet.

PRIVACY ▲ HIGH

Allows an individual to be authenticated anonymously.

SECURITY ◁▷ MEDIUM

Passwords of sufficient length with safeguards such as limiting number of attempts to prevent guessing can be quite secure. However, the reliance of this mechanism on individuals makes it less secure. It is susceptible to social engineering attacks such as phishing and poor password practices such as re-use across different services and easy to guess passwords. The use of password managers, software that creates and remembers secure passwords, can mitigate some of these risks.

SCALABILITY ▲ HIGH

Proven at internet scale.

MATURITY ▲ HIGH

They have been in use since the early days of the internet.

COST ▼ LOW

As a knowledge factor, there is no cost associated with dedicated hardware or software for the use of passwords.

Secret Questions

A set of questions preset by an individual is used as a test for authentication.

USAGE

Secret questions are typically used for account recovery in case of a forgotten password/lockout, or as a secondary factor for authentication.

PRIVACY ◁▷ MEDIUM

The questions used sometimes contain unrelated private information.

SECURITY ▼ LOW

Sometimes the questions used are common knowledge, susceptible to social engineering attacks where a victim is tricked into revealing answers.

SCALABILITY ▲ HIGH

Proven at internet scale.

MATURITY ▲ HIGH

They have been in use since the early days of the internet.

COST ▼ LOW

As a knowledge factor, there is no cost associated with dedicated hardware or software for the use of secret questions.

DIGITAL IDENTITY ARTIFACTS

An identity artifact is a document or object, which can be both physical or digital, that is issued to an Individual at the end of the process of Identification, that facilitates in establishing their Identity. It usually includes a registration number assigned to the Individual.

Identity artifacts can contain attributes about an individual (either in writing or encoded digitally) or digital credentials (for authentication and digital signatures).

Some of these require dedicated hardware (card readers) to be installed at the point of service delivery to access the attributes and credentials stored in the artifact, while others can interface directly with personal electronics (phones, computers) over USB, Bluetooth, or NFC, making them useful for online service delivery to individuals.

USAGE

Identity artifacts are primarily used as a possession factor for authentication of individuals, typically along with a knowledge (PIN) or biometric (fingerprint) second factor to unlock the credentials digitally stored on the artifact. They are sometimes used as third-party attested factors for identification. The digital credentials stored on the artifact are also used for digital signing in some cases.

SECURITY

Limited by the physical security of the artifact. They can be stolen or lost. A second factor, in the form of a PIN, biometric, or manual photo verification is used to limit these risks. This is an appropriate use of biometrics, as they are stored and matched on the device as opposed to being collected in a central database.

PRIVACY

There may be private information printed on the artifact that may not be necessary for authentication. Digital artifacts can store many credentials, even one per-service. This allows individuals to remain anonymous or share the minimal set of attributes required to be authenticated. The WebAuthn standard features such privacy enhancements.

SCALABILITY ▲ HIGH

Scalability of identity credentials is generally high.

MATURITY

Depends on the technology chosen.

COST

Depends on the technology chosen.

The different digital identity artifacts are:

1. Microchip-based (Smart cards)
2. Contactless Cards
3. QR Code
4. Biometric System-on-Card (BSoC)
5. Security Keys
6. Smartphones & Computers

Microchip-based (Smart Cards)

An integrated circuit chip that is capable of processing and storing information is embedded in physical documents.

A PIN is used to decode the data present on the chip. A dedicated chip reader is required to access this information. Additional software may be required for decoding and verification of information.

USAGE

Smart cards are used for identification, authentication, and digital signing at physical points of service, where card readers are present. They are deployed in the payments industry, public ID schemes, and biometric passports.

SECURITY △ MEDIUM TO HIGH

Barring implementation flaws, chip cards are considered secure. The PIN can be observed, intercepted by a malicious party.

SCALABILITY ▲ HIGH

Over 10 billion were issued as of 2015 [0].

MATURITY ▲ HIGH

Widely deployed in payments, SIM cards for mobile phones, and national ID schemes. Interoperable standards exist [1].

COST ▽ LOW TO MEDIUM

Supported cards need to be issued, scanners and decoding software required.

[0] *https://technology.informa.com/582859/smart-card-ic-shipments-to-reach-128-billion-units-in-2020*

[1] ISO/IEC 7816 *iso.org/obp/ui/#iso:std:iso-iec:7816:-8:en*

Contactless Cards

Similar to smart cards, microchips can be powered from a distance and information (usually not PII) is transmitted through radio waves.

These can be active (possess an internal power source) or passive (powered by an external scanner). There are two scanning mechanisms:

Radio-frequency identification (RFID) Passive RFID tags used in cards can be scanned up to a distance of 3 meters [0].

Near Field Communication (NFC) A newer set of standards, similar to RFID, but allows two-way communication (between two active tags) and operates at a shorter range (up to 10 centimeters) [1].

USAGE

Contactless cards and tags are used in payments (for low-value transactions), ticketing, toll booths, controlling physical access to rooms, buildings.

SECURITY ▼ LOW

Information can be remotely accessed by malicious parties using high-powered scanners [0]. Mitigations include transmitting a one-time token along with account details to limit repeated use of any stolen information (replay attack) and analytics based fraud detection for suspicious transactions [2]. RFID-proof sheaths are also used to prevent this form of theft.

SCALABILITY ▲ HIGH

As of 2018, 370 million contactless payment cards exist across 111 countries [2]. NFC chips are commonly found in smartphones [3]. No known scalability challenges exist.

MATURITY ◁▷ MEDIUM

Usage dates back to 2000s, interoperable standards exist [4].

COST ◁▷ MEDIUM

Supported cards need to be issued, scanners and decoding software required.

[0] *https://hackaday.com/2013/11/03/rfid-reader-snoops-cards-from-3-feet-away*
[1] *https://electronics.howstuffworks.com/difference-between-rfid-and-nfc.htm*
[2] *https://newsroom.mastercard.com/2018/01/17/dispelling-the-myths-the-reality-about-contactless-security-2*
[3] *https://en.wikipedia.org/wiki/List_of_NFC-enabled_mobile_devices*
[4] *https://www.iso.org/standard/73599.html*

QR Code

A QR Code is a type of machine-readable optical label.

Information is encoded using Quick Response (QR) codes for convenient scanning of documents. A digital signature of the issuing authority can also be encoded to verify the source of the information.

USAGE

QR codes are used for quick, but insecure, document scanning.

SECURITY ▼ LOW

Can be trivially copied and intercepted but provides quick, cost-effective machine-based scanning.

PRIVACY

Depends on the information being encoded. Utility is limited to encoding of pseudonymous and not personally identifiable information.

SCALABILITY ▲ HIGH

Proven at internet scale.

MATURITY ◁▷ MEDIUM

Interoperable standards exist [0].

COST ▼ LOW

The cost of issuing a QR is negligible, can be scanned with low-cost smartphone cameras.

[0] ISO/IEC 18004 *https://www.iso.org/standard/62021.html*

Biometric System-on-Card (BSoC) [0]

Biometric system-on-card (BSoC) is a type of smart card with an embedded fingerprint scanner which forms a built-in second authentication factor.

Storage and matching of fingerprints happens on the card (offline authentication) before the credential or attributes stored on the card are released. It can be made compatible with standard smart card readers including EMV (microchip) readers, which are widely deployed by the payments industry [1].

USAGE

This technology is seeing pilot deployments in the payments industry.

SECURITY ▲ HIGH

Built-in multi-factor authentication. Similar to smart cards. Susceptible to flaws of fingerprint scanning.

SCALABILITY ▲ HIGH

Unproven but scalability properties should be similar to smart cards.

MATURITY ▼ LOW

Pilot programs from Visa and Mastercard exist that are limited to select banks [1] [2].

COST ▲ HIGH

Embedding a fingerprint reader in each card can be costly.

[0] *https://www.thalesgroup.com/en/markets/digital-identity-and-security/banking-payment/cards/emv-biometric-card*
[1] *https://www.mastercard.us/content/dam/mccom/en-us/documents/biometric-card-merchant-faq.pdf*
[2] *https://usa.visa.com/visa-everywhere/security/biometric-payment-card.html*

Security Keys [0]

Security keys are physical authentication devices which can grant access once connected to a computer or mobile device.

These are similar to smart cards, contactless cards, and biometric system-on-cards.

The hardware is usually shaped like a thumb drive and consists of:

1. a microchip for cryptographic operations and verification.
2. storage for key material, biometrics, and other credentials.
3. USB, NFC chip or, a Bluetooth chip for interfacing with a computer or mobile device.
4. optionally, a fingerprint scanner for offline authentication.

Standard protocols such as FIDO CTAP, U2F, WebAuthn and, PIV are available, but proprietary solutions also exist.

USAGE

Security Keys are currently deployed in niche high-security use-cases like authentication of employees in organisations and by high-risk individuals using web services.
New web standards and support in popular web browsers and consumer devices is making this technology available more widely.

SECURITY ▲ HIGH

The use of dedicated hardware authenticators is considered an industry best-practice [1]. These are typically used as a second factor, but new standards such as WebAuthn aim to use built-in MFA in such keys to replace passwords entirely.

SCALABILITY ▲ HIGH

Has seen deployment in large organisations.

MATURITY ◁▷ MEDIUM

Open standards exist. Increasing support from online services.

COST △ MEDIUM TO HIGH

Dedicated hardware keys can be costly. Additional cost associated with security features such as fingerprint readers and convenience features such as NFC and Bluetooth.

[0] *https://www.wired.co.uk/article/best-security-keys*
[1] *https://landing.google.com/advancedprotection/*

Smartphones & Computers

Smartphone and computer applications are used to store and secure digital credentials which are used to access online services.

Similar to other digital identity artifacts, smartphones are equipped with:

1. Processors and storage for cryptographic operations and key material,
2. Interfaces such as USB, Bluetooth and NFC, and
3. Biometric authentication mechanisms such as fingerprint and face recognition for local authentication.

The credentials that represent an individual's digital identity are sometimes supplied through a SIM card provided by a cell service operator, or directly established with the authenticating entity through the internet.

USAGE

Smartphones and computers are used for identification, authentication, and authorisation in scenarios where the mode of access to a service is through the Internet. They have been deployed in public ID schemes [0], e-commerce, payments, and a wide range of online services.

SECURITY ◁▷ MEDIUM

Even though smartphones offer the standard security features found in hardware authenticators, they offer less security as they are directly connected to the internet, increasing the attack surface for malicious actors.

SCALABILITY ▲ HIGH

Proven at internet scale.

MATURITY ◁▷ MEDIUM

Open standards exist. Increasing support from online services.

COST ▲ HIGH

Personal computing devices cost more than other identity credentials, but people may already possess them.

[0] *https://digitalid.design/research-maps/estonia.html*

DIGITAL ID WORKFLOWS

System workflows that describe how identification, authentication, and authorisation are carried out in digital ID systems.

The various workflows are:

1. Multi-factor Authentication (MFA)
2. Single Sign-on (SSO) / Federated Authentication
3. Self Sovereign Identity (SSI) / Decentralised ID
4. Access Control

Multi-factor Authentication (MFA)

Is the practice of requiring two (also known as 2FA) or more factors for authentication.

Different kinds of factors (knowledge, possession, inherence) must be used (also known as multimodal authentication). For example, requiring two possession factors is not considered multi-factor authentication.

USAGE

Used to strengthen confidence in authentication in cases where a single factor does not provide sufficient assurance. For instance, in online banking.

SECURITY ▲ HIGH

It depends on the security of the underlying factors used, but generally this is considered to be a best-practice [0].

SCALABILITY ▲ HIGH

It depends on scalability of underlying factors used, but generally, there are no scalability concerns.

MATURITY ▲ HIGH

Well known. Open standards exist [1].

COST ▽ LOW TO MEDIUM

Higher than single factor authentication.

[0] *https://www.nist.gov/itl/applied-cybersecurity/tig/back-basics-multi-factor-authentication*

[1] *https://www.yubico.com/authentication-standards/fido-u2f/*

Single Sign-on (SSO) / Federated Authentication [0] [1] [2]

Refers to a set of mechanisms where one entity (called identity provider) is responsible for authenticating an individual, after which they are automatically authenticated by a range of applications or services (relying parties) that recognize this provider.

The individual may need to provide consent to the identity provider to allow the relying party to authenticate them and recieve any accompanying identity attributes.

There are three common flows in single sign-on authentication:

1. the relying party directly contacts an individual's identity provider to check authentication.
2. the identity provider issues a digitally signed token to the individual which they present to a relying party. The relying party, in turn, independently verifies this using a pre-shared public key without contacting the identity provider.
3. an intermediate entity (called broker), which is a server operated by a third party, can mediate exchange of credentials. This design is commonly seen in federated ID systems that connect multiple identity providers to multiple relying parties.

USAGE

Used for authenticating individuals to access online services.

SECURITY

Depends on the security of the identity provider, standard and implementations used, and the architectural model chosen (centralised, federated, decentralised). Centralisation is a concern for SSO systems as a popular identity provider could yield disproportionate power over issuance and revocation of identities, and form a honeypot of data that is a lucrative target for criminals. Additionally, the compromise of a single credential used for SSO can affect many disparate services.

PRIVACY

Since multiple different entities (relying parties, identity providers, brokers) are involved in the authentication process, the personal data and metadata about individuals visible to each party must be taken into account. Personal data includes the credentials and any other identifying information used for authentication. Metadata includes the time and frequency of authentication, the identity providers associated with the individual, and what relying parties are being accessed by them.

SCALABILITY ▲ HIGH

Proven at internet scale.

MATURITY ▲ HIGH

Open, interoperable standards exist. Used in organisational intranets since a long time.

COST ▼ LOW

There are costs associated with remunerating commercial identity providers and identity brokers or setting them up in-house.

[0] *https://auth0.com/docs/sso/current*
[1] *https://developers.google.com/identity*
[2] *https://developers.facebook.com/docs/facebook-login/*

Self Sovereign Identity (SSI) / Decentralised ID [0] [1]

These are umbrella terms used by software vendors to describe a wide range of improvements proposed in new digital ID solutions sold by them.

Depending on the system configuration, these may include improvements such as:

Source of identity/trust Decentralised digital ID systems allow the individual attributes that comprise a digital ID (name, address, date of birth, email, etc.) to be sourced from multiple, disparate entities (identity providers) in an interoperable manner.

Decentralised storage Individuals can decide where to store the credentials and identity attributes issued to them by identity providers. This can either be on their own devices, or on personal data stores managed by vendors of such systems.

Sharing of identity attributes Self-sovereign ID systems allow individuals to selectively share only the attributes of their identity that are required by the service (relying party) they are trying to access i.e. data minimisation.

Control By placing a record of issuance of a digital credential on a public or shared decentralised database (such as a blockchain), such systems claim to decentralise control over an individual's digital ID. In practice, this means that a digital credential cannot be revoked or refuted by any single intermediary, apart from the issuing authority.

Availability and metadata Services that consume digital credentials (relying parties) can independently consult the public or shared databases to verify the credentials presented to them. Identity providers do not need to remain online and do not learn about the metadata related to usage of the credential (like when and where it is being used). This is an improvement over federated architectures where identity providers play a more active role in authentication.

USAGE

They are not currently in use outside of small pilot deployments. They are intended to be used for identification, authentication, and authorisation in an interoperable manner across identity providers and relying parties in both public and private sectors.

[0] *https://www.w3.org/TR/did-core/*
[1] *https://www.w3.org/TR/vc-data-model/*

SECURITY

Concerns include:

- **Maturity** The security of such systems has not been proven, especially for the unprecedented levels of sensitive, personally identifiable information that the proposed systems, which would be integrated across public and private sectors, would need to manage. This is a particularly important factor while considering suitability for public sector use.
- **Complexity** The complex architecture of such systems can present additional attack surface for malicious actors. These systems bundle many independent improvements, and the incremental benefits of each one over existing solutions must be individually evaluated.
- **Re-centralisation** Over time, the digital ID market could converge to a few popular vendors for both storage of credentials, and for managing and sharing them. This would diminish many of the improvements purported by such systems.

PRIVACY

Similar to federated systems, many different entities are involved in issuing, consuming, transferring, managing, and storing digital IDs in decentralised systems. The level of access to private data and metadata given to each intermediary should be carefully considered.

- **Metadata** There is not enough study on how metadata is handled by such systems, particularly around the use of immutable databases like blockchains. Records stored on the blockchain are de-identified but could be re-identified, and used in concert with tracking mechanisms such as web tracking. System vendors and other network participants could link identity providers and relying parties associated with an individual, and the time/frequency of authentication.
- **Increased Data Collection** Systems that seamlessly interoperate between public and private entities can enhance data collection by technology companies, many of which have vested financial interests in data collection and have historically operated under meaningless, coercive notions of user consent.

SCALABILITY

Unproven. Depends on system configuration.

MATURITY ▼ LOW

Not yet deployed on a wide scale.

COST

Depends on system configuration.

Access Control

A set of mechanisms used to define who has access to what data and services in the context of a digital ID system.

The various access control techniques are [0] [1] :

Mandatory Access Control Defines an approach where access to each object or action is controlled.

Discretionary Access Control Defines an approach where the level of access allowed to each object or action is set by the owner associated with it, as opposed to a single, central authority.

Role-Based Access Control In this mechanism, roles are assigned to various actors in a system with different levels of permissions and privileges allotted to different roles.

Rule-Based Access Control Each object and action has an explicitly defined set of policies mandating what each actor can and cannot do in the context of a particular system.

Attribute-Based Access Control A set of rules and policies is dynamically evaluated based on the attributes of a particular actor and the relationships between them. eXtensible Access Control Markup Language (XACML) is a programming language used to define these rules.

USAGE

Used to constrain access to data and online services so that it is only available to authorised actors.

SECURITY

Access control is a standard security measure for any internet application and must be incorporated into the design.

SCALABILITY ▲ HIGH

Attribute-based access control, which allows fine-grained access control based on individual user attributes is more computationally expensive.

MATURITY ▲ HIGH

Mature implementations of access control techniques exist.

[0] *https://www.techotopia.com/index.php/Mandatory,_Discretionary,_Role_and_Rule_Based_Access_Control*
[1] *https://www.imperva.com/learn/data-security/role-based-access-control-rbac/*

DIGITAL ID STANDARDS

Standardised software specifications that define methods, interfaces, communication protocols for digital ID systems.

These are typically accompanied by open-source implementations in the form of software libraries. Standardisation also implies some degree of interoperability between implementations.

USAGE

These standards define ways to represent, manage, and share aspects of an individual's digital identity online. They form common languages for identity providers, relying parties, digital ID system vendors, and individual devices to interoperate. These standards are primarily used for authentication and authorisation, but some also support management of identification information.

SECURITY

We do not give a security rating for the standards described below as the security of software depends on implementational and operational factors and is not, in most cases, inherent to the standards, which are theoretically sound. However, we do highlight security and privacy concerns, where applicable.

SCALABILITY ▲ HIGH

The scalability of software standards is generally high by design. However, implementation-specific issues may arise.

COST ▼ LOW

The cost of software procurement is generally low as free, open-source implementations of these standards exist.

The different Digital ID Standards are as follows:

1. Security Assertion Markup Language (SAML)
2. OpenID and OpenID Connect
3. Decentralized Identifiers (DIDs) & Verifiable Credentials
4. Web Authentication (WebAuthn)
5. Client-to-Authenticator Protocol (CTAP) & Universal 2nd Factor (U2F)
6. Personal Identity Verification (PIV)
7. OAuth
8. User Managed Access (UMA)
9. Transparency and Consent Framework (TCF) & Consent Management Platforms (CMPs)

Security Assertion Markup Language (SAML) [0] [1]

SAML is a federated authentication standard that was developed in early to mid-2000s by the Organization of the Advancement of Structured Information Standards (OASIS).

SAML defines data structures, called 'assertions', for exchange of signed credentials and other attributes between identity providers and relying parties, and a set of procedures for their verification. This protocol uses the XML (eXtensible Markup Language) format for data exchange. The current major version is 2.0.

USAGE

It is deployed to manage user authentication in universities, governments, and other large organisations. It is also used in federated national ID schemes.

SECURITY & PRIVACY

Depends on system configuration. See Single Sign-on for general considerations.

SCALABILITY ▲ HIGH

Proven at internet scale.

MATURITY ◁▷ MEDIUM

Open standard. Mature open-source implementations exist. Widely supported.

[0] *https://www.cloudflare.com/learning/access-management/what-is-saml/*
[1] *https://www.okta.com/identity-101/whats-the-difference-between-oauth-openid-connect-and-saml/*

OpenID and OpenID Connect [0]

OpenID and its successor OpenID Connect are standards that were developed in the late 2000s/early 2010s by the OpenID Foundation. Similar to SAML, OpenID provides federated authentication that enables single sign-on.

OpenID defines a common set of data structures, and signing and verification methods in the standard JWT (JSON Web Token) format for the exchange of authentication information and additional attributes between an identity provider and relying parties.

The current version, OpenID Connect features an authentication layer over the popular authorisation framework OAuth 2.0, providing the benefits of both.

USAGE

OpenID is a federated authentication standard. Unlike SAML, which is popular in enterprise environments, OpenID Connect saw adoption in consumer facing applications with support from large identity providers such as Google and Microsoft for authentication. However, it has not seen widespread use. It is also used in federated national ID schemes.

SECURITY & PRIVACY

Depends on system configuration. See Single Sign-on for general considerations.

SCALABILITY ▲ HIGH

Proven at internet scale.

MATURITY ◁▷ MEDIUM

Open standard. Mature open-source implementations exist. Widely supported.

[0] *https://openid.net/connect/faq/*

Decentralized Identifiers (DIDs) & Verifiable Credentials

These are independent but complementary standards [0] [1] that define protocols and interfaces for proposed decentralised digital identity systems to interoperate.

They are being developed and standardised at the Worldwide Web Consortium (W3C) led by organisations under the Decentralized Identity Foundation [2].

A Decentralized Identifier (DID) is a globally unique identifier, akin to a URL, created and owned by an individual. The standard envisions this identifier being used to locate individuals across compatible digital identity systems, which can then provide relying parties and identity providers further ways to interact with them.

The Verifiable Credentials specification defines procedures to issue attributes that comprise a digital identity, to a DID, and to verify them. It expects DIDs to be stored in an Identifier Registry (usually a blockchain, but can also be a centralized database). An identity provider can issue a digitally signed credential to an individual, which is linked to their DID. The individual can present this credential to a relying party, who can verify its authenticity, and the validity of the issuer and the individual by consulting the registry. This specification supports selective sharing of identity attributes, in line with the principles of data minimisation.

USAGE

Not in use outside of small-scale pilot deployments.

SECURITY & PRIVACY

Depends on system configuration. See Self Sovereign Identity (SSI) / Decentralised ID for general considerations.

SCALABILITY ▽ LOW TO MEDIUM

Current blockchain architectures are not considered scalable [3]. Proposed improvements [4] are unproven in practice.

MATURITY ▼ LOW

These standards are still in development and have not seen widespread adoption although there appears to be interest from commercial vendors to use them.

COST ◁▷ MEDIUM

The complexity and novelty of such systems could mean higher software procurement and hardware costs. Current blockchain architectures also present high computational requirements.

[0] *https://www.w3.org/TR/did-core/*

[1] *https://www.w3.org/TR/vc-data-model/*

[2] *https://identity.foundation/*

[3] *https://en.wikipedia.org/wiki/Bitcoin_scalability_problem*

[4] *https://identity.foundation/sidetree/spec/*

Web Authentication (WebAuthn) [0] [1] [2]

This is a recent Worldwide Web Consortium (W3C) standard [0], developed by the FIDO consortium, that aims to reduce reliance on passwords on the web.

It defines an API for enrolment and authentication which is based on public key cryptography (PKI). All major web browsers support this standard [3].

This standard is essentially a possession factor — demonstrating possession of a private key is used for authentication. However, it allows relying parties to require stronger forms of authentication by defining an abstract object called 'Authenticator', which manages cryptographic operations and storage of the private key. An authenticator can simply be software, or dedicated hardware chips such as Trusted Platform Modules (TPMs) present in phones and computers, or even physical tokens over protocols such as CTAP and U2F. The authenticators can support additional inherence/knowledge factors such as biometrics or PINs for higher assurance. Common examples of Authenticators are Windows Hello, Apple TouchID, and Yubico security keys.

This is considered a part of the 'FIDO2' standard. Universal Authentication Framework (UAF) was a previous iteration of this and did not achieve widespread use or standardisation.

USAGE

Used for authentication on the web. Not widely deployed as of yet but the standard has been adopted by all major web browsers.

SECURITY

More secure than passwords. Offers protection from phishing and weak, re-used passwords. Sensitive information is stored on the authenticator and is never transmitted. Depends on security of Authenticator chosen. Authenticators with built-in MFA offer high assurance and aim to replace passwords on the web.

PRIVACY

Allows individuals to be authenticated anonymously.

SCALABILITY

Not yet seen large scale deployment yet so scalability is unproven.

MATURITY ▼ LOW

This is a relatively new standard.

COST ▽ LOW TO MEDIUM

Hardware-based Authenticators can be expensive. Dedicated chips for cryptographic operations such as TPMs are usually only found on higher-end devices. However, a cost-effective software-based Authenticator could also provide a reasonable level of security as a second factor.

[0] *https://www.w3.org/TR/webauthn/*
[1] *https://developer.mozilla.org/en-US/docs/Web/API/Web_Authentication_API*
[2] *https://webauthn.guide/*
[3] *https://caniuse.com/#feat=webauthn*

Client-to-Authenticator Protocol (CTAP) & Universal 2nd Factor (U2F) [0] [1] [2] [3]

These are communication protocols for web browsers and other applications to interface with external authenticators such as hardware-based security keys over various media (USB, NFC, Bluetooth).

The current version is CTAP2 and it is part of the 'FIDO2' standard along with WebAuthn.

CTAP2 supports additional user verification (like biometrics, PINs) whereas its predecessor U2F (also called CTAP1) was designed to be used as a second factor.

USAGE

They are used for hardware-based authentication in online services.

SECURITY

Hardware-based authentication is considered an industry best practice [4].

PRIVACY

Hardware authenticators allow individuals to remain anonymous while being a strong authentication measure.

SCALABILITY

Not yet seen large scale deployment yet so scalability is unproven.

MATURITY ▽ LOW TO MEDIUM

CTAP2 is a relatively new standard. However, it is backwards compatible with U2F which is more mature and was moderately popular [0].

COST ◁▷ MEDIUM

There are costs associated with the hardware authenticators that support these protocols.

[0] *https://www.tomshardware.com/news/us-government-adopts-fido-u2f,35447.html*
[1] *https://www.theverge.com/2019/2/22/18235173/the-best-hardware-security-keys-yubico-titan-key-u2f*
[2] *https://doubleoctopus.com/security-wiki/protocol/client-to-authenticator-protocol/*
[3] *https://www.okta.com/blog/2019/01/understanding-fido-standards-your-go-to-guide/*
[4] *https://landing.google.com/advancedprotection/*

Personal Identity Verification (PIV) [0]

This open standard was developed by the National Institute of Standards and Technology (NIST) as part of its Federal Information Processing Standards (FIPS) for use within the US government. However, it is also supported by commercial authentication products [1].

PIV specifies standards for credentials, which is typically a smartcard but security keys are also supported. It allows for authentication mechanisms like PKI, PINs, biometrics, photographs, and other unique identifiers. It also supports cryptographic signing and verification.

USAGE

PIV is a protocol for hardware-based authenticators (smartcards, security keys, etc.). It is primarily used within the US government and by its contractors.

SECURITY

It is designed for centralized authentication in organisations and not for consumer use. It supports MFA.

SCALABILITY ◁▷ MEDIUM

Proven at organisational scale.

MATURITY ◁▷ MEDIUM

Open implementations exist. Has not seen adoption outside US government and its contractors.

COST ◁▷ MEDIUM

There is cost associated with smartcards or security keys used in this standard.

[0] *https://piv.idmanagement.gov/elements/*
[1] *https://developers.yubico.com/PIV/Introduction/YubiKey_and_PIV.html*

OAuth [0] [1]

OAuth is a popular federated authorisation standard developed at the Internet Engineering Task Force (IETF).

It allows for access delegation i.e. an individual can allow a third party to access a resource or a service online on their behalf. This is achieved through the use of a temporary credential, called an access token which is issued by a web service on behalf of an individual. The token is given to a relying party which can then use it to access specific resources that the individual has authorised. The framework also allows for automatic expiration of tokens and revocation of access.

OAuth is sometimes incorrectly used for authentication by using an access token as proof of authentication of an individual. The OpenID Connect protocol is an extension to OAuth that allows for passing of identity assertions for authentication in addition to authorisation.

The current version of OAuth is 2.0.

USAGE

OAuth is an authorisation framework widely used by web services (including large ones like Facebook, Google, and Microsoft) as a standard, interoperable way for their users to share their information with third-party services.

SECURITY

The OAuth framework has been criticised [3][4] for being too flexible to accommodate a large variety of business use-cases, making it difficult for developers to implement a secure solution.

PRIVACY

OAuth reveals metadata about individuals to entities involved in the transaction. Relying parties learn about the identity providers associated with an individual, and identity providers learn about the relying parties and the time/frequency of access.

SCALABILITY ▲ HIGH

Proven at internet scale.

MATURITY ◁▷ MEDIUM

Open-source implementations exist. Has undergone formal security analysis from researchers [2].

[0] *https://developer.okta.com/docs/concepts/auth-overview/*

[1] *https://oauth.net/articles/authentication/*

[2] What's Wrong with OAuth2? | Identiverse 2018 *https://www.youtube.com/watch?v=OLwz7pIXOWQ*

[3] *https://web.archive.org/web/20130325140509/http://hueniverse.com/2012/07/oauth-2-0-and-the-road-to-hell/*

[4] *https://dl.acm.org/doi/10.1145/2976749.2978385*

User Managed Access (UMA) [0] [1] [2]

User Managed Access (UMA) is a federated authorisation standard built on top of the OAuth protocol and is developed by the Kantara Initiative.

It tweaks the OAuth protocol to allow for certain use-cases. It separates the point of authorisation from the point of access to a resource, allowing for authorisation to multiple different services (federation) to be controlled from a central point, like a dashboard accessible to an individual. Additionally, while OAuth focusses on delegating access to other services, UMA enables person-to-person sharing of authorised resources. It also allows for policy-based authorisation that does not require explicit consent from an individual at the time of request of access to a resource.

The current version of UMA is 2.0.

USAGE

Outside of small-scale deployments, UMA has not seen much use yet.

PRIVACY

UMA takes a user-centric approach to access delegation, giving individuals more control on how to share their data. The metadata generated is similar to OAuth, where relying parties (which can be services or other individuals) learn about the identity providers associated with an individual and vice-versa. It introduces a new component — an authorisation server, which presents an additional threat from a privacy and security perspective.

SCALABILITY

Not yet seen large scale deployment yet so scalability is unproven.

MATURITY ▼ LOW

Some open implementations exist but this protocol is relatively new.

COST ◁▷ MEDIUM

The dedicated authorisation server defined by these specifications will present a higher operational cost than OAuth.

[0] *https://ldapwiki.com/wiki/User-Managed%20Access*
[1] *https://medium.com/@dewni.matheesha/user-managed-access-uma-2-0-bcecb1d535b3*
[2] *https://kantarainitiative.org/confluence/display/uma/UMA+Implementations*

Transparency and Consent Framework (TCF) & Consent Management Platforms (CMPs) [0]

Transparency and Consent Framework (TCF) is a collection of policies and open technical standards for consent management developed by the Interactive Advertising Bureau (IAB) Europe.

Consent Management Platforms (CMPs) are commercial software solutions that help web publishers display notices about data collection and obtain consent. They are a common fixture on websites today in the form of 'Manage Cookies' banners.

TCF requires advertisement vendors and CMPs to register with the IAB. This central registry is used by publishers to view what vendors are part of the ad network and how they intend to comply with regulations, and to select which ones appear in their consent collection user interfaces. The technical standards specify common data exchange protocols and formats for participating CMPs and vendors to collect consent in an interoperable manner.

USAGE

TCF is used by digital advertisers, vendors, and publishers to comply with data protection regulation requiring personal data processors to obtain and record consent.

PRIVACY

Led by advertisers whose profits are tied to increased data collection that enables targeted advertising, commercial incentives of these platforms are not aligned with maintaining user privacy. A study [1] found that among the most popular CMPs on websites in the UK, only 11.8% were found to meet minimal consent requirements under European law. It also reported the use of dark patterns — manipulative design practices that make it harder to register objections than to give consent [2] .

SCALABILITY ▲ HIGH

Proven at internet scale.

MATURITY ▼ LOW

These frameworks and platforms are relatively new.

COST ▽ LOW TO MEDIUM

There is cost associated with registering with the IAB as a TCF compatible vendor or CMP.

[0] *https://iabeurope.eu/wp-content/uploads/2019/08/TCF-Fact-Sheet_General.pdf*

[1] *https://arxiv.org/pdf/2001.02479.pdf*

[2] *https://www.fastcompany.com/90452333/why-you-still-cant-escape-dark-patterns*

GLOSSARY

RELYING PARTY

A Relying Party is an entity that uses the Authentication mechanism provided by an Identity System to verify the Identity of an Individual, in order to process a transaction or grant access to a system, or information, or a service. Based on the nature and purpose of the Identity System, relying parties can be both government bodies or private actors.

IDENTITY PROVIDER

An Identity Provider is an entity that provides an Authentication mechanism to an Identity System. It is trusted by relying parties to identify an individual an authenticate their identity when access to a system, or information, or a service is requested. Identity providers may also be government bodies or private actors.

BLOCKCHAIN

A Blockchain is a database that contains a permanent record of everything it stores (immutability) and is maintained by many independent entities, none of which can individually dictate its contents (decentralization).

FALSE POSITIVE RATE OR FALSE MATCH RATE

False Positive Rate or False Match Rate is the probability of an incorrect input being accepted as a match, and False Negative Rate is the probability of a correct input being wrongly rejected. These terms are used to describe the accuracy of biometric systems.