Innovation Insight for Many Flavors of Authentication Token

Published 30 March 2023 - ID G00778753 - 31 min read

By Ant Allan, James Hoover

Authentication tokens provide the commonest way of implementing MFA to address the weaknesses of passwords and reduce ATO risk, but token-based methods are increasingly defeated. Security and risk management leaders must act swiftly to mitigate risks and plan for migration to phishing-resistant MFA.

Overview

Key Findings

- One-time password (OTP) hardware tokens, and especially phone-as-a-token methods, dominate many multifactor authentication (MFA) use cases, but these methods are increasingly defeated by phishing (broadly defined) and other attacks.
- X.509 authentication tokens (smart cards and USB tokens) are established high-trust, phishing-resistant MFA options for workforce PC and network login in risk-averse and regulated industries. However, they have high overheads and do not support other use cases well.
- Fast IDentity Online (FIDO) options, including Windows Hello for Business and FIDO2 security keys, that enable passwordless MFA are gaining traction for some employee use cases. But strategic adoption of these options is hampered by uneven support across use cases and the scarcity of true roaming smartphone authenticators.
- Customer MFA (“strong customer authentication”) is complicated by the range and diversity of customers, the varied penetration of enabling technologies (such as smartphones and passkeys), and high-friction enrolment processes.

Recommendations

Security and risk management leaders responsible for identity and access management (IAM) should:

- Ensure user authentication methods (with or without tokens) are fit for purpose by evaluating, across different use cases, total cost of ownership (TCO), user experience (UX) and other needs and constraints, as well as authentication strength (including resistance to phishing and other attacks).

- Reduce the potential vulnerabilities of legacy implementations by disinvesting from known-weak legacy out-of-band (OOB) modes, such as SMS, implementing compensating controls to protect other incumbent methods, and migrating to more effective, phishing-resistant methods, using passwordless methods whenever possible.

- Plan to improve consistency across multiple use cases by investing strategically in centralized authentication services supporting FIDO2 (including passkeys for customer authentication). Seek tactical opportunities to invest in the short term as suitable authenticators become available.

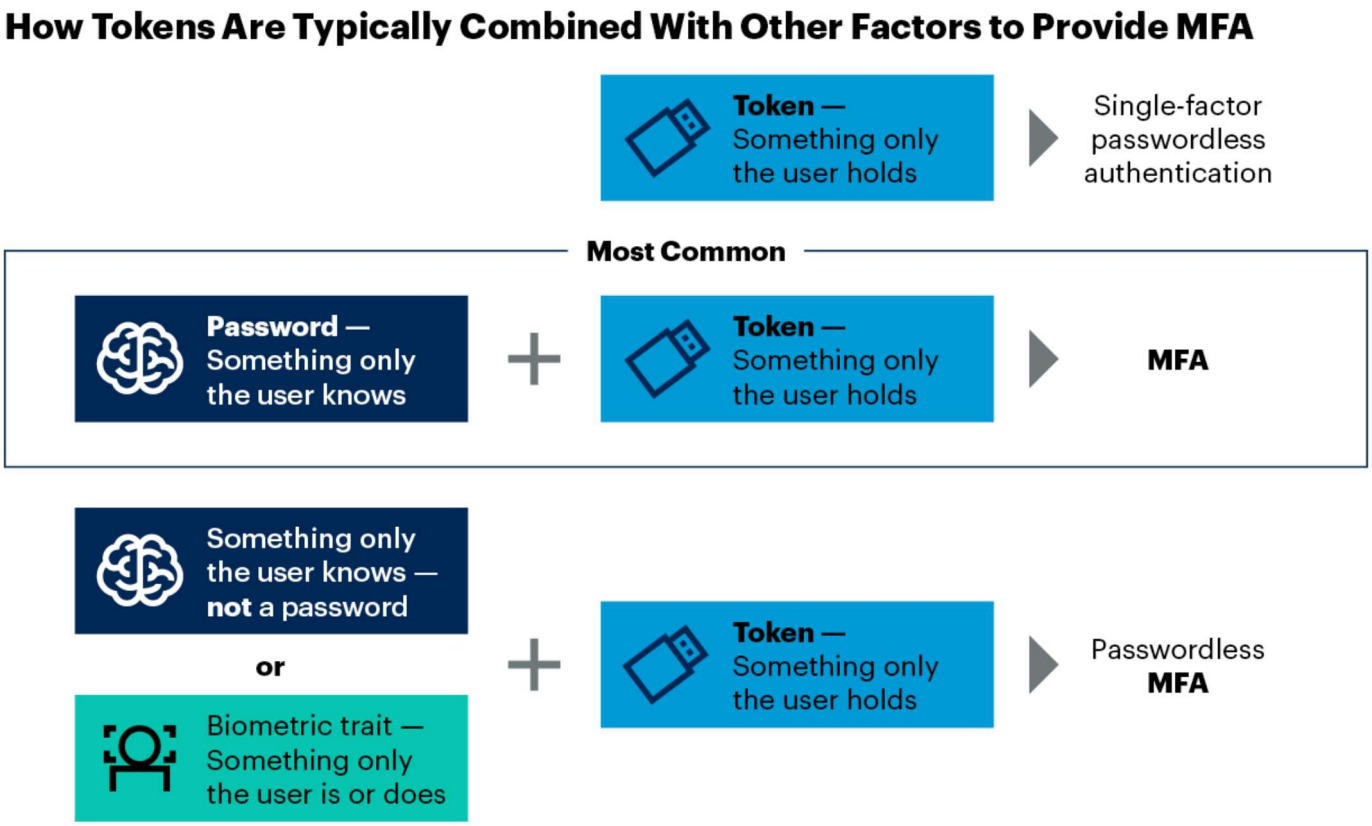
- Enhance the security of enrolment/credentialing and account recovery processes by investing in appropriate identity proofing and affirmation methods. Lower barriers to customers' adoption of new methods by simplifying these processes and optimizing the customer UX.

Introduction

The primary goal of user authentication is to provide sufficient credence [1] in an identity claim to bring the risk of account takeover (ATO) within an organization's risk tolerance. [2]

Passwords are ubiquitous, but notoriously weak. Thus new tools and technologies are needed to reduce or mitigate the risk of ATO. By far the most common approach is to add some kind of authentication token (a "possession" factor) to provide MFA (see Figure 1). [3]

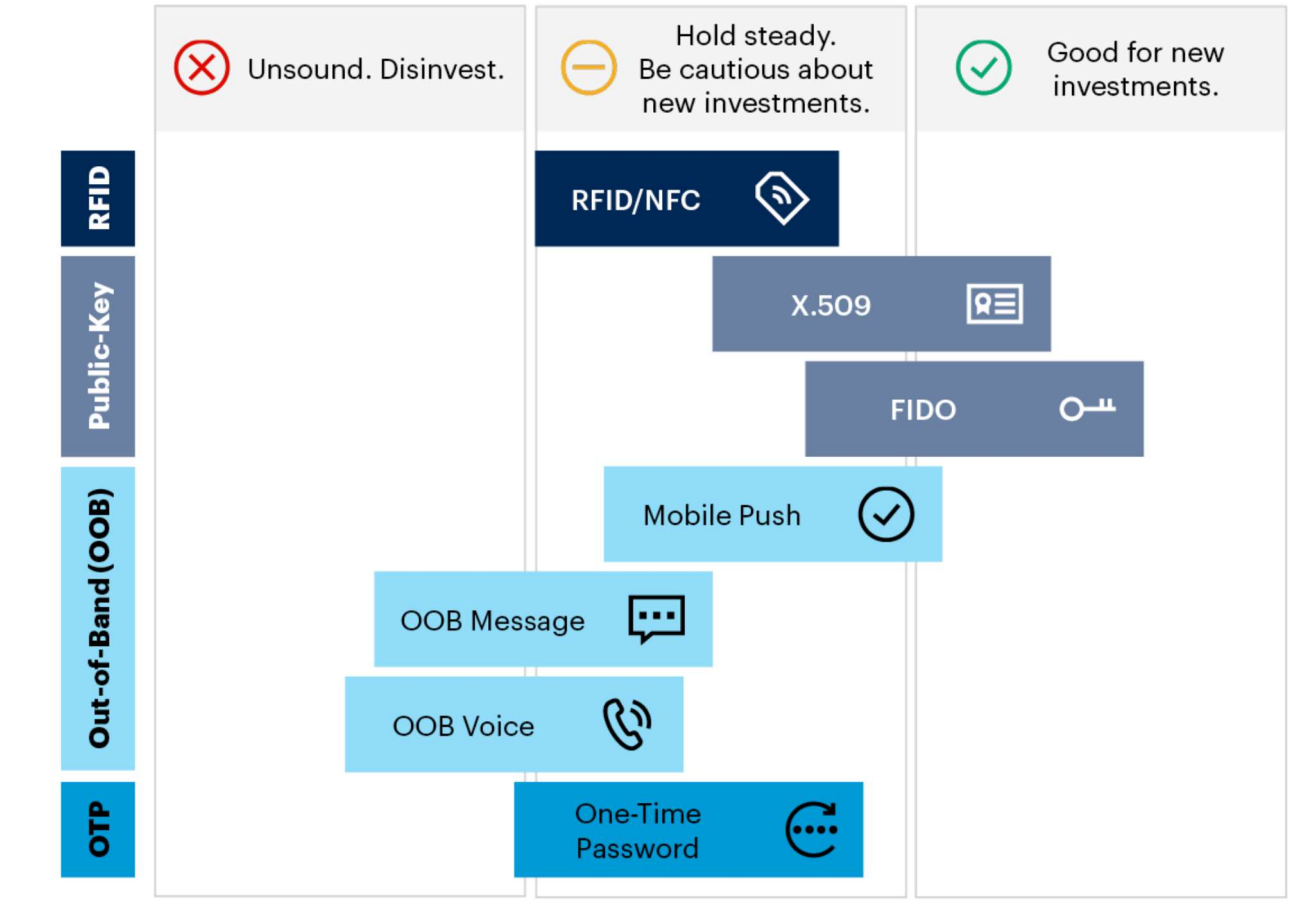
Figure 1: How Tokens Are Typically Combined With Other Factors to Provide MFA



What kinds of tokens are widely used? Are they providing the expected value? What innovations are most important? How can IAM and other security and risk management leaders choose among the options to best meet current and future needs? In this research we provide analysis and recommendations to help optimize future investments (see also Figure 2).

Figure 2: The Strategic Value of Different Flavors of Authentication Token

The Strategic Value of Different Flavors of Authentication Token



Source: Gartner
778753_C

Gartner

Description

An authentication token is a physical device or a data object containing a credential that is formally bound to an identity established to provide a specific person with access to digital assets. It is, ostensibly at least, in the sole possession of that person and thus provides evidence for the credibility of a claim to that identity.

Figure 3 illustrates where tokens lie within a simple taxonomy of authentication methods. [4]

Figure 3: Tokens Within a Simple Taxonomy of Authentication Methods

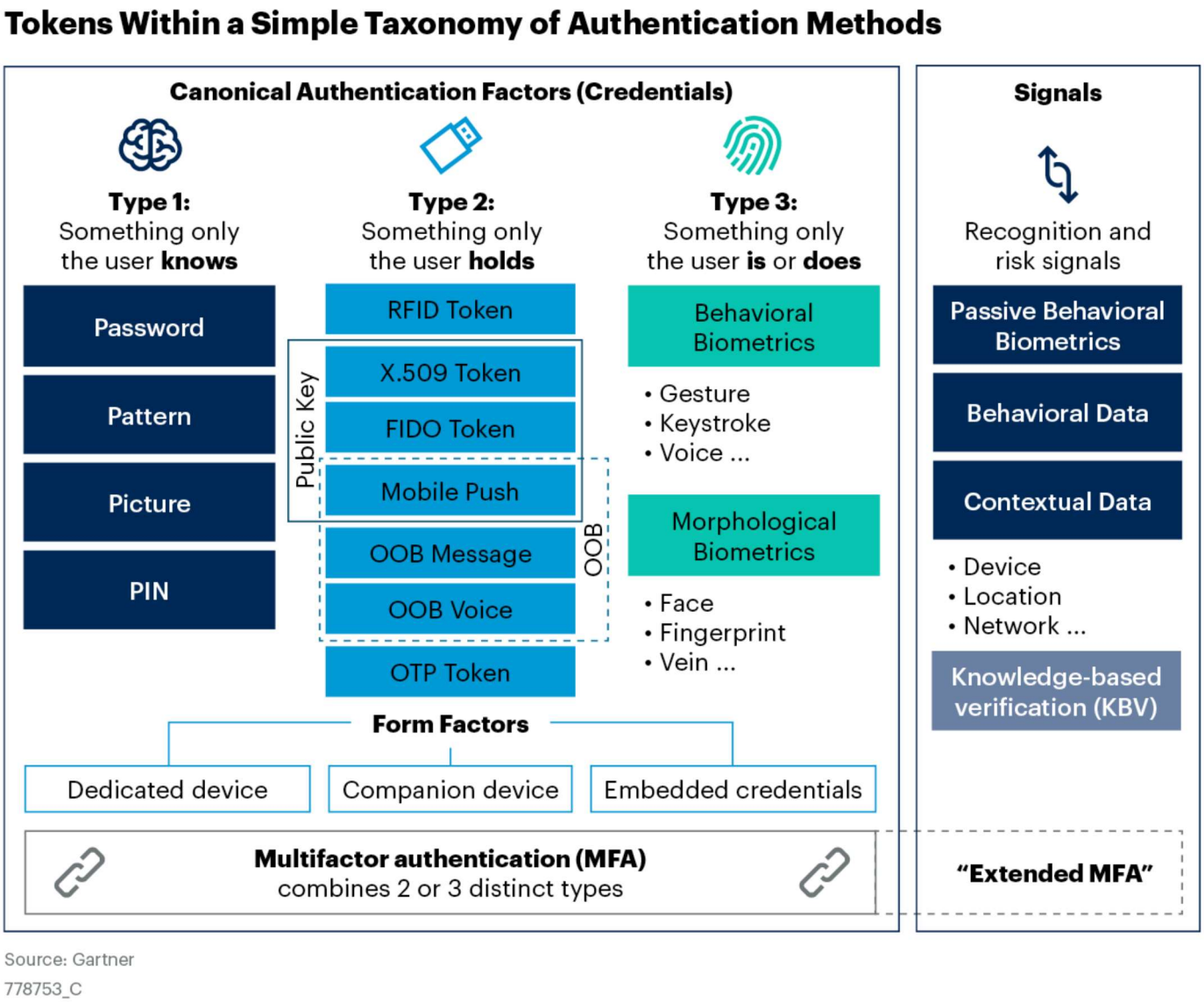


Figure 4 shows the principal flavors of token. [5]

Figure 4: The Principal Flavors of Authentication Tokens

The Principal Flavors of Authentication Tokens

		Hardware Tokens	Phone-as-a-Token and More	Software Tokens or Integral Hardware Tokens
		Dedicated Device	**Companion Device**	**Embedded Credentials**
RFID	RFID/ NFC	Building access cards	Building access apps	
Public-Key	X.509	Smart cards, USB tokens (e.g., YubiKeys), microSD	App embedding X.509 credentials	X.509 credentials ("soft certs")
	FIDO	FIDO2 security keys (e.g., YubiKeys)	FIDO2 roaming authenticator app or passkeys	FIDO2 platform authenticator (e.g., WHfB, passkeys)
Out-of-Band (OOB)	Mobile Push	"Legacy Modes"	App (typ. with X.509 credentials)	App (typ. with X.509 credentials)
	OOB Message		Messaging app or native messaging (incl. text-to-speech)	Messaging app or native messaging
	OOB Voice		"Phone" app, cellphone, landline phone	"Phone" app, VoIP
OTP	One-Time Password	Display tokens; USB tokens (e.g., YubiKeys); RCA	App	App or browser plug-in
		External authenticator (discrete device)		Endpoint device
		Form Factors: ID-1 smart cards, microSD, various proprietary USB and display tokens	**Form Factors:** Typically smartphones; "any" phone for most legacy OOB modes	**Form Factors:** PCs, tablets, smartphones, &c.; may use a secure enclave on the device

Source: Gartner
RCA = remote chip authentication; WHfB = Windows Hello for Business
778753_C

Gartner

We categorize tokens by mode as follows (from the bottom up in Figure 4):

- **OTP tokens:** These use an algorithm to generate an OTP based on a "seed," typically a secret cryptographic key (see Note 1).
- **OOB authentication:** These methods exchange authentication information between users and authentication servers via a different channel ("band") from that between endpoints and targets (see Note 2). Mobile push methods typically use public-key cryptography, but we discuss them in this category.
- **Public-key tokens:** These use public-key cryptography based on public/private key pairs (see Note 3). This category includes adolescent FIDO2 methods, including passkeys.
- **RFID tokens:** These use simple identification data, transmitted to a reader, when triggered by an interrogative radio-frequency pulse. Examples include building access cards and smartphone

apps). They may use 13.56MHz near-field communication (NFC) as well as legacy 125kHz RFID technology, but are distinct from contactless smartcards (public-key tokens).

Each mode has two or three different instantiations characterized by the device holding the credentials, as follows: [6]

- **Dedicated device:** The token is a discrete physical device expressly embedding the credentials (a "hardware token").
- **Companion device:** The token is a general-purpose physical device, such as a smartphone, distinct from the endpoint device a person is using. [7]
- **Embedded credentials:** The token is a data object (containing the credentials) located in the endpoint device a person is using to gain access to protected assets. [8]

A smartphone, used as a token supporting access from a PC, may be used to access the same assets as an endpoint device in its own right. In this case, the token "flips" from the "companion device" to the "embedded credentials" category. This has implications that we discuss in the Risks section later.

A multiprotocol token (device or app) can support two or more distinct modes, consolidating authentication across diverse use cases and enabling migration between methods. [9]

Benefits and Uses

Authentication tokens are sometimes used for single-factor passwordless authentication (see Figure 1). For example, OOB SMS modes are used for passwordless customer authentication to online banking and other services in China and India. [10]

However, they are predominantly used in combination with a legacy password, a device or a local PIN or (less often) a biometric method to provide MFA, as shown in Figure 5. PINs are most commonly combined with public-key tokens; passwords are most commonly combined with other flavors.

Figure 5: How Different Flavors of Token Are Combined With Other Factors

How Different Flavors of Token Are Combined With Other Factors

		Dedicated Device	Companion Device	Embedded Credentials
RFID	RFID/NFC	**Legacy Password** 3PB	**Legacy Password** DNB or 3PB	
Public-Key	X.509	Legacy Password **Device PIN via endpoint** "Match-on-Card"	Legacy Password **Local PIN** DNB or 3PB	Legacy Password **Local PIN** DNB or 3PB
Public-Key	FIDO	Legacy Password **Device PIN via endpoint** "Match-on-Card"	Legacy Password **Local PIN** DNB or 3PB	Legacy Password **Local PIN** DNB or 3PB
Out-of-Band (OOB)	Mobile Push	X	**Legacy Password** Local PIN DNB or 3PB	**Legacy Password** Local PIN DNB or 3PB
Out-of-Band (OOB)	OOB Message	X	**Legacy Password** Password Response	**Legacy Password** Password Response
Out-of-Band (OOB)	OOB Voice	X	**Legacy Password** Password Response 3PB (voice recognition)	**Legacy Password** Password Response 3PB (voice recognition)
OTP	One-Time Password	**Legacy Password** Device PIN (PINpad)	**Legacy Password** Local PIN DNB or 3PB	**Legacy Password** Local PIN DNB or 3PB

Key
Knowledge Factor
Biometric Factor
Bold text shows the commonest combination
DNB: device-native biometrics
3PB: third-party biometrics

Source: Gartner
778753_C

Gartner

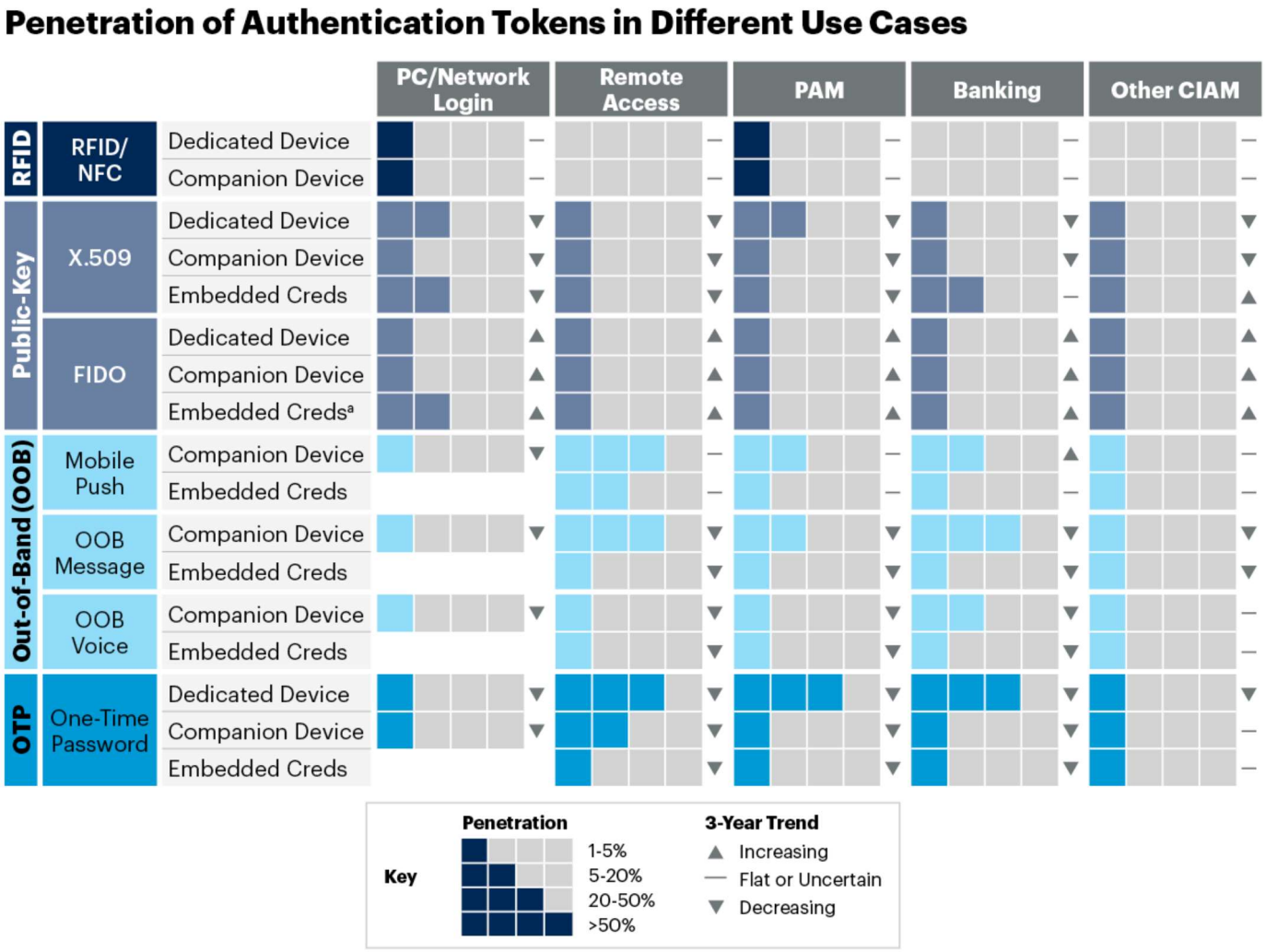
When combined with a password, it is usual for the password to be evaluated first, with the token being used in a second authentication step; thus, this kind of MFA is often called "two-step authentication" or "two-step verification." [11] In contrast, a PIN is usually combined with a token in a single step.

In some instances, the second step may be deferred until the person accesses a high-risk asset or submits a high-value transaction, or, more generally, the risk or threat level passes a threshold. This is an example of "step-up authentication." [12]

In all cases, the principal benefit of any authentication token is that it creates an additional obstacle for an attacker (and one that presents a different kind of challenge from the other authentication factor), thus reducing the risk of ATO.

Different flavors of token have greater market penetration for some use cases than others (see Figure 6). Broadly, X.509 tokens are more widely used for Windows PC and network login, whereas OTP tokens and OOB methods are more widely used everywhere else. Note that remote access encompasses SaaS applications, zero-trust network access (ZTNA) and virtual desktop infrastructure (VDI), as well as VPNs.

Figure 6: Penetration of Authentication Tokens in Different Use Cases



Source: Gartner
[a] Including Windows Hello or WHfB and passkeys
778753_C

Gartner

Risks

Attacks Against Token-Based MFA

Authentication tokens may be vulnerable to a variety of attacks, some broadly defined as “phishing” attacks: [13]

- OOB SMS or voice modes can be attacked by SIM swaps, malware and attacks against a telco's SS7 infrastructure.
- Man-in-the-middle (MITM) attacks against OTP methods (hardware tokens as well as apps) have been used for many years. Recent examples of tools facilitating such attacks include EvilProxy, Evilginx2 and Modlishka.
- Prompt-bombing or MFA fatigue attacks try to wear down users by deluging them with mobile push authentication prompts; users may eventually give in and tap the "accept" button, completing login for the attacker. [14]

Recommended actions:

- Mitigate attacks by means of:
  - Additional features (such as contextualized messaging and session binding). [15]
  - Secure configuration (choking login attempts, for example).
  - Location intelligence (for example, triangulating the locations of a login and an authenticating phone).
- Plan a migration to natively "phishing resistant" methods, such as public-key tokens; prefer FIDO2 tokens to X.509 tokens if net new investment is required.

Other attacks target credentialing and account recovery. Here attackers exploit weaknesses in the administrative processes for issuing and replacing tokens:

- The exposure arises because organizations' investments in tokens have not been matched by investments in identity proofing. [16] This weakness can be exploited by attackers (for example, to enroll their own phones for mobile push).
- Password reset processes often rely on the same additional factor used in MFA, thus weakening MFA. For example, if attackers can compromise the phone-as-a-token method, they can use it to change the password and then use that password plus the phone for ATO.

Recommended action:

- Refresh enrollment and recovery workflows that incorporate suitable identity proofing and affirmation tools. [17] Do not rely on knowledge-based verification and email-delivered OTP. [18]

See also Gartner's forthcoming research on How to Respond to the 2023 Cyberthreat Landscape.

Check-Box Compliance

How to Build a Robust, Defensible Security Program That Enables Business Growth and Agility recommends "making controls decisions based on specific risk and risk appetite rather than on check-box compliance." However, authentication method selection is still more often driven by regulatory need than by thoughtful risk assessment.

A lower-cost, lower-trust token might satisfy a check-box compliance approach, but leave the organization with a greater exposure to ATO and other attacks than alternatives would. However, UX considerations can rule out higher-trust alternatives. [19]

Recommended action:

- Base selection decisions on specific risks and the organization's risk appetite, balancing security risks against operational and business risks — for example, balance ATO and fraud risks against customer retention risks. [20]

"Scope Creep" and Changing Scenarios

Organizations often extend the use of authentication methods to use cases beyond their original scope, for which they may not provide the same value in terms of trust or UX. For example:

- OTP hardware tokens were very widely used to support remote access from PCs, but many Gartner clients found that users complained (even more) when they had to use them for remote access from their smartphones.
- Phone-as-a-token methods are the more widely used alternative to hardware tokens. They provide a better UX in all use cases. However, this approach provides lower confidence for mobile use cases, as the instantiation "flips" from companion device to embedded credentials: **A device cannot be its own companion.**
- Authentication tokens might be rolled out to a larger population of users with different UX expectations and tolerances than the original cohort. [21]
- New, higher-risk applications might be introduced within an MFA regime that now provides insufficient credence. Or data center applications might be migrated to the cloud, increasing exposure to ATO and other attacks.

Sometimes authentication token choices can be undermined by people's behavior, which may be beyond the organization's control. [22]

Recommended action:

- Continuously reassess risk and trust, UX and TCO across different use cases, and adapt to changes in need.

User Resistance

Employees and customers alike may resist a requirement to use authentication tokens at all, or just object to the use of a particular kind of token.

Attitudes vary so widely across populations that they can be contradictory. For example, some users may prefer mobile push to an OTP hardware token because of the UX benefits, while other users may object to using a personal phone at all or having to install an app just to be able to authenticate.

Customer authentication is especially sensitive to a variety of issues. Customers are put off by:

- Having to carry a specialized device (such as an OTP hardware token or an EMV card reader for remote chip authentication).
- Lack of suitable technology (such as a smartphone or a passkeys-enabled device).
- High-friction enrollment ceremonies (such as the installation of an app or the setting up of passkeys, [23] even if the method offers a great day-to-day UX.

Recommended actions:

- Overcome employees' resistance by making clear the need for MFA and choosing methods and authentication flows that offer a superior UX (see Quick Answer: How to Overcome Employees' Resistance to Multifactor Authentication).
- Incentivise customers to migrate to a new method by:
  - Highlighting value — clearly articulate and personalize the security and UX benefits.
  - Minimizing hurdles — make the option easy to find and enrollment as simple as possible.

Caveat:

- Do not leave weaker methods available to customers who have enrolled for a stronger method. Doing so just leaves everyone equally exposed to ATO attacks.

Alternatives

Among orthodox authentication methods (that is, those based on curated credentials), the single-factor alternatives to tokens lie within the "Type 1" or "Type 3" categories in Figure 3 — "knowledge" methods (with the exception of PINs that are local to a device or app) and biometric methods.

Passwords are ubiquitous and their associated risks and overheads are well-known; methods based on patterns and pictures are rare; biometric methods are uncommon.

But tokens are rarely used alone; they are far more often implemented as an element of MFA (as shown in Figure 1), so the only way to provide MFA without tokens is to combine some kind of knowledge plus a biometric mode. However, Gartner sees very few examples of this style of MFA.

Further alternatives arise when we consider recognition and risk signals as well (see Figure 3). A broader, modern approach to authentication combines signals with credentials in a dynamic way (see Shift Focus From MFA to Continuous Adaptive Trust).

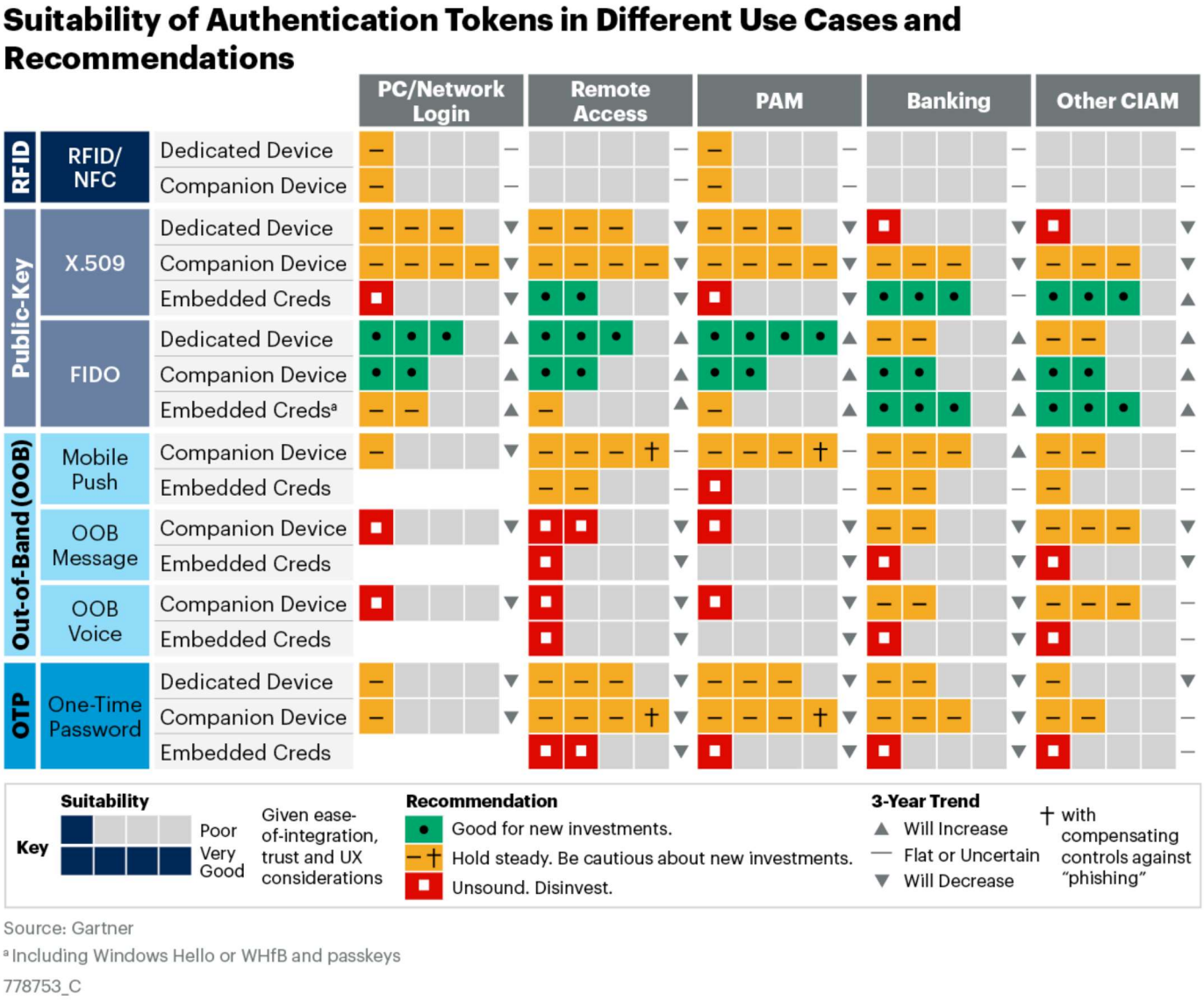
Recommendations

IAM and other security and risk management leaders should:

- Ensure user authentication methods (with or without tokens) are fit for purpose by evaluating, across different use cases, TCO, UX and other needs and constraints, as well as authentication strength (including resistance to phishing and other attacks).
- Reduce the potential vulnerabilities of legacy implementations by:
  - Disinvesting from known-weak legacy OOB modes, such as SMS.
  - Implementing compensating controls to protect other incumbent methods.
  - Migrating toward more effective, phishing-resistant methods.
  - Minimizing dependency on passwords as an authentication factor, alone or in MFA.
- Plan to improve consistency across multiple use cases by:
  - Preparing for strategic investments in centralized authentication services supporting FIDO2 (including passkeys for customer authentication).
  - Seeking tactical opportunities to invest in the near term as suitable authenticators become available.
- Enhance the security of enrolment/credentialing and account recovery processes by investing in appropriate identity proofing and affirmation methods (likely requiring specialist third-party tools). [17]
- Lower barriers to customers' adoption of new methods by simplifying enrollment processes and optimizing the customer UX. (See the recommended actions under "User Resistance" in the "Risks" section.)

Figure 7 summarizes Gartner's guidance about the use of different flavors of token in different use cases.

Figure 7: Suitability of Authentication Tokens in Different Use Cases and Recommendations



Source: Gartner

[a] Including Windows Hello or WHfB and passkeys

778753_C

Gartner

Representative Providers

Authentication tokens are widely available from more than 100 vendors. Those listed below are the ones most often mentioned by Gartner clients (see also Market Guide for User Authentication):

- Cisco (Duo Security)
- FEITIAN Technologies
- HYPR

- Microsoft (Windows Hello for Business, Azure Active Directory Premium)
- Okta
- Ping Identity
- RSA
- Secret Double Octopus
- Telesign
- Thales (Thales Digital Identity and Security)

Evidence

In addition to the sources cited below, this research is based on many hundreds of interactions with vendors and end-user organizations over the past two years.

[1] "Credence" is the formal epistemological term for the strength or degree of belief in a proposition (for example, an identity claim); see R. Pettigrew, Epistemic Utility Arguments for Probabilism, *The Stanford Encyclopedia of Philosophy*, Winter 2019 Edition. In an authentication context, this term is implicitly reflected in the "credentials" that form the basis of orthodox authentication methods. Both derive from the Latin verb *crēdo* ("I believe," "I trust in," "I rely on," "I accept as true"). It is broadly interchangeable with "assurance," "confidence" and "trust" in idiomatic usage.

[2] Note that risk tolerance and risk appetite are often treated as being synonymous, but that Gartner distinguishes between them, as discussed in Buyer's Guide for User Authentication.

[3] We use the term "MFA" throughout for combinations of two or more factors. MFA using three factors is rare, and the term is widely used as a synonym for "two-factor authentication" (2FA).

[4] A token epitomizes the second of three canonical types of authentication factor — very simply, "something only you hold" among "something only you know, hold and are (or do)" — established in 1975. The definitive publication here is Federal Information Processing Standards Publication 41, Computer Security Guidelines for Implementing the Privacy Act of 1974. IAM Leaders' Guide to User Authentication discusses a broader taxonomy that includes recognition and risk signals, as well as orthodox, credential-based methods.

However, "token" is an overloaded term in the fields of IAM and security. Some definitions also include "something only you know," such as a password. Early editions of NIST SP800-63 "Electronic Authentication Guidelines" followed this definition, but the most recent edition uses the term "authenticator" instead (still encompassing "memorized secrets").

"Token" is also used for the identity assertions exchanged in single-sign-on (SSO) and federation protocols; other terms for this are "access token," "security token" and "session token."

MFA can be extended (augmented) with recognition and risk signals, as discussed in IAM Leaders' Guide to User Authentication and, in more depth, in Shift Focus From MFA to Continuous Adaptive Trust. MFA using four (or more) factors is not possible in the canonical model, but vendors often claim device identity, location or other contextual data as additional factors.

We treat knowledge-based verification (KBV) using "security questions" as a recognition signal, rather than as "something only you know," as it is typically based on information that is often freely shared with others or available in public sources (including social media).

[5] By "principal flavors," we mean the kinds that are proven and widely used or emerging but strategically important.

[6] The boundaries between these categories are sometimes fuzzy. For example, a USB memory stick holding credentials as a data object is better considered a hardware device than a companion device, even though it is a general-purpose device and not dedicated to authentication.

[7] The most common examples are OTP and OOB modes that fall within what Gartner defined as "phone-as-a-token authentication" (see Technology Insight for Phone-as-a-Token Authentication). Since that category was introduced in 2015, authentication using a smartphone as an X.509 or FIDO token has become generally available.

[8] The token may be held in normal data storage (a "software token") or in a hardware element (an "integral hardware token"). A hardware element holding the credentials may be a hardware Trusted Platform Module (TPM), a SIM card, a secure enclave or a system-on-chip (SoC). SoCs such as Microsoft Pluton and Apple M1 offer more advanced protection due to tight OSs and hardware integration (mitigating MITM and other attacks).

U.S. federal guidelines (NIST SP800-63B Digital Identity Guidelines: Authentication and Life Cycle Management) recognize the use of a hardware TPM as a "device authenticator" rather than a "software authenticator;" this was an explicit comment in the 2017-12-01 errata. This puts it on the same footing as a dedicated hardware device, even though it is not a physically discrete token.

[9] During the past two decades, several vendors offered "hybrid tokens" that functioned as OTP display tokens and X.509 USB tokens, and some smart card vendors added OTP displays to their cards. However, neither of these approaches gained much traction in the market. Recently, we have seen increased client interest in USB and wireless tokens from FEITIAN Technologies, Yubico and other vendors that can be used as X.509, OTP and FIDO2 tokens. We also see some smart card vendors adding FIDO2 capability to their cards.

Hybrid tokens can ease migration to FIDO2 and provide continued support for X.509 credentials used for encryption or digital signature services.

Mobile push and OTP are frequently implemented in the same smartphone app. We also see some vendors (such as Entrust) extending their apps to support X.509 as well. We expect that an increasing number of vendors will add FIDO2 roaming authenticator functionality to their apps within the coming 12 months.

[10] Single-factor and multifactor passwordless authentication options are discussed in Take 3 Steps Toward Passwordless Authentication and, in greater depth, in You, Too, Can Start Enjoying the Benefits of Passwordless Authentication Today.

[11] Examples of service providers using the term "two-step verification" include Google (Protect Your Business With 2-Step Verification) and the Sony PlayStation Network (Two-Step Verification).

[12] Microsoft Azure AD Premium Conditional Access provides a common example of step-up authentication. Note that it usually invokes only the second factor (for example, OOB SMS or mobile push), so it is not the "full" MFA — the password is still in force. However, this option is still referred to in documentation as "require MFA."

[13] The use of the term "phishing" in this context was established primarily by the U.S. Office of Management and Budget's memorandum M-22-09 Moving the U.S. Government Toward Zero Trust Cybersecurity Principles. This memo notes that "sophisticated phishing attacks … can convincingly spoof official applications and involve dynamic interactions with users." However, it goes on to say, "Users can be fooled into providing a one-time code or responding to a security prompt that grants the attacker count access." But the latter (an example of "prompt bombing") would be the result of an attacker logging into an official application with the target's user ID and password. Thus, the definition and scope of "phishing" (and therefore "phishing resistant") are broad and fuzzy. However, the term is a useful shorthand for a class of attacks that OTP and OOB methods are vulnerable to.

[14] This is the actor vector that has had the most rapid increase, and it is increasingly mentioned in Gartner client interactions. The recent (2022) attack against Uber is the best-known example.

[15] Examples of session binding, to confirm that the person with the phone is the one logging in, include:

- **Number matching:** A code is displayed on the login pane and the user is prompted to enter this code (or to choose it from a short list) in the authenticator app on their phone.
- **QR code:** The user is prompted by the authenticator app to use their phone's camera to scan a QR code displayed on the login pane.

[16] Market Guide for Identity Proofing and Affirmation defines identity proofing as the combination of activities during an interaction that brings a real-world identity claim within organizational risk tolerances by providing an assurance that (a) the real-world identity exists and (b) the individual claiming the identity is the true owner of that identity and is genuinely present during the process.

"Password reset"-type approaches do not provide the right level of trust for the issuance or enrollment of tokens that are meant to elevate trust.

[17] In a few instances, this capability may be part of an authentication provider's toolset, especially an access management vendor's (natively or via partners). Otherwise, this will require investment in third-party tools (see Market Guide for Identity Proofing and Affirmation). Gartner has seen a small but increasing number of clients exploring the use of such tools in this use case.

[18] These methods may be acceptable for password reset, but, where tokens are used for MFA, independent, higher-trust methods are necessary (see Quick Answer: How Can I Securely Reset Employee Passwords or Recover Accounts?).

[19] OOB SMS is vulnerable to multiple attacks, including spoofing, SIM swap, mobile malware and Signaling System 7 (SS7) attacks. In light of such vulnerabilities, some regulators are urging banks to find a "path to reduce reliance on text-based one-time passcodes (SMS OTP)" — see, for example, U.K. Finance Communication on Requirements for Strong Customer Authentication, 28 January 2020. However, good alternatives are elusive, as many customers are reluctant to migrate to alternatives that may not provide a better UX and may involve awkward enrollment ceremonies.

[20] Buyer's Guide for User Authentication sets out a framework for selecting authentication methods that balances ATO risk reduction against TCO and UX, including diversity, equity and inclusion considerations, across employee and customer populations.

[21] For example, when an organization extends MFA from a small number of remote users to all employees now working from home. This change in scope also increases the TCO, but organizations are far more aware of that.

[22] Two examples we have seen also "flip" a phone-as-a-token method from companion device to embedded credentials, lowering trust: (a) using voice over IP (VoIP) to receive OOB voice calls; (b) using a messenger app (such as macOS Messages) to receive SMS texts.

[23] Passkeys are increasingly supported by service providers, but are not always supported well. Gartner sees service providers (a) failing to make it obvious to customers that passkeys are supported; (b) hiding the option among the security settings behind another label (for example, "security key," without even mentioning FIDO2); (c) failing to customize the enrolment flow to distinguish it from that used for FIDO2 security keys (for example, retaining irrelevant prompts like "now insert your security key"). Hanko hosts a demo site, passkeys.io, that shows how simple enrollment can be.

Acronym Key and Glossary Terms

2FA	two-factor authentication
AAL	Authenticator Assurance Level

ATO	account takeover
CAP	Chip Authentication Program
CR	challenge response
CIAM	customer identity and access management
CTAP	Client to Authenticator Protocol
DPA	Dynamic Passcode Authentication
FIDO	Fast IDentity Online
KBV	knowledge-based verification
MFA	multifactor authentication
MITM	man in the middle
NFC	near-field communication
OOB	out of band
OTA	over the air
OTP	one-time password
OTT	over the top
RCA	remote chip authentication
RP	relying party
SCA	strong customer authentication

SoC	system-on-chip
SS7	Signaling System 7
SSO	single sign-on
TCO	total cost of ownership
TPM	Trusted Platform Module
U2F	Universal Second Factor
UAF	Universal Authentication Framework
USSD	Unstructured Supplementary Service Data
UX	user experience
VoIP	voice over Internet Protocol
WHfB	Windows Hello for Business

Note 1. OTP Tokens

OTPs can be time-, event- or challenge-response (CR)-based. The venerable RSA SecurID tokens use a proprietary time-based algorithm; most other vendors' tokens now use algorithms developed by the Initiative for Open Authentication (OATH). CR modes can be used for transaction authorization.

Legacy OTP hardware tokens have a display showing the current OTP, which the user transcribes to the login pane. The typical form factor is a chunky "key fob." Some (larger, flatter) tokens have numeric keypads for PIN or challenge entry. A few have optical inputs to read QR-code (or similar) challenges.

A more modern form factor is a USB key that effectively generates and types the OTP at the touch of a button (the PC sees it as a USB keyboard). These are smaller than display tokens and provide better UX. YubiKeys are the best-known examples, and strongly favored by Gartner clients, but they are not unique.

OTP software tokens are available for PC desktops, as web browser plug-ins, or as smartphone apps enabling the use of phones as companion devices. This last type has been the most common, but

hybrid apps for both OTP and mobile push are now more common (see later).

Remote Chip Authentication (RCA)

RCA — which Mastercard calls Chip Authentication Program (CAP) and Visa calls Dynamic Passcode Authentication (DPA) — is used almost exclusively in banking. The bank provides a handheld card reader with a display and a keypad to each customer to support online and (sometimes) mobile banking.

RCA has two basic functions:

- **Authentication:** The customer puts an EMV payment card in the reader and enters a PIN to generate an OTP. (The PIN is the same as they would use at an ATM or point of service.)
- **Transaction authorization:** The customer additionally enters transaction details to generate an authorization (or verification) code.

Note that it is the payment card that is the customer's token. The reader is generic and can be used by anyone (for example, it could be shared by spouses/partners or colleagues at an overseas conference).

Note 2. OOB Authentication

These methods typically use a phone (a smartphone, feature phone or landline phone) as the physical token.

Mobile push uses a dedicated app.

OOB messaging (most commonly using SMS) and OOB voice use generic apps or infrastructure.

Mobile Push

Mobile push is an OTP-less mode that uses a data channel over the air (OTA) or via wireless connectivity to exchange authentication information with a smartphone app. A push notification wakes an app that prompts the user to, for example, tap a button to confirm login and generate a response.

Mobile push apps typically embed public-key credentials, used to sign the user's response and provide data integrity and data origin authentication. Such apps might thus be considered a variety of public-key token or cryptographic authenticator (see, for example, 10 Reasons to Love Passwordless #2: NIST Compliance).

Many mobile push apps also function as OTP apps to provide a fallback option when the phone has no connectivity or to satisfy user preferences.

Some vendors integrate an additional authentication factor: a local PIN; a device-native biometric mode (such as Touch ID on Apple iPhone) or, more rarely, a third-party biometric mode. This can provide passwordless authentication.

OOB Message

Most OOB message modes use generic messaging protocols (typically SMS) to deliver an OTP, which the recipient enters in the same way as an OTP generated by a token. The OTP can just be a nonce password, but may be generated algorithmically; it is sometimes called a one-time PIN or one-time code.

Some implementations are OTP-less: The recipient need only acknowledge the message (this is often called "two-way messaging") or respond with a password for single-step MFA. This password is often called a PIN, but rather than being local to the phone, it is verified against a centralized data store.

OOB modes using other messaging protocols are far less common. They include:

- **Unstructured Supplementary Service Data (USSD) modes,** which make use of the signaling channel of mobile networks. The authentication flow resembles mobile push, but OOB USSD modes do not require an app and so can be used with any mobile phone. Boloro is the best-known vendor.
- **Over-the-top (OTT) messaging modes** via instant messaging services provided by third parties, such as WhatsApp (Meta). We see this very rarely.

OOB Email Is Deprecated

OOB email modes are also used. However, email does not prove possession of a specific device; a person's access to email typically depends only on a password. Thus, combining a password and OOB email "collapses" to two instances of the same factor (that is, two passwords), which does not count as MFA.

Thus, these modes are deprecated in some standards. For example, NIST SP800-63B Digital Identity Guidelines: Authentication and Lifecycle Management states: "Methods that do not prove possession of a specific device, such as voice-over-IP (VoIP) or email, SHALL NOT be used for out-of-band authentication."

OOB Voice

Common OOB voice modes deliver an OTP via an automated voice call. In some implementations, the OTP is displayed on screen and then captured via the phone keypad or interactive voice response.

Missed-Call Authentication

Some vendors, including communications platform as a service (CPaaS) vendors, offer a "missed call" mode. In this mode the voice call is made from a varying number but the call is terminated before the user has time to answer it. The user checks the number and enters the last few digits of the calling number on the login pane with any OTP method.

This is cheaper than other OOB voice modes as the calls are not completed, but the UX is poor. It does not appear to offer any security advantages. We see this approach mostly in Southeast Asia and the eastern Mediterranean.

Note 3. Public-Key Tokens

This category historically includes tokens based on X.509 credentials (a key pair and a certificate that binds the public key to a person). FIDO authenticators also make use of public-key cryptography, but not the X.509 framework or standards.

NIST SP800-63B Digital Identity Guidelines — Authentication and Lifecycle Management calls these "cryptographic authenticators" (formerly "cryptographic tokens").

X.509 Tokens

X.509 credentials comprise a public/private key pair and a certificate that binds the public key to a person. Their use is therefore sometimes called certificate-based authentication (CBA).

Smart cards are the best-known form factor (although not all smart cards are X.509 tokens). Other hardware form factors are common, especially USB keys (including YubiKeys), but also, for example, microSD cards.

Embedded credentials ("soft certs") are also common. Increasingly, these are held in hardware TPMs, as well as emerging SoC hardware elements that combine CPU and secure hardware (for example, the Apple M1 chip). A few vendors (Entrust, for example) support the use of smartphones as companion devices.

FIDO Tokens

FIDO tokens implement authentication protocols developed by the Fast IDentity Online (FIDO) Alliance.

Although the two first-generation protocols — Universal Authentication Framework (UAF) and Universal Second Factor (U2F) — are still in play, the majority of new FIDO deployments use FIDO2. FIDO2 includes the enabling W3C Web Authentication (WebAuthn) web standard.

A FIDO2 internal or platform authenticator embedding credentials for one or more relying parties (RPs) can be native to the OS or installed on the device. In conjunction with a local "gesture" (such as a PIN or a biometric method) it provides passwordless MFA to a website via WebAuthn-conformant browsers.

There are two kinds of FIDO2 external or roaming authenticators, which provide a transportable authentication method (that is, users do not need to have FIDO2 platform authenticators on every device they use). These are:

- A FIDO2 security key, which is a hardware authenticator conforming to the FIDO Client to Authenticator Protocol (CTAP) and which stores credentials for one or more RPs. It can be used:
  - As a single factor.
  - In conjunction with a legacy password for MFA.
  - With a local-to-the-key PIN or biometric method for passwordless MFA. We expect that this mode will dominate in the majority of use cases.
- A smartphone or other device with a FIDO2 platform authenticator (native as, for example, in Android or iOS, or in a third-party app) can be used as a FIDO2 external authenticator in the same way as a FIDO2 security key.

Use of a smartphone or other device as a FIDO2 roaming authenticator was part of the original description of FIDO2. However, Gartner has seen far less progress with this than we had expected in 2021. The first-to-market FIDO2 apps (from Akamai and IDmelon, and most recently from HYPR) require proprietary desktop software or dongles, which limits their utility among enterprises seeking fully roaming authenticators.

In 2022, the FIDO Alliance announced an update to WebAuthn to enable a user to use their phone in this way to login to an app or website on a nearby device (regardless of the OS and browser that the two devices run). This is suggestive of shortcomings in the original specification.

The main benefit of the update is to enable the use of a smartphone with native FIDO2 capability (that is, passkeys; see below) for customer authentication. However, Gartner expects that the update will foster wide availability of fully transportable apps for employee authentication. At least one vendor has this on its roadmap for 2023.

Windows 10 and 11

Microsoft Windows 10 and 11 support Windows Hello for Business (WHfB), which is essentially Microsoft's implementation of a FIDO2 software authenticator (wrapped with proprietary stuff), and FIDO2 security keys:

- WHfB enables passwordless login to a Windows 10/11 PC via a local PIN or a biometric method (most Gartner clients we have spoken to are considering using face rather than fingerprint). These can be combined for "multifactor unlock," but this provides weaker MFA than that incorporating FIDO2 cryptographic credentials.

- Like any FIDO2 authenticator, WHfB always provides MFA for network and downstream access:
  - WHfB with a hardware TPM meets NIST SP 800-63B requirements for an Authenticator Assurance Level (AAL) 3 multifactor cryptographic device (that is, hardware authenticator; §5.1.9).
  - WHfB with a software TPM meets NIST requirements for an AAL2 multifactor cryptographic software authenticator (§5.1.8).
- WHfB login from an Azure AD joined (or, in some flows, hybrid joined) PC counts as MFA in Azure AD Premium Conditional Access, so the user can skip further authentication prompts (invoking Azure AD MFA or an alternative). However, Azure AD MFA is required to enroll for WHfB.
- **FIDO2 security key** support enforces the use of a local-to-the-key PIN or biometric method (FEITIAN Technologies, OCTATCO, Yubico and others offer keys with embedded fingerprint sensors). This can be used for login from an Azure AD joined or hybrid joined PC.

Passkeys

In 2022, the FIDO Alliance announced an update to WebAuthn to support multidevice FIDO2 credentials, enabling a user to automatically access their FIDO2 passkey on all their devices without having to separately enroll each device for every RP.

When a customer is asked to log in to a website from a passkeys-enabled device (phone or computer), they enable the login with the same biometric method or PIN that they use to unlock their device.

Passkeys may be suitable as a password replacement for employees, but Gartner is wary of their value for MFA, as an enterprise has no control over enrollment and configuration (in the current iteration of the standard). We recommend WHfB, FIDO2 security keys and proprietary FIDO2 authenticator apps (especially as fully roaming variants become available in the coming 12 months).