Exploring Myths, Misconceptions, and Best Practices for SMS-Based 2FA

How Trusted Engagements Help Improve Experiences

sinch

Two-factor authentication (2FA) is now a **common means of increasing account security** used by social networks, mobile commerce solutions, and financial institutions, among others. SMS is an accepted solution for 2FA. However, to be effective, 2FA over SMS has stringent service requirements, putting greater demands on messaging providers to ensure timely delivery while also maintaining service integrity. This paper will explore myths, misconceptions, and best practices for SMS-based 2FA.

About the author

William Dudley has almost 30 years of experience building and managing telecommunications network infrastructure. He defines global strategy and solutions within the mobile ecosystem, focusing on solutions for messaging (SMS, MMS, RCS, social and chat apps), mobile-enabled online security, next-generation networks (5G, LTE, IPX), and consumer engagement through mobile channels. As a mobile evangelist, Mr. Dudley communicates through both internal and external publications, social media, and is active in industry groups. You may follow him on Twitter at **@wdudley2009**.

Understanding two-factor authentication

Two-factor authentication, also called multifactor authentication (MFA), is a scheme that requires at least two authentication factors, including:

- A knowledge factor (something only the user knows, such as a password or PIN)
- A possession factor (something only the user possesses, such as an ATM card, mobile device, or smart card)
- An inheritance factor (some feature inherent in the user, such as a fingerprint or retina pattern)

2FA has been in common use for many years in situations in which the user must be accurately identified. ATM usage is a good illustration. The user presents an ATM card to the machine (something the user possesses) and then enters a secret PIN (something the user knows). Without both of these authentication factors working, the user cannot withdraw cash from the account. Another example of 2FA use is immigration control with biometrics. The traveler has a passport (something only he or she possesses), as well as a unique fingerprint (a feature inherent in the user) to match to the passport.

The widest use of 2FA in the mobile space is for one-time passwords (OTPs). OTPs are typically sent via SMS to a mobile device. Alternately, the OTP can be sent to an e-mail address on file. This is used as an additional authentication step using a random string of digits (for example, a PIN), as well as a means to change passwords or reset lost passwords.

The goal of 2FA, especially within the online and mobile space, is to reduce instances of online fraud involving monetary or identity theft. While there are variations in the manner in which an individual can be authenticated, the strength of 2FA lies in its implementation – that is, the strength of the two factors that are used to identify the individual.

When 2FA is deployed in a mobile ecosystem, there must be a variety of safeguards in place to allow the implementation to function securely. At the most basic level, a 2FA implementation should ensure a high degree of likelihood that the message containing the token will arrive to the mobile device. Unfortunately, some implementations do not meet this fundamental requirement.

Two-factor authentication through SMS

One of the most popular means to deliver OTPs or 2FA authentication tokens (PINs, alphanumeric codes, and so on) is to send them to a registered mobile device (something the user has).
Over the last 10 years, many social networks, online shopping businesses, and financial institutions are using 2FA as the standard method for resetting passwords, authorizing users, and validating transactions.
Delivery of these 2FA tokens over SMS is typically reliable and quick, and it uses a medium that virtually every mobile device can support.

While there have been well-documented cases of fraud using 2FA-based SMS through fraudulent SIM-swap or network hacking, for the most part these are not common.
For some high-value authentication use cases, there are other, higher security methods to provide 2FA; however, for many more common use cases, 2FA through SMS can be effective and secure.

Delivery of the OTP through SMS is simple and reliable because SMS is ubiquitous across mobile devices, and no special tokens or cryptographic keys need to be shared between the mobile device and the server.
There are alternatives to 2FA through SMS such as the Google Authenticator app (based on time-based, one-time password [TOTP] standards), which can provide even greater security; however, these do require a smartphone, which are not necessarily widely available in many regions.

Requirements for reliable delivery

Delivery of a 2FA message through SMS requires a reliable and quick delivery channel. This means that the pathway to the destination mobile operator should be as direct as possible through approved routes.
Most operators today require that senders of application-to-person (A2P)–type traffic (of which 2FA and OTPs are clear examples) use approved routes – either through a direct IP connection (typically via Short Message Peer-to-Peer [SMPP] protocol) or through the use of approved sending of global titles for Signaling System 7 (SS7) delivery.

The originating address of the 2FA message should be either a short code, if this is supported or required by the destination country, or a long code (standard International Telecommunication Union [ITU] E.164 telephone number).

Short codes and long codes

Sending 2FA messages to subscribers across the world presents some complex situations:

- Some countries require short codes only. Because short codes do not cross national boundaries, for each country that requires a short code, an enterprise's short codes should be acquired and connected to each of the mobile network operators (MNOs) – and to the mobile virtual network operators (MVNOs), if available.
- For MNOs in countries that do accept long codes, acceptance may depend on whether the long code is local – that is, with the national country code – or has a different country code. Additionally, for countries that accept long codes, the route that message traffic takes to connect to MNOs may be a factor.

Short-code countries

In many countries, including the United States, short codes are the norm. U.S. MNOs scrutinize and approve each short code and its campaigns. Short codes used by enterprises for 2FA and OTPs are commonly approved, and they provide reliably high delivery rates. In many of these countries, long codes for business texting are also now feasible; however, we recommend that for 2FA purposes, the sender ID should always be a properly vetted and approved short code, should a choice between short codes and long codes be available.

Over the last few years, some organizations have been attempting to deliver 2FA, PIN codes, and OTPs among related applications through long codes utilizing the national person-to-person (P2P) network. These are not approved routes and are often blocked as spam by interoperator hub providers as well as MNOs using various antispam filters.

Use of long codes in countries that support only short codes for A2P traffic is not without controversy. Some aggregators have tried to define 2FA and OTP traffic as P2P, because the traffic was initiated by the subscriber; however, this scenario is strongly held to be a classic A2P use case by MNOs and aggregators. The reality is that 2FA-type messages cannot be replied to, thus eliminating the concept of using two-way P2P in these national messaging ecosystems. This was a method used by some aggregators to try and circumvent approved A2P routes.

Long-code countries

Long codes are acceptable for 2FA and OTPs in countries where they are approved for A2P routes. In these cases, the A2P traffic flows over approved routes (through accepted global titles in the SS7 network) to reach MNO subscribers. For many multinational enterprises, long codes provide an efficient manner to reach subscribers in numerous countries. A2P messaging aggregators with a strong global reach can provide information regarding the rates that each MNO charges as well as information regarding local regulations. The A2P messaging aggregators can, many times, pay wholesale rates to the MNOs and can therefore resell that access as well as apply all appropriate mobile number portability solutions, to ensure the message is delivered to the correct MNO for the subscriber.

The use of approved A2P routes to reach MNOs is paramount for the quick and reliable delivery of 2FA and OTP tokens to subscribers. Understandably, most of these expire after a short period of time (the majority in 10 minutes or less). Consequently, it's essential for the message to be delivered quickly from the enterprise servers to the subscriber.

Unfortunately, some A2P aggregators who purport to provide global MNO connectivity will hand off the messages to intermediaries, who hand them off to still other intermediaries, who may refile the message to be delivered through SIM farms. SIM farms use subscriber identity modules (SIMs) to provide their services. They usually deploy hundreds, if not thousands, of SIM cards in automated messaging gateways. Incoming messages are resent, many times changing the originating number, into the global P2P SMS network. Messages are then sent back into the SS7 network or through messaging hubs to be delivered to the end subscriber. These messages are highly likely to be blocked as spam or as unapproved A2P traffic over P2P delivery routes. Consequently, enterprises that focus on "lowcost" messaging aggregators may find that very few of their 2FA and OTP messages actually reach the end subscriber.

Summary: getting the best delivery of 2FA messages

2FA and related messages should always use the highest-quality SMS routes available from well-known, high-quality messaging aggregators. As we've shown, there is plenty of misinformation about the methods that should be used to deliver 2FA messages, which can result in ill-informed choices.

The key point is that these messages should be delivered quickly and accurately, using approved messaging routes. Trying to cut costs by using low-quality routes or cut-rate aggregators will often end up costing more in customer frustration and the resulting lost revenues.

Learn more

To discover more, contact your Sinch representative, visit us **online**, or join **our community**.

www.sinch.com

Sinch brings businesses and people closer with tools enabling personal engagement. Its leading cloud communications platform lets businesses reach every mobile phone on the planet, in seconds or less, through mobile messaging, voice and video. Sinch is a trusted software provider to mobile operators, and its platform powers business-critical communications for many of the world's largest companies.