Protecting SMS messages used in critical business processes

Security advice for organisations using text messages to communicate with end users

Introduction

Short Messaging Service (SMS), commonly called text messages, are a convenient way for organisations to communicate with customers, staff and others. However, SMS messages are also frequently used to send highly sensitive data. For example, one-time passcodes used to access critical systems and services as part of a multi-factor authentication (MFA) system.

Unfortunately, while the qualities of SMS make it a valuable business tool, the technology was never intended to be used to transmit high risk content. Consequently, there are a number of inherent weaknesses in the ecosystem which support SMS.

These weaknesses mean that, where the value of the message content is of interest to bad actors, they are increasingly attempting to exploit SMS.

Advice

This guidance does not rule out the use of SMS for transmitting sensitive data. Instead, we advise that you should understand how your organisation uses SMS, and determine whether to put in place additional controls. To aid with that process we outline common use cases for SMS, relevant threats to the technology and possible measures to defend against them.

Mobile telecoms companies are aware of the problems with SMS and are actively working to close vulnerabilities. However, these are complex issues and it may be impossible to fully compensate for the inherent weaknesses of the system. So, any organisation using SMS must have a clear understanding of how and where the technology is used and take steps to mitigate or reduce associated risks, where appropriate.

Audience

This advice is aimed at technical staff at UK enterprises using SMS, particularly system designers and engineers.

Non-UK telecoms companies may have characteristics which would reduce the effectiveness of some of the suggested controls. For example, for the UK 'roaming' means the handset is (or appears to be) out of the country. In the USA and in other countries, it is possible for a handset to roam on another network whilst still in the country.

Why SMS is popular

The Short Messaging Service (SMS) was originally developed as an engineering signalling system. It was not designed as a method for transmitting secure messages.

SMS has a number of qualities which make it attractive for business use:

- **ubiquity** – the vast majority of mobile phones globally support the SMS protocol making it easy/cheap to develop services
- **familiarity** – consumers understand SMS
- **timely** – SMS messages generally get delivered, globally, within a few seconds
- **inexpensive** – relatively low cost to use
- **reliability** – the store and forward nature of SMS means it is often seen as a 'fire and forget service'

Organisations, particularly Banks, use SMS for the following purposes:

- **to send information to customers**
- **to send one-time passcodes to customers**
- **to confirm a questionable transaction**

General threat protection advice

1. Know your estate

Before you can determine which protections you should put in place, you must first understand exactly where and how your organisation uses SMS. You should then assess the level of risk associated with these business process.

For example, simple updates on the progress of a product application are likely to be deemed 'low risk.' Their value to an attacker is unlikely to justify the effort needed to subvert the SMS sending process. This situation is unlikely to warrant investment in additional protective controls.

Conversely, a one-time passcode sent by SMS and used to authorise new payments could be deemed 'high risk'. This type of message would be high value to an attacker and therefore justify investment in additional controls.

Suggested Controls

You should create and maintain a formal record of how and where your organisation uses SMS.

Even if the risks are deemed to be low, this inventory is vital in order to rapidly assess the impacts of new or increased attacks against SMS.

2. Consider alternatives to SMS

There are many ways by which SMS can be compromised and full defence against such attacks is not possible.

Suggested Controls

In some cases there may be alternatives to SMS, such as the Push Notifications offered by the iOS and Android ecosystems.

3. Protect the integrity of customer phone numbers

Whilst many attacks against SMS are complex and technical in nature, it's also possible to subvert SMS-reliant tasks by targeting the underlying database.

For example, access to the database which holds customer's genuine mobile numbers would allow an attacker to alter records, diverting SMS messages to a number under their control.

Suggested Controls

Before allowing phone numbers to be amended, there should always be a robust process of customer authentication, ensuring that only the legitimate owner of that number can change it.

When a phone number is updated, a message notifying of the change should be sent to the old phone number, asking the customer to make contact if they did not make this change. It may also be prudent to send this message by some means other than SMS, email for example. This checking process should itself be protected from social engineering attacks of the form: "I didn't ask for the change, please change it back to xxxxxxx."

When a phone number is updated, the change should be time stamped. When the number is used to send a 'high value' SMS, the enterprise should make a risk-based decision on the age of the phone number (typically older numbers are likely to be more trustworthy)

You should treat the customer's mobile phone number as having value. As such, when displayed (for example as confirmation of the number an SMS will be sent to), the number should be partially masked so that it's of no value to an attacker, but still recognisable by the customer.

The organisation should identify, and if appropriate take action, where the same mobile number appears to be in use by different customers.

This could be an indicator of fraudulent use or could be the reallocation by a telecoms company of an old number. In the latter case it is important from a reputational risk perspective not to send the former holder's message to the new owner of the number.

SMS attacks and compensating controls

There are increasing numbers of attacks on SMS, where it's being used as part of a high value business process, such as sending a passcode to a banking customer.

Attacks on SMS typically see ‘take over’ of the phone number, or the International Mobile Subscriber Identity (IMSI), a globally unique code that identifies a mobile network subscriber. This allows the attacker to receive, and potentially reply to, SMS messages intended for the genuine customer.

This section outlines the methods used for some of the most common SMS attacks. It also suggests controls that could be applied to reduce the risk of compromise and increase the chance of detection.

1. Defend against SIM Swaps

Attackers use social engineering to convince mobile phone retailers into transferring a genuine customer’s phone number (MSISDN) to a new SIM. They are then able to receive all SMS messages (and voice calls) sent to that customer’s phone number.

Suggested Controls

A number of telecoms companies, SMS aggregators and other commercial organisations offer the ability to query the mobile networks to ascertain if the SIM has been (recently) swapped.

Prior to sending a high risk SMS, organisations should gather SIM swap data from mobile networks. With this data, a risk based decision can be made on whether it is ‘safe to send’ the message. Such a decision is likely to be based on the elapsed time since the swap. The shorter the time, the higher the likelihood of a compromise.

Points to note

- SIM swapping does not necessarily mean that an attack is in progress, for example a SIM will be swapped when a customer changes handset to one using a different physically sized SIM.
- Not all networks offer this data. This is particularly true of second tier operators who do not run their own infrastructure. These Mobile Virtual

Network Operators (MVNOs) rent capacity from Mobile Network Owners (MNOs)

- Not all networks return a time stamp for the SIM swap event.
- These and other information requests from networks are likely to carry a fee.

2. Defend against SS7 attacks

Signalling System 7 (SS7) is the ‘glue’ that allows mobile networks to operate. It also permits separate networks to interoperate - for example when a handset from *operator A* is roaming on *operator B*’s network, in a different country.

SS7 is a legacy protocol and was not built with the current interconnected and adversarial world of cyber security in mind. In particular, the use of SIGTRAN (SS7 over IP) has given many more actors, both good and bad, access to SS7.

There have been a small number of attacks where bad actors have manipulated the SS7 network to make it appear as if the customer’s device was roaming on a foreign network and are thus able to receive SMSs intended for the genuine customer. These attacks are targeted at specific users and are typically short in duration - long enough to capture the SMS and short enough to avoid detection.

Whilst there is debate about how easy it would be to carry out SS7 attacks at scale, the possibility cannot be ruled out.

Suggested Controls

A number of telecoms companies, SMS aggregators and other commercial organisations offer the ability to query mobile networks to ascertain if the phone (IMSI) is roaming. Roaming is a potential IOC (indicator of compromise) but could also be legitimate.

Enterprises should consider gathering roaming data from mobile networks prior to sending a high risk SMS.

- The SS7 attack on a given subscriber is likely to be short-lived, so there could be a ‘race condition’ between the enterprise and the attacker. It may therefore be prudent to check roaming status a second time, soon after the first - probably immediately after the message has been sent.

- The subscriber may be legitimately roaming. If the roaming check is positive, enterprises could adopt one of the following strategies using a risk-based approach to how ‘deeply’ this is executed.

Option 1: Do not allow the high risk SMS when the IMSI is roaming

This could have a big impact on customers who are genuinely roaming. However, the number who are roaming and need to use the function supported by SMS is likely to be low. So, this may be an acceptable trade-off.

Note: *Option 2 will only be possible where the customer is using an application supplied by the enterprise on the device which receives the SMS*

Option 2: Geolocate the handset using the App

The logic for this approach could work in one of two ways.

Compare the geolocated country to it's roaming status:

- If the handset is geolocated in its expected home country but reports itself as 'roaming,' this is almost certainly a fraud (SS7) attack
- If the handset is geolocated outside of the home country and reports itself as 'roaming,' this is likely to be a genuine ‘roam’

Or for greater assurance of the location of the handset, compare geolocated country with roaming country:

- Query the network for roaming country information
- If same as geolocation then likely to be legitimate
- If different to geolocation then likely to be an attack

3. Defend against Malware attacks

There have been a number of reported cases where the handset has been infected with malware which receives the SMS (hiding it from the genuine customer) and forwards it to the attacker.

Suggested Controls

Occurrences of SMS-redirecting malware are relatively low, however enterprises should implement best practice in helping customers protect themselves as much as possible.

Enterprises should undertake customer security awareness activity.

Enterprises which have an app on the device should consider monitoring for jailbreaking/rooting. If a device has been jailbroken/rooted, a risk-based decision could be taken not to trust high value SMSs sent to it.

General purpose SMS protection techniques

This section outlines measures and techniques which can be used to protect SMS messages, and the processes which rely upon them.

1. Use data from the device to make risk-based decisions on the 'safety' of SMS

Smartphones are able to provide potentially useful data which can help in assessing the risk of the use of SMS. Two particularly useful items are Device ID and the physical location of the device.

Device ID

If the content from the SMS (e.g. a onetime code) is being entered on a device which you have previously tied to the intended customer, then it is highly likely that this is the genuine use. Depending on the risk of the activity being undertaken and the amount of user interaction desired, enterprises may decide that simple matching of the identifier is sufficient proof of device possession, without the need to send an SMS.

If your enterprise has had a previous 'relationship' with the customer's device, you could store its unique identifier on the device *and* 'server side'. For example, when the user signs up for an account, or during a previous transaction you make a record of an Android device's IMEI.

If the content from the SMS (e.g. a onetime code) is being entered on a device which you have previously tied to the intended customer, then it is highly likely that this is the genuine use.

Note: *Due to privacy concerns, it is possible that Google may deprecate access to the IMEI in a future OS releases. This being so, you should store a different unique identifier in the Android Keystore.*

Use geolocation/geofencing

The physical location of the device can assist in determining whether the content of an SMS is being used by the intended individual.

For example if your organisation stores address details for its users, proximity to these addresses (within a defined 'geofence') would be a good indicator that the SMS had been delivered to the legitimate device and user.

Geolocation could be used in combination with a query of the device's roaming status. Impossible combinations, would indicate compromise.

Note: *Geolocation is not infallible, particularly on Android, as it is technically possible for a bad actor to spoof geolocation coordinates. Dependant on the use case, you may not wish to prevent SMS use outside a given 'geofence'. You could instead, use the location information as part of a more comprehensive risk scoring approach.*

2. Ensure third parties protect SMS too

Many enterprises rely on third parties to operate high value, high risk SMS services. An example of this would be a bank using SMS one-time passcode as part of 3D-Secure (Verified by Visa, Mastercard Securecode etc).

You should make sure that these 3rd party vendors apply appropriate controls to identify/prevent fraudulent use of SMS.

3. Understand your organisation's SMS integration points

Many third party organisations offer services for sending SMSs on behalf of enterprises. Not all of these have direct connection relationships with Mobile

Network Owners (MNOs) and use one or more fourth parties to pass SMS traffic to MNOs.

In general, so called *Tier 1 SMS aggregators* who have a direct connection with MNOs reduce the risk of SMS compromise. These aggregators can ask MNOs to block messages which have a sender claiming to be from the enterprise but which have not originated from the aggregator.

PUBLISHED

6 November 2019

REVIEWED

6 November 2019

VERSION

1.0

WRITTEN FOR

Small & medium sized organisations

Public sector

Large organisations

Cyber security professionals