Messaging, Malware and Mobile Anti-Abuse Working Group

M3AAWG Mobile Messaging Best Practices for Service Providers

Updated August 2015

For information on M3AAWG and other mobile-related anti-abuse work, please see www.m3aawg.org.

Table of Contents

1 Introduction

1.1 Objectives

The objective of this document is to provide a set of industry best practices to help mitigate the abuse of mobile messaging (i.e., SMS, MMS and RCS) and connected services. Specifically, this document defines a set of voluntary best practices for service providers and vendors to assist in the creation and maintenance of the highest practical levels of trust and security attainable in an open, globally-interconnected messaging environment.

1.2 Scope

The scope of this document is Mobile Messaging Services, including mobile SMS, MMS, RCS and connected messaging services such as landline SIP services that use a unique E.164 [E.164] phone number as an account identifier and network address. It includes Person-to-Person (P2P) as well as Application-to-Person (A2P) messaging streams.

1.3 Document Introduction and Overview

Text messaging services have, compared to email and other Internet-based messaging services, been lightly abused until recently. However, as abuse has increased, many service providers are struggling to devise an appropriate set of defenses, especially since there is relatively little published experience in this area.

This document is designed to help the reader identify defense mechanisms for protecting text messaging services. It begins with a text messaging service description. Next, it briefly discusses key security threats. Following this, there are sections outlining three key areas to consider in defense design:

- Service design
- Technical defense
- Collaboration with other ecosystem members

Text messaging shares many characteristics with email and other open messaging services. Many of the defense techniques discussed in this document are borrowed from more mature domains such as email and webmail, where the techniques developed from years of experience have proven useful. However, there are several significant service characteristics (e.g., addresses, authentication and media) that are important in text messaging abuse and defense. These differences are identified in the service description section and are taken advantage of in the defense recommendations of the service design, technical defense and collaboration sections.

2 Text Messaging Service Description

Text Messaging Services is a set of interoperable services that permit the exchange of messages with arbitrary content (i.e., not only plain text in the traditional sense) between designated endpoints – there are over 4 billion users worldwide. Text messaging began as a store-and-forward service capable of sending up to 128 byte text messages between mobile phones and has expanded to cover additional endpoints, such as landline phones and PCs, and extended message sizes and media types.

Technically, some text services (e.g., SMS and MMS) are store-and-forward, while others (e.g., RCS' video calling) utilize end-to-end sessions. Text messaging includes both one-to-one and one-to-many messaging between endpoints routed across the global telephone network based on E.164 [E.164] phone numbers.

2.1 Addressing

The address space includes:

- E.164 telephone numbers, typically uniquely assigned to a person and/or an organization
- Private numbering spaces, such as Short Codes (typically usable only within a country and/or a service provider's network)

Phone numbers are, in most jurisdictions, a scarce resource. Therefore they are only rarely permanently assigned and are often shared. Typically, phone numbers are indefinitely leased from a service provider and may be reassigned to a new account within days of the lease termination. While in most cases an end-user device is assigned a unique telephone number, some interconnected applications use pools of phone numbers shared by multiple user accounts. In these cases a non-phone number identity is typically mapped to a conversational thread defined by a pair of phone numbers, one or both of which may not be unique to a user.

Private numbering spaces such as Short Codes are typically used for the convenience of dialing fewer digits or to facilitate types of messaging such as application-to-person (A2P) messaging that may be incompatible with a mobile endpoint.

2.2 Endpoints and Protocols

While the most common endpoint is an end-user device assigned a unique telephone number, some interconnected applications use pools of phone numbers shared by multiple user accounts. The types of endpoints and interfaces include:

- Mobile phones with traditional SMS messaging over SS7-based signaling
- Mobile phones with SIP messaging
- Landline phone-type devices (which may not have an IP address)
- Desktop computers
- Email gateways
- Web gateways
- Voice gateways
- Applications
- Social messaging gateways
- Other endpoints

2.3 Message Forwarding and Cycles

Some telephone networks, as well as many end-user devices and connected services, support forwarding of text messages, possibly to multiple destinations. Text "forwarding" is often (i.e., in the case of SMS) implemented by creating a new message with an identical body. However, envelope information such as an immutable message ID, path information and the original sender's address may be lost. As a result, it is possible to create infinite loops that can be difficult to control.

2.4 Interconnection Models

The text messaging network is loosely structured and is composed of a number of direct inter-carrier links with a large number of aggregation and interchange hubs. The predominant interconnection types are:

- A service provider using one of several available inter-carrier vendors (ICVs) to act as its message transfer point based on multilateral agreements
- Multiple service providers using a single ICV based on multilateral agreements
- Service providers directly interconnected pursuant to bilateral agreements
- Aggregators providing smaller service providers and/or A2P end users with access to the inter-carrier network

2.5 Routing

Numbering databases that are shared with telephony services are used for routing SMS messages. In many national networks, auxiliary databases such as E.164 Number Mapping (ENUM) and the National Portability Administration Center (NPAC) are widely used for inter-carrier routing. ENUM lookups are used to find the destination carrier that serves a destination phone number.

Text messages may traverse one or more hops between service providers. Next-hop routing is typically controlled by a service provider's manual routing tables. Text message routing lacks the uniform automatic routing that is characteristic of IP networks and the Internet.

Messages to and/or from private address spaces are not globally routable and may only be routed between service providers by special agreement.

3 Threat and Counter-Threat Strategy

3.1 Threat Summary

The threats to messaging services and their users are many, including denial of service, unintentional flooding, theft of service, spam, phishing, fraud, malware distribution and cramming. Sources of abuse include attacker-purchased devices and services (including those fraudulently obtained); stolen computer or Web accounts; and infected, misconfigured or compromised systems and devices. For a more detailed threat description, please see Operation Safety-Net: Best Practices to Address Online, Mobile, and Telephony Threats jointly published by the London Action Plan and M^{3}AAWG [OSN].

As of 2014, the most common type of mobile message is SMS wholly or partially transported over SS7 signaling networks. However, mobile messaging (and particularly SMS) is facing a number of major developments:

- Increasing market share of smartphones
- Broader adoption of IP-based messaging
- Increased availability of free or low-cost services
- Increasing anonymity provided by Internet-based access to services
- An increase in application-to-person (A2P) services in E.164 numbering spaces previously occupied exclusively by person-to-person (P2P) streams

This combination of increased accessibility of mobile messaging networks and endpoints that are more easily compromised contribute to greater aggregate threat exposure. While mobile malware presents a sizeable potential threat, especially in markets where smartphones are typically rooted or jailbroken, the predominant messaging threats as of 2014 are originating from attacker-owned mobile devices and non-mobile accounts.

Two key factors promoting the growth of text messaging abuse are the increased accessibility and anonymity provided by Internet-based user accounts.

3.2 Defense Pragmas

To mitigate exploitation, service providers need to:

- ***Develop agile and robust defenses.*** Most attackers are able to rapidly change their attack methods and defenses must adapt or fail. Much of an attacker's energy is spent in crafting assaults and circumventing defenses. Similarly, defenders must continue to evolve and develop their defenses. The contest between attacker and defender depends largely on the ability of one actor to out scale the other, forcing one's opponent to waste energy and resources in an unsustainable way. Defenses that require substantial effort in adapting to simple changes in attack patterns are unlikely to be successful. Mobile providers must endeavor to create automatically adapting defenses and to strategically build their resistance around the attack characteristics that are the most difficult or expensive for attackers to change.
- ***Increase apparent defense complexity and minimize intermediate feedback***. It is much more difficult for attackers to analyze and evade a series of defenses if they cannot be decomposed into individual layers and algorithms. The value of this approach becomes apparent in considering how we debug programs: Coders normally decompose complex programs into simpler, independently- analyzed routines and then observe intermediate monitoring points. Do not give attackers more information than necessary to resolve false positives. Maximize the effort that an attacker must spend to decompose the defense layers and circumvent the defenses.
- ***Decrease motivations for, and achievability of, abuse goals.*** Nearly all attacks are intended to achieve an identifiable goal (e.g., sending messages to phish for bank account credentials) and are constrained by resource limitations (e.g., the number of infected devices under an attacker's control). If defenders are able to reshape the ecosystem such that a specific attack becomes impractical, the attacker will likely stop or find a different service to attack. If attackers are able to continually adapt and circumvent service provider defenses, service providers operating under governmental regulations may wish to consider petitioning their regulators for authority to shape services in ways that disrupt an attacker's ability to successfully monetize sending large volumes of unwanted messages, making abuse unprofitable.

4 Service Design Practices

In designing a messaging service, it is highly advisable to analyze – and to attempt to minimize – opportunities for potential abuse. Services can be designed with flexible restrictions, such as selective use of CAPTCHA, that are based on potentially suspicious behavior relative to an account's normal or historical baseline. Baseline characteristics can include elements such as message rates, ratio of origination to termination, and other statistical indicators of potential abuse. This section lists a number of considerations and techniques to help design a service that will make it difficult for an abuser to send a high volume of unwanted messages.

4.1 Message Quarantine

Message quarantines that allow for delay and retrospective processing, as well as user-accessible "spam" folders, may be used to improve defense accuracy. However, most mobile messaging clients do not provide native access to a spam folder. External Web-based adjunct spam folders may be provided, but are not commonly available.

Where used, spam folders may have user-interface features that allow recipients to challenge mistaken spam classifications.

4.2 Application-to-Person Messaging from E.164 Phone Numbers

Both globally dialable E.164 and service-provider-network local addresses (e.g., Short Codes) are technically capable of sending and receiving messages. However, many networks and services are incompatible with A2P messaging that originates from or terminates to E.164 addresses. One of the key technical issues behind this incompatibility is protection from A2P spam. Many network anti-abuse defenses, developed around email and other services, are designed to prevent high-volume, bulk, unwanted or predominantly unidirectional messaging – attributes that are uncharacteristic of P2P messaging.

Additionally, many service providers have user agreements and/or inter-service-provider agreements that prohibit A2P messaging. For these reasons, it is a best practice to NOT use E.164 phone numbers for A2P messaging unless it can be verified that such activity conforms to the agreements and policies of affected users and service providers.

Additionally, as discussed elsewhere in more detail in this document (see both 4.8.1 Opt-In and 4.9 Spamprogation Policy), it is a best practice to ensure that application-originated messages are sent with the appropriate consent of all recipients.

4.3 Account Creation

Account creation policies are an important part of preventing abuse. This includes both devising appropriate policies and analyzing abuse trends. Tracking features or characteristics of account creation events – and looking for patterns that can be correlated with abuse – should be an ongoing activity. That data can then be used to take action against maliciously created accounts and for predicting future abuse before it occurs. Examples of account creation policies that can help limit abuse include:

- ***Preventing automatic account creation.*** Given the person-to-person nature of text messaging, it is appropriate to ensure that the registrant is a person, not an automaton. Challenges such as CAPTCHA, puzzles or questions that separate human from automata are recommended. In implementing such challenges, it is important to consider accommodations for the disabled.
- ***Requiring and recording user identification.*** Secure authentication is preferred; for example, using secure credentials such as a government-issued identity document containing the user's photo. Ideally those documents should be directly confirmed against online data to preclude use of forged identity documents. Identities can also be tied to other, weaker, identifiers. For example, accounts can be tied to a user's phone number via voice, touch-tone or text validation. Tying a user's new account to their email is weaker still.

  In descending order of strength, authentication processes include:

  - Secure user identification (e.g., government-issued identification); if practical, confirmed against an authoritative online database
  - Account tied to a valid credit or debit card
  - Account tied to an existing phone number
  - Account tied to an existing email account

- ***Collecting and validating contact information,*** such as the customer's claimed address.
- ***Setting account privileges commensurate with, and tailored to, the strength of the user identification and validation.***
- ***Creating a unique account for each user.***
- ***Obtaining user consent in an End User License agreement during the account creation process*** that explicitly permits the provider to:
    - Filter messages, where needed
    - Restrict inappropriate use (e.g., bulk messaging, A2P messaging, annoying other users, disrupting communications, illegal activity or sending solicitations without consent)
    - Account termination/suspension/restriction as appropriate when needed to control abuse
    - Permission to share reported abuse data with other parties, subject to appropriate restrictions (for example, anonymized data may be shared with an originating service provider)
- ***Limiting the number of accounts that can be created based on similar identities within a short time period.*** "Similar identities" might be defined to include the account creation IP addresses, phone numbers, single sign-on (SSO) provider account names, credit card numbers or email addresses. Be mindful that rate limiting based on IP address may more severely impact users who are connected from behind proxies and NAT, including so-called "carrier grade NAT." Requiring a validated credit card provides some level of protection against unlimited account creation, but stolen credit cards can still be used.
- ***Limiting the number of messages that a new account can send.*** For example, limit new accounts to 20 messages in the first 24 hours after they are created.
- ***Monitoring and restricting new accounts' sending patterns,*** such as accounts that predominantly send rather than receive messages, and adaptively restricting sending with message limits and/or per-message CAPTCHA.
- ***Limiting the number of recipients per message*** to a reasonably low number, e.g. 20 or 30 rather than 1,000 per message.
- ***Using the age of an account as a characteristic in determining trust.*** Newer accounts may be more likely to have been created for malicious use than established accounts that have been sending for a significant period without complaints or other abuse indications. Note that this may not be true of valid accounts that have been recently compromised.
- ***Monitoring black markets for the sale of bulk accounts.***

4.4 Login

As with account creation, it is important to have a set of security policies and mechanisms in place at user login. The following guidelines and policies are applicable:

- ***Each session must have a secure user authentication mechanism,*** such as a password, but where appropriate including client certificates, tokens, biometrics or other multifactor authentication mechanisms.
- ***Authentication is required for each session*** and may be handled by application-layer or lower security layers (e.g., GSM/UMTS/LTE registration) with strong authentication.

- ***Login may have flexible and dynamic security features*** and may include additional tests as needed, such as IP reputation, history and CAPTCHA.
- ***Limit the number of login attempts that can be initiated from similar sources*** and against a given account within a short time period to deter brute force attacks. Be mindful that rate limiting based on IP address may more severely impact users who are behind proxies and NAT, and that non-malicious applications configured with outdated passwords may execute retries.

4.5 Message Origination

As with account creation and login, message sending policies are also needed to mitigate abuse. A few guidelines follow:

- It is a best practice for service providers to include in their terms of use an obligation for all senders to obtain explicit written recipient consent prior to any automated or commercial message origination. Note that this typically requires senders who use mailing lists to access and use deactivation lists to prevent sending of messages to numbers reassigned to a non-consenting recipient.
- Clearly identify the originator by MSISDN or Short Code in 'from' address fields wherever possible.
- Spoofing should not be permitted in mobile messaging unless lawful (as is permitted in some jurisdictions), and explicitly agreed to by both originating and terminating service providers.
- Restrict message origination rates and numbers of recipients to levels consistent with the authorized use, account history and potential for abuse. CAPTCHA may be applied uniformly or selectively to deter abuse.
- Do not allow multiple account holders to share pools of one or more MSISDNs for message origination unless the recipients of such messages have consented and have been provided with a suitable opt-out mechanism. This mechanism should allow a single opt-out action to apply to the entire pool of originating MSISDNs.

4.6 Message Monitoring and Control (in Origination, Transport and Termination)

The best practices in this section apply to three phases of messaging: 1) Origination in the service provider environment, 2) Transport within and between service providers, and 3) Termination.

Where permitted by law, and with any necessary subscriber consent, a combination of content (as permitted by law), volumetric statistics, account history, location, strength of authentication, CAPTCHA and/or other methods should be used to monitor, detect abuse and police the messaging stream.

Where practical, provide recipients with some control over spam filters; for example, personal block and allow lists. The service provider should ensure that features necessary to protect the service, as opposed to safeguarding only the subscriber, may not be disabled by subscribers.

4.7 Automated Accounts – Verification and Mailing List Maintenance

Text messaging accounts are identified by phone numbers, which are frequently recycled. One cannot assume that a phone number is assigned to the same person that owned it a year ago. In situations where consent of the recipient is needed or personal information is transmitted, it is important to maintain mailing lists.

It may be necessary to have the consent of the current owner of a recipient phone number to legitimately message them. In this case, there needs to be a notification mechanism advising of recycled phone numbers (i.e., account deactivation feeds). Before sending potentially sensitive personal information, it may be necessary to revalidate the recipient before sending.

Additionally, it is important to provide an unsubscribe mechanism. It is a best practice to notify users how they may unsubscribe (e.g., by replying with the word "STOP") and to process deactivation requests and unsubscribe feeds within 24 hours.

4.8 Opt-In/Opt-Out

Services that support bulk messaging should require explicit, written recipient opt-in and permissive (easily accomplished) opt-out. This section discusses best practices for both opt-in and opt-out.

4.8.1 Opt-In

Opt-in mechanisms should be provided for subscribed or automated messaging, whether commercial or non-commercial.

The opt-in process should require the subscriber to prove control of the destination account. It should be considered that in many networks it is possible to spoof an originating address. Therefore the preferred confirmation method is to require a subscriber to prove receipt of a communication sent to their text messaging address. For example, this can be accomplished by originating a subscription confirmation code communication directed to the subscriber's text messaging address and requiring a confirmation that proves receipt, either by reply or by entry on a website.

Note that opt-in mechanisms that originate calls, messages or invitations may be abused to monetize, harass or deny service. They should be protected from automated triggers by CAPTCHA and/or sender and recipient rate limits, as needed to minimize abuse potential. Additionally, defensive provisions may be needed to protect a service provider from inadvertently incurring charges as a result of being asked to send to premium phone numbers.

In networks where spoofing is difficult, a simple subscription message, such as "START" from an authenticated SMS client, is sufficient.

4.8.2 Opt-Out

Automated services should provide an opt-out mechanism accessible to subscribers from a text messaging client. Subscribers should be able to opt out of any subscription by replying with "STOP" or similar text from the text messaging client. Additional opt-out mechanisms (e.g., a website unsubscribe page) are acceptable provided that authentication and measures to prevent harassment and DoS attacks from unwanted traffic to the subscriber are employed.

Any services that utilize pools of sending addresses should ensure that a single opt-out will apply to the entire service, not just messages from a single pool address.

4.9 Spamprogation Policy

While it is reasonable for one person to independently decide to send a message to a friend recommending an app or service, some otherwise legitimate applications may spam a user's entire list of contacts – or even spam random lists – with invitations purportedly sent by the device's owner and intent on convincing the contacts that the owner wants them to use it. This behavior may even be concealed from, or not easily restricted by, a sending device's owner. Both the sending of unwanted messages and lack of adequate user control are not acceptable.

The legality of propagating endorsements via spam (spamprogation) might rely on whether the primary purpose of the message is commercial solicitation or if it is the users' personal desire for enhanced social contact with people they already know. Complicating this for service providers, who need to decide

whether or not to permit these activities or to block such messages, are six aspects that have impeded the classification of spamprogation messages:

- Whether these constitute automated or bulk messaging sent from individual accounts
- Compliance with terms and conditions of service (e.g., permissibility of gatewaying, relaying, automated, and/or bulk messaging)
- Who is the sender – is it the app, the account holder, or both?
- Whether the messages' primary purposes are commercial or non-commercial
- Whether the account holder is aware of the contents and the recipient of each message that is sent
- The presence or absence of recipient consent

While most messages may be sent without the recipient's permission, there currently are no practical means that a service provider can use to discern consent to receive a specific message. Unfortunately, spam filter technology is not likely to advance to the point that it can accurately discern recipient consent in such a distributed sending environment.

While some jurisdictions and services may be subject to laws, regulations and guidelines that are more restrictive, the following best practice is the minimum recommended: In cases where there is no explicit recipient consent, invitation messages should be sent only if all of the following conditions are met:

- Senders are aware of and consent to the application or account accessing their contact lists.
- The sender has taken some explicit action to authorize or initiate sending, such as checking a box, for **each** recipient. Any mechanism that includes a 'select all' capability does not comply with this guideline.
- The sender is aware of the content that will be sent.
- The sender's primary intent is non-commercial; for example, enhanced social interaction.

4.10 Privacy

Service providers should publish and follow privacy policies consistent with applicable laws and regulations.

4.11 Spoofing

A spoofed identity is an identity (in this case, typically a phone number) other than the identity assigned, possibly with layers of delegation, by the service provider of record to the applicable end user account.

Some spoofing is well intentioned and benign, such as a person using their wireless number to originate calls from a "unified messaging" VoIP service. In other cases, it may include toll fraud, impersonation of a specific party, and arbitrary or random identity assumption to cloak one's own identity.

Where permitted by law, it is a best practice to not use spoofed identities in text messaging. While permissible in some jurisdictions, spoofing should not be permitted in mobile messaging unless lawful and explicitly agreed to by both originating and terminating service providers.

Further it is a best practice for a service provider to prevent spoofing of identities in accounts under its control unless sufficient evidence of authority to spoof has been presented. In these cases it is important to understand that text messaging identities (i.e., phone numbers) are leased identities and are NOT permanent; phone numbers are routinely recycled to other parties. In general, there need to be mechanisms in place that ensure the authority to spoof has been issued by the current holder of the spoofed identity. This might be accomplished by exchange of provisioning events, such as account cancellations, or other mechanisms. Relying on a one-time verification of authority from some time in the past will generally be insufficient.

4.12 Forwarding and Gatewaying

Some services or phone numbers may not have traditional text messaging clients. For these and other reasons, it may be necessary or pragmatic to:

- Forward messages to another phone number
- Gateway messages to some other service (e.g., voice, email)
- Not deliver messages
- Provide feedback to the sending service provider

Text messaging, as widely implemented, lacks extended features that are needed for seamless forwarding, including gatewaying to other services. Specifically, there are no explicit loop detection and forwarding chain recording features.

For these reasons, automatic forwarding may lead to message loops and overload and to unexpected behaviors, such as replies from an unexpected address breaking validation mechanisms and/or thread associations. Industry standards work is needed to facilitate safe forwarding. The following guidelines are recommended:

- It is a best practice to limit automatic forwarding to those destinations authorized by both the forwarder and forwardee accounts.
- Except in special cases where network and/or user agent features can be depended on to prevent forwarding loops, any network that supports forwarding should, where practical, employ loop-detection mechanisms.
- To control and break forwarding loops, it is a best practice for text messaging services to restrict the rates at which an account may send and/or receive messages.
- It is a best practice to NOT permit a user account, user-controlled device or application to act as a message relay or gateway.

4.13 Ack vs Nak

The question of whether to acknowledge (Ack) or reject (Nak) unwanted messages that are rejected or blocked by a service provider's anti-abuse defenses is one that has been extensively discussed in the context of email. Reasons for rejection might include excessive message rates (i.e., volume), invalid addresses, unacceptable content, and of course, other circumstances. There are advantages and disadvantages to both policies and there is no single policy that is best across all service provider environments. Instead, a policy should be formulated for each situation based on an awareness of the impact of that situation. There must be a balance between the advantages of transparency in resolving delivery issues and resistance to disclosing defense policies to attackers. Here are several key considerations:

- Naks may provide an attacker with information that is helpful in improving attacks. They may provide evidence of:
    - Rate (i.e., volume) limits
    - Invalid addresses
    - Blocked content
- The transparency provided by Naks might help rapidly resolve false positives and other delivery issues, including inactive or incorrect addresses.

5 Technical Defense Practices

5.1 Anti-Spoofing

To help combat spoofing, receiving service providers may develop anti-spoofing defenses based on ingress points, the reputation of the originating service provider, whether or not a specific message's originating party's address is served by that originating service provider, and the existence of agreements that authorize spoofing.

These defenses may cause messages with spoofed identities to be blocked, delayed or subject to quarantine. They may cause indications of an unvalidated sending identity by the receiving service provider (e.g., if they do not originate from the sending MSISDNs ENUM provider listing).

Receiving service providers should, to the extent practical, present an end-user indication of trust in the sender's identity, if applicable.

5.2 Anti-Phishing

Text messaging anti-phishing best practices are quite similar to those of other messaging areas such as email. While the key defining characteristic of phishing is social engineering, mitigation efforts are generally focused on technical means to prevent spoofing and to prevent the phishing message from reaching the end user. Such efforts may be complicated by privacy concerns and regulatory constraints.

Authentication of users and message origination are the first opportunity for mitigating potential phishing-related abuse. Conscientious operators prevent malicious messaging from being emitted though their systems, whether intra-system or for delivery to other operators. Operators should provide mechanisms for end users to report text phishing in a manner that captures relevant details, including the originator of the message.

Operators may be constrained in blocking text messages due to regulatory requirements. Where possible, blocking should be initiated based on unique characteristics of the message such as a malicious URL. In addition, blocking of access to the malicious URL at the network and/or DNS levels should be initiated as well as submission to browser block feeds. Where sources of phishing can be identified, blocking of the source should be considered, if allowed, subject to contractual and regulatory constraints and where impact to legitimate users is minimal or non-existent.

Service providers should include provisions in their terms of service and contracts that allow a necessary and sufficient range of actions in monitoring messaging and in preventing and mitigating phishing abuse.

5.3 Phone Number Reputation and Blocklists

One nascent defense deployed against unwanted messaging is blocking of messages based on sender reputation. Lists of barred phone numbers are used to block any message that appears to originate from a barred sender.

Blocking policies may be enforced on the network and/or through user agents such as downloadable third-party messaging user agents. They may be scoped to protect an individual (e.g., personal blacklists) or the entire network.

As of the writing of this document, few phone-number based blocklists are freely available. Most are proprietary in content, format and/or distribution protocols and mechanisms. M^{3}AAWG encourages

service providers to consider developing and/or offering their customers the ability to use phone-number based blocklists if they do not already. Any offered blocklist should provide end-users with a mechanism to remove inappropriately included numbers from the blocklist.

5.4 Content Filters

Content-based filters examine the contents of messages, attempting to identify malicious or unwanted content based on a wide variety of methods, including the presence of key words or known malicious URLs. To be useful, content filters must block most spam while blocking almost no "ham," or legitimate content. They typically contain necessarily proprietary technology, a "secret sauce" unknown to attackers. Content filters are typically relatively expensive to develop and maintenance generally requires a feed of new rules and fingerprints. They are widely used in messaging defense and network-based solutions are commercially available from several vendors.

Unlike many defense mechanisms that take enforcement actions against accounts, content filters enable enforcement actions (such as traffic blocking) to be taken against individual messages. Content filters are typically effective in identifying all accounts sending similar messages. They are most effective in situations where similar content is used to originate attacks from a large number of (possibly spoofed) addresses (e.g., botted phones). They are least effective in cases where rapidly changing content originates from a smaller set of more stable addresses. In many environments, abuse may be adequately controlled without content filters.

6 Abuse Detection, Analysis and Mitigation Practices

6.1 Analytical Repository

In cases where messages are blocked, it is important to understand why. If messages are mistakenly blocked by content filters, analysis of mis-blocked messages may be the only practical means to identify and rectify the issues. For these reasons, especially where content filters are used, it is a best practice to maintain a repository or log of blocked messages, provided that doing so is consistent with applicable laws, user agreements and service provider policies.

Message retention creates potential privacy issues that must be balanced with the need to maintain effective defenses. Important points that should be considered are the components of messages to be retained, the duration of retention and access to the retained messages. For example, if content filters are not used, it may be unnecessary to retain message bodies. Consider retaining cryptographic hashes representing message bodies or addresses instead of these components themselves. If hash-based methods are employed, retention of cryptographic hashes (e.g., the MD-5 cryptographic hash function) or fingerprints of message parts (e.g., the message body) may be an adequate substitute for full message content.

Another option is for the repository to only store messages that have abuse indications or are culled from the service provider's own spam trap accounts and/or reporting services.

6.2 User Feedback

Enabling users to report unwanted messages or other abusive activities to their service provider is a highly recommended best practice. Users, unlike most automata, are able to rapidly discern spam from "ham," even in a network where legitimate and malicious activity share similar volumetric and other behavioral characteristics. It should be noted, however, that humans are fallible and subject to error and may also be malicious. Feedback systems should employ mechanisms to identify and appropriately down-weight inaccurate user feedback. This may be accomplished via reporter reputation and correlation mechanisms. Complaint validation mechanisms (e.g., validating that the accused actually messaged the accuser) may also be used where possible.

In order to be most effective, end-user feedback should be as close to real-time as possible and should feed into a data store that can be mined for trends and patterns. This allows the service operator to see what end users are finding objectionable and to correlate the volume and characteristics of specific complaints with abuse events. These trends can then feed into message filtering, firewall or other enforcement systems and processes.

6.2.1 Designing the User Feedback Interface

In designing the User Feedback system, it is important to make the spam reporting process as accessible and effortless as practical, while minimizing potential accidental or intentionally false reports. As feedback systems may themselves be abused, user authentication is highly recommended. The system's user interface experts should determine the best way to allow users to send abuse reports. The general principles are:

- Make it easy for the end user to know how to report abuse.
- Make it easy for the end user to send an abuse report as soon as they observe the abuse.
- If reasonable, reduce the number of clicks or other actions required to send the abuse report.
- It is reasonable to ask the user to categorize the type of abuse into a small list of available categories or to click an "Are you sure you intended to report this?" confirmation button. This additional categorization information may make the abuse reports more usable, so the extra clicks and actions may be worthwhile.

6.2.2 User Feedback in the User Agent – "This Is Spam" Button

The "This Is Spam" button or link in the user agent should be positioned on the screen where the user might first see the abuse so that the user can easily report it.

In 2012 the Open Mobile Alliance (OMA) standards organization completed work on an open industry standard format and protocol for reporting SMS and MMS spam: The OMA Mobile Spam Reporting (SpamRep) V1.0 enabler [SpamRep]. This standard was developed for reporting unwanted mobile messages from a mobile device to a network, but has also been used to convey information in inter-service-provider feedback loops. The use of an open industry standard interface is especially important in the mobile ecosystem as there are a large number of service providers and user agents; multiple and conflicting proprietary interfaces do not facilitate interoperability. SpamRep is currently the sole open standards organization specification for mobile spam reporting, although a number of proprietary implementations also exist.

In systems where it is not possible to change the user agent to automatically report via a "This Is Spam" type button or link, an alternative method should be used to send an abuse report. The alternative system should try to retain as many of the original characteristics of the abuse message as possible, including associated metadata that is useful in tracing the origin of the message. For example, in email abuse reporting when there is no spam button, users will often forward messages as an attachment to a spam reporting email address. Using an attachment preserves the headers of the original email, which contain the characteristics of the abuse that may be helpful in filtering or taking down malicious accounts or predicting future abuse. In text messaging, the "This Is Spam" reporting button operates at the application level and is able to collect many abuse characteristics automatically from the phone and forward all of these characteristics in a single abuse report.

An example of an alternative system (for phones and messaging applications that do not support a "This Is Spam" button) is the 7726 reporting method advocated by M^{3}AAWG and the GSMA; 7726 spells SPAM using the letters on a normal phone keypad (7=S, 7=P, 2=A, 6=M). The 7726 reporting standard

allows handset users to forward the content of their SMS and MMS spam reports to the Short Code 7726. In doing so, some of the features of the abuse are not included in the spam report, such as the MSISDN/MDN, Short Code or email address that originated the abusive message. The 7726 system attempts to collect this piece of data by sending a response SMS to the end user's phone asking for the phone number of the sender of the malicious message. The end user can then optionally provide the sender information by typing the sender's phone number into a second text message.

The following example illustrates a spam reporting SMS conversation using the 7726 Short Code between an end user who has received spam and mobile operator "Telecom XYZ":

From:	End User
To:	7726
Message:	Instantly! Need $959 cash loan in just 2 hours without a credit check, if so call 1-800-555-1234 to apply now.
To:	End User
From:	7726
Message:	XYZ Free Msg: Thank you for reporting spam. Please reply with the spammer's "from" number or email address and help us stop spam.
To:	7726
From:	End User
Message:	212-555-8989
To:	End User
From:	7726
Message:	XYZ Free Msg: Thank you, we appreciate your assistance.

6.2.3 End User False Positive Feedback

If the system is filtering or deleting messages which are considered to be spam or abuse, then the mobile operator may also want to offer the end user access to a quarantine or spam folder with a "This Is Not Spam" button or link. This allows end users to view the messages that have been marked as abusive and gives them an opportunity to offer feedback if they consider a decision on a specific message to be incorrect or a "false positive" (i.e., they believe the message was falsely identified as abusive).

This quarantine or spam folder could be given to all users as a standard offering or it could be an optional feature that users choose to turn on if they are concerned about messages being incorrectly filtered as spam. Alternatively, it can be turned on by default for a percentage of users to encourage reviewing for False Positives based on the needs of the provider's filtering system or anti-abuse staff.

6.2.4 Not All User Feedback Is Reliable

Not all end-user feedback is reliable on its own. End users may make mistakes or have alternative reasons for reporting activity as abusive. End-user feedback should be considered within the framework of other complaints and data collected about the abuse. The development of a reporter trust rating system is useful for automatically determining valid feedback versus erroneous or malicious feedback.

6.2.5 Honeypot and Grayspace Feedback

A honeypot is an account that is created as a "trap" for the purpose of detecting, deflecting or counteracting unauthorized use of the service. It usually involves an account that should not be seeing activity, as it has been specifically created to be discovered by spammers, or should only be seeing certain limited types of activity. In this way, any activity that differs from what is expected can be treated as probable misuse and can be mined for characteristics that can be fed into abuse detection and filtering systems.

A similar concept to honeypots is grayspace and darkspace. Grayspace is comprised of active phone numbers that are incapable of receiving P2P text messages. Examples are non-SMS landlines and devices such as some tablets that have no SMS user agents. Messages to these devices may, within the constraints of applicable national laws and subscriber privacy expectations, be noted as anomalous and used to detect abusive messaging.

Darkspace is inactive or non-existent phone numbers (such as those that do not follow numbering conventions). Receipt of a significant volume of similar undeliverable and/or unviewable messages may be indicative of unsolicited messaging. In these cases it may be appropriate to analyze the content, envelope and delivery characteristics, and take appropriate enforcement actions against the identified messages and any related messages.

It should be noted that the text messaging address space (phone numbers) is different from many other namespaces such as email addresses. The address space is densely populated and addresses are subject to rapid reassignment. Phone numbers are often mistyped, so phone number honeypots will receive more accidental and non-abusive traffic than honeypots in other namespaces. For these reasons, honeypots and grayspace traps are most effective when used in conjunction with other detection techniques, including volumetric analysis and end-user feedback.

End-user feedback suffers from delays, as minutes to days may elapse before an unwanted message is reported by end users. In contrast, honeypots and grayspace traps may detect unwanted traffic as soon as it is delivered.

6.3 Complaint and Incident Management

To best deal with abuse, and minimize threats to the service and its users, a service provider should have automated systems and a process in place to deal with abuse. This should include:

- Having a (preferably automated) mechanism for receiving complaints from users. These may include reporting methods such as forwarding to a Short Code (e.g., 7726), abuse reporting webpages, buttons in user agents, abuse reporting email addresses and other mechanisms.
- Having a (preferably automated) mechanism for receiving complaints from peer service providers. These may include reporting methods such as forwarding to a Short Code (e.g., 7726), abuse-reporting websites, automated feedback loop mechanisms, reserved inter-provider abuse-reporting phone numbers, abuse-reporting email addresses and other solutions.
- Have a person on staff who can discuss incidents with peer service providers.

- Have a process and systems for tracing abusive text messages, specifically:
    - Determine whether the messages originated inside one's own service.
    - If internally-originated, identify the originating mobile device, phone number, account, geolocation and/or IP address. With some services, it may be necessary to compare the origination information to detect spoofing.
    - If externally-originated, determine:
        - The point of ingress into one's own service provider network; e.g., SS7 point code, service center, Call Session Control Function (CSCF) address
        - The peer service provider the message was received from
        - The service provider responsible for the originating phone number. Note that in some cases, a primary telephone company may resell phone numbers to other services. In these cases, it may be necessary to use a third-party lookup service and/or contact the primary telephone company for assignment information.
- Have a responsive process in place for managing incidents.

7 Collaboration and Education

7.1 Data Sharing

Data sharing between service providers generates a comprehensive understanding of the abuse in progress and provides for more effective mitigation of abusive activities.

For instance, if the abuse is generated within company A's network but is targeted toward Company Z, Company A may not be immediately aware of the situation. It is likely that the impacted end users at Company Z may be the only ones, or at least the first parties, to notice this unwanted traffic.

Company Z may have some limited ability to detect and filter the abuse coming from A, but Company A is likely the only company with the power to mitigate abuse at the source by shutting down accounts or taking other appropriate action. Information from Z might help A mitigate more of the abuse in less time. Therefore, it is generally in Company Z's best interest to share data with Company A, provided Company A can be trusted to use the data appropriately.

Other entities also potentially benefit since many abuse sources target multiple providers rather than just a single company. This means that when companies work together to defeat those abuse sources, the entire ecosystem benefits.

Data sharing has privacy implications that may need to be addressed. Privacy protections may be complex and are highly dependent on national laws; on the other hand, some national laws require data sharing. With a combination of subscriber consent, redaction and data obfuscation, some data sharing can usually occur. Cryptographic solutions, such as a one-way hash of data elements, can be used in lieu of sending message envelope or body information. This may avoid privacy issues while allowing only organizations in the original delivery path to decode an abuse complaint and identify an accused message or sender. Some reporting mechanisms, such as SpamRep, contain optional features for obtaining subscriber consent to share complaint data.

Subject to applicable laws, end-user license agreements and privacy expectations, it is a best practice to share abusive content broadly with trusted industry and government partners to facilitate blocking of messages and enforcement actions against accounts, servers, domains, call-to-action and/or other resources being exploited by abusers.

For additional information on data sharing, the IETF RFC 6449 [Feedback] is an excellent reference. Although written specifically for email, most of its principles are also applicable to text messaging.

7.2 Industry Forum Participation

It is recommended that service provider representatives with operational, product or development responsibilities for text messaging services participate in industry forums such as M[3]AAWG. Participation in the formal sessions, smaller vetted workshops and private discussions at these forums contribute to the sharing of threat knowledge and to abuse monitoring, and enhances the industry's defense resources, strategy and techniques. Such groups are responsible for much of the collaborative defense systems, efforts and liaison needed to successfully protect text messaging from abuse.

Appendix A – References and Resources

References

[E.164] List of ITU-T Recommendation E.164 Dialing Procedures as of 15 December 2011, http://www.itu.int/dms_pub/itu-t/opb/sp/T-SP-E.164C-2011-PDF-E.pdf

[Feedback] IETF RFC 6449 Complaint Feedback Loop Operational Recommendations, https://tools.ietf.org/html/rfc6449

[OSN] Operation Safety-Net: Best Practices to Address Online, Mobile, and Telephony Threats (M^{3}AAWG and London Action Plan), https://www.m3aawg.org/sites/default/files/M3AAWG_LAP-79652_IC_Operation-Safety-Net_2-BPs2015-06.pdf

[Phish] Anti-Phishing Best Practices for ISPs and Mailbox Providers, Version 2.01, June 2015, (M^{3}AAWG and the Anti-Phishing Working Group–APWG), https://www.m3aawg.org/sites/default/files/M3AAWG_AWPG_Anti_Phishing_Best_Practices-2015-06.pdf

[SpamRep] Mobile Spam Reporting Technical Specification, Approved Version 1.0 OMA-TS-SpamRep-V1_0-20120619-A–19 Jun 2012, http://technical.openmobilealliance.org/Technical/release_program/docs/SpamRep/V1_0-20120619-A/OMA-TS-SpamRep-V1_0-20120619-A.pdf

Other Resources

Abuse Desk Common Practices-MAAWG, http://www.maawg.org/sites/maawg/files/news/MAAWG_Abuse_Desk_Common_Practices.pdf

MAAWG Best Common Practices for Mitigating Abuse of Web Messaging Systems, http://www.maawg.org/sites/maawg/files/news/MAAWG_Web_Messaging%20BCP_2010-08.pdf

M^{3}AAWG Best Practices for the Use of a Walled Garden, Version 2.0, March 2015, https://www.m3aawg.org/sites/default/files/document/M3AAWG_Walled_Garden_BCP_Ver2_2015-03_0.pdf

MAAWG Email Anti-Abuse Product Evaluation Best Current Practices, http://www.maawg.org/sites/maawg/files/news/MAAWG_Anti-abuse_Product_Evaluation_BCP.pdf

Appendix B - Abbreviations and Definitions

A2P – Application to Person messaging

Bot - a remotely controlled device used for unauthorized purposes

Botting - the practice of infecting or otherwise gaining control of another party's device

Cramming – The act of subscribing an end user to a premium service for which he or she is billed without the end user's consent

ESME – External Short Message Entity

GSM – Global Standard for Mobile Communication

ICV – Inter Carrier Vendor – vendors providing connectivity between wireless subscribers, networks and services. For clarity and continuity, the historical term "*Carrier"* is used, but ICVs provide their services to all service providers.

MIN – Mobile Identification Number

MMS – Multimedia Messaging Service

MO – Mobile Originated

MS – Mobile Station

MSISDN – Mobile Station ISDN number

MT – Mobile Terminated

NANP – The North American Numbering Plan serving 20 North American NPA-NXX – represents area code and exchange of the North American Numbering Plan

P2P – Person to Person messaging

Service Provider – Any entity that makes a messaging service available to consumers through the use of telephone numbers

RCS – Rich Communication Suite

SIP - Session Initiation Protocol is a signaling communications protocol, widely used for controlling multimedia communication sessions such as voice and video calls over Internet Protocol (IP) networks

SME - Short Message Entity

SMPP – Short Message Peer-to-Peer Protocol

SMS – Short Message Service

SMSC – Short Message Service Center

SpamRep – An Open Mobile Alliance standard for reporting unwanted messages

SS7 – Signaling System 7, a telephony signaling protocol suite

Acknowledgments

Primary author:

Alex Bobotek, M[3]AAWG Voice and Telephony Abuse SIG Co-Chair, M3AAWG Board Member

Contributing authors:

Michael Hammer, American Greetings, Inc.
Angela Knox, Cloudmark, Inc.
Joe St Sauver, M[3]AAWG Senior Technical Advisor
Anonymous contributor