Characterizing the Security of the SMS Ecosystem with Public Gateways

BRADLEY REAVES, North Carolina State University
LUIS VARGAS, NOLEN SCAIFE, DAVE TIAN, LOGAN BLUE, PATRICK TRAYNOR, and KEVIN R. B. BUTLER, University of Florida

Recent years have seen the Short Message Service (SMS) become a critical component of the security infrastructure, assisting with tasks including identity verification and second-factor authentication. At the same time, this messaging infrastructure has become dramatically more open and connected to public networks than ever before. However, the implications of this openness, the security practices of benign services, and the malicious misuse of this ecosystem are not well understood. In this article, we provide a comprehensive longitudinal study to answer these questions, analyzing over 900,000 text messages sent to public online SMS gateways over the course of 28 months. From this data, we uncover the geographical distribution of spam messages, study SMS as a transmission medium of malicious content, and find that changes in benign and malicious behaviors in the SMS ecosystem have been minimal during our collection period. The key takeaways of this research show many services sending sensitive security-based messages through an unencrypted medium, implementing low entropy solutions for one-use codes, and behaviors indicating that public gateways are primarily used for evading account creation policies that require verified phone numbers. This latter finding has significant implications for combating phone-verified account fraud and demonstrates that such evasion will continue to be difficult to detect and prevent.

CCS Concepts: • **Security and privacy → Mobile and wireless security**;

Additional Key Words and Phrases: Multifactor authentication, SMS, SMS abuse, SMS spam

ACM Reference format:
Bradley Reaves, Luis Vargas, Nolen Scaife, Dave Tian, Logan Blue, Patrick Traynor, and Kevin R. B. Butler. 2018. Characterizing the Security of the SMS Ecosystem with Public Gateways. *ACM Trans. Priv. Secur.* 22, 1, Article 2 (December 2018), 31 pages.
https://doi.org/10.1145/3268932

1 INTRODUCTION

Text messaging has become an integral part of modern communications. First deployed in the late 1990s, the Short Messaging Service (SMS) now delivers upwards of 4.2 trillion messages around the world each year (The Open University 2014). Because of its ubiquity and its perception as providing a secondary channel bound tightly to a user's identity, a range of organizations have

This work was supported in part by the U.S. National Science Foundation under Grants No. CNS-1526718, No. CNS-1464087, No. CNS-1540217, No. CNS-1542018, and No. CNS-1464088.
Authors' addresses: B. Reaves, 890 Oval Drive, Raleigh, NC 27695-8206; email: bgreaves@ncsu.edu; L. Vargas, N. Scaife, D. Tian, L. Blue, P. Trayor, and K. R. B. Butler, E301 CSE Building, PO Box 116120, Gainesville, FL 32611; emails: {lfvargas14, scaife, daveti, bluel, traynor, butler}@ufl.edu.

2471-2566/2018/12-ART2 $15.00
https://doi.org/10.1145/3268932

implemented security infrastructure that take advantage of SMS in the form of one-time codes for two-factor authentication (Aloul et al. 2009; DeFigueiredo 2011; Duo Mobile 2015) and account validation (Thomas et al. 2013).

The text messaging ecosystem has evolved dramatically since its inception and now includes a much wider range of participants and channels by which messages are delivered to phones. Whereas phone numbers once indicated a specific mobile device as an endpoint and were costly to acquire, text messages may now pass through a range of different domains that never touch a cellular network before being delivered to a non-cellular endpoint. Moreover, these systems allow users to send and receive messages for free or low cost using numbers not necessarily tied to a mobile device, specific geographic area, or even a single customer. As such, they violate many of the assumptions upon which the previously mentioned security services were founded.

In this article, we perform the first longitudinal security study of the modern text messaging ecosystem. Because of the public nature of many SMS gateways (i.e., messages are simply posted to their websites), we are able to gain significant insight into how a broad range of companies are implementing SMS-based services as an important part of their security infrastructure. Moreover, these systems allow us to see the ways in which defenses such as phone-verified accounts (PVAs) are successfully being circumvented in the wild. Our work makes the following contributions:

- **Largest Public Analysis of SMS Data:** While others have looked at aspects of SMS security in the past (Delany et al. 2012; Dmitrienko et al. 2014), ours is the largest and longest study to date. Our analysis tracks over 600 phone numbers in 31 countries over the course of 28 months, resulting in a dataset of 900,655 messages. This dataset, which is double the size of our previous study (Reaves et al. 2016a), allows us to reason about the messaging ecosystem as a whole and allows us to determine whether previous observations represent steady-state problems or are instead temporary issues.
- **Evaluation of Security Posture of Benign Services:** We observe how a range of popular services use SMS as part of their security architecture. While we find many services that attempt to operate in a secure fashion, we identify a surprising number of other services that send sensitive information in the clear (e.g., credit card numbers and passwords), include identifying information, and use low entropy numbers for their one-use codes. Because there is no guarantee that this channel is indeed separate, such observations create the potential for attacks. Additionally, during our collection period, NIST deprecated the use of SMS authentication (Grassi et al. 2016). However, we show that many services have not complied with this recommendation.
- **Characterization of Malicious Behavior via SMS Gateways:** We cluster and characterize the lifetime, volume, language, location, and content of the traffic seen in SMS gateways. Our analysis uncovers numerous malicious behaviors, including bulk targeted spam campaigns and phishing. Most critically, our data shows that these systems are being used to support phone-verified account fraud, and the ways in which these systems are used makes proposed mitigations from previous work (Thomas et al. 2014) largely ineffective.
- **Longitudinal Analysis:** Having the longest running SMS dataset to date, we then analyzed both malicious and benign behaviors over the course of two years. We find these behaviors to be largely stable and showing no significant changes over time, implying that these behaviors will likely continue for the foreseeable future.

We note the very fact that some users are willing to intentionally direct text messages to public portals is obviously dangerous. We do not address this phenomenon and instead focus on the risks of compromise of the SMS channel. Because these messages are known by the recipient to be publicly available, this dataset would naturally not be entirely representative of *all* SMS activity

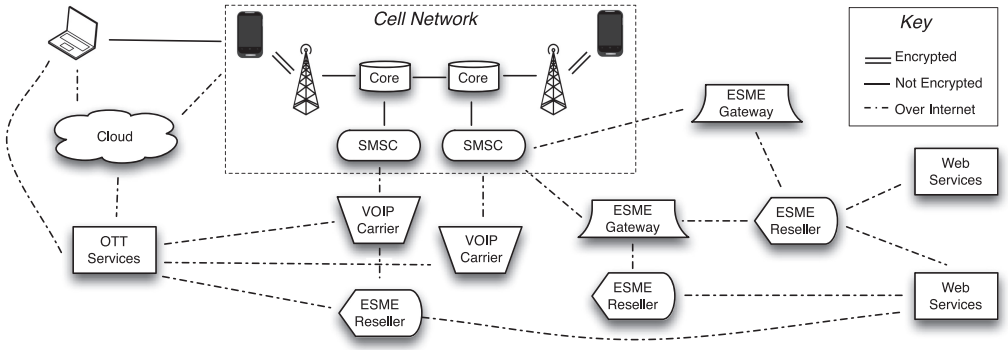


Fig. 1. The modern SMS ecosystem includes a wide variety of non-traditional carriers, ESME gateways and resellers, and OTT services. This evolution challenges old assumptions (e.g., phone numbers represent mobile devices tied to a single identity) and creates new opportunities for interception.

of a typical user. Nevertheless, this dataset enables the first public insights into issues such as PVA scams, SMS spam, and sensitive information sent by legitimate services. Furthermore, this data is widely available to the community for continued evaluation and measurement in the future.

The remainder of the article is organized as follows: Section 2 discusses the modern SMS ecosystem, which includes and extends beyond traditional cellular infrastructure; Section 3 discusses our collection and analysis methodology; Section 4 characterizes our dataset; Section 5 discusses our analysis on legitimate usage of SMS via the gateways; Section 6 discusses the malicious behaviors seen in our dataset. Section 7 analyzes related work, and Section 8 provides concluding remarks.

2 THE MODERN SMS ECOSYSTEM

In this section, we describe at a high level how text messages are sent and received, with an emphasis on developments that have greatly expanded the SMS ecosystem.

Figure 1 shows the components of the modern SMS ecosystem. Short Messaging Service Centers (SMSCs) route messages through carrier networks and are the heart of the SMS system (Traynor et al. 2008). These entities receive inbound text messages and handle delivery of these messages to mobile users in the network using a store-and-forward regime similar to email. When a mobile device sends or receives a text message, the message is encrypted between the phone and the base station serving the phone; however, once inside the core network the message is typically not encrypted.

Text messages[1] are not just sent between individuals but also by parties external to the network known as External Short Message Entities (ESMEs). ESMEs form an entire industry dedicated to facilitating the sending and receiving of messages for large-scale organizations for purposes as diverse as emergency alerts, donations to charities, or receiving one-time passwords (Traynor 2012). These ESMEs act as gatekeepers and interfaces to SMS. Some have direct connections to SMSCs in carrier networks via SMPP (Short Message Peer-to-Peer) (SMS Forum 2003), while others resell such access purchased from other ESMEs. For example, the VoIP carrier Bandwidth provides SMS access to many third-party services. Recently, startups like Twilio (2015), Nexmo (2015), and Plivo (2015) serve as ESMEs and provide easy-to-deploy, low-cost voice and SMS services.

Just as SMS distribution has evolved over the past two decades, how end users receive SMS has evolved as well. Originally, SMS were only delivered to mobile phones or to ESMEs. With the

[1]In this article, we use SMS and "text message" interchangeably.

advent of smartphones, this ecosystem is changing rapidly. Over-the-top networks like Burner (2015), Pinger (2015), and Google Voice (2015) provide SMS and voice services over data networks (including cellular data). Many of these services contract out to third-party ESMEs for service and do not actually act as ESMEs themselves. Additionally, messages that are delivered to a mobile device may not remain restricted to that device. Systems like Apple Continuity (2015), Google Voice, Pushbullet (2015), and MightyText (2015) use local wireless networks or cloud services to store and sync SMS from the receiving device to the user's other devices. Millions of subscribers use these services to transfer their messages from their localized mobile device to be stored in the cloud.

The modern SMS ecosystem has the security consequence that a single SMS may be processed by many different entities—not just carriers—who *in toto* present a broad attack surface. Essentially, the attack surface of the bulk collection of text messages has grown to include various types of ESMEs and many other end devices such as laptops and tablets. Attacks against these systems may be technical in nature and take a form similar to publicized data breaches (Krebs 2014, 2015a, 2015b; U.S. Office of Personnel Management 2015). Additionally, redirection attacks on the SS7 network can reroute calls and SMS messages to an endpoint controlled by an adversary without knowledge to the user (Karsten Nohl 2016; Luca Melette 2016). More specifically, such attacks can be carried out by any independent party that is willing to purchase access to the SS7 network. Once inside, the attacker can spoof call forwarding requests to redirect the calls/messages to themselves or eavesdrop into a conversation (Peeters et al. 2018). These attacks on the SS7 core are becoming more prevalent and easier to exploit than ever. Social engineering attacks are also possible. Mobile Transaction Authentication Numbers (mTANs)[2] have been stolen using SIM Swap attacks (Tims 2015), where an attacker impersonates the victim to a carrier to receive a SIM card for the victim's account, allowing the attacker to intercept security-sensitive messages. Attackers have also compromised accounts protected by one-time-passwords delivered over SMS by impersonating the victim to set up number forwarding to an attacker-controlled device (Campbell-Dollaghan 2014). Accordingly, it is worth determining what data are sent via SMS so that the consequences of future compromise are well understood.

This work measures how different entities implement security mechanisms via text messages through the use of public SMS gateways. As such, we are able to observe a wide array of services and their behavior through time. Additionally, because these gateways provide phone numbers to anonymous users, we are also able to measure the extent to which such gateways are being used for malicious purposes. This combined measurement will help to provide the research community with a more accurate and informed picture of the security of this space.

3 METHODOLOGY

In this section, we describe the origins of our dataset, discuss some limitations of the data, discuss supplementary sources that give us additional insights into our SMS dataset, and finally describe the techniques we use to analyze the dataset.

3.1 Public Gateways

In the previous section, we noted that there are a number of organizations that process text messages, including carriers, ESMEs, resellers, and value-added services like message syncing. Within the category of ESMEs lie a niche class of operator: public SMS gateways. Many third-party entities (including cellular carriers) provide external public interfaces to send text messages but not

[2]mTANs are used to authenticate mobile banking transactions via SMS in many countries, including Germany, South Africa, and Russia.

Table 1. SMS Gateways We Analyzed and the Number of Messages Collected from Each

Site	Messages	Phone #s
(1) receivesmsonline.net	120,082	55
(2) receive-sms-online.info	180,865	129
(3) receive-sms-now.com	83,086	61
(4) hs3x.com	132,042	110
(5) receivesmsonline.com	149,767	138
(6) receivefreesms.com	83,691	134
(7) receive-sms-online.com	138,755	32
(8) e-receivesms.com	12,367	19

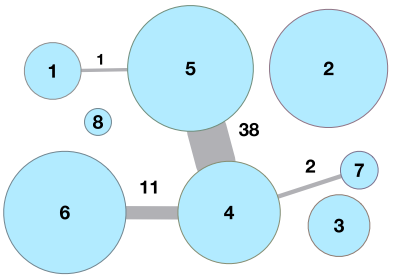


Fig. 2. This figure indicates the relative amount of phone numbers owned by gateways 1–8 (vertices), and edges indicate the count of shared phone numbers. Many gateways concurrently share phone numbers, indicating collaboration by operators.

receive them. Example use cases include the convenience of an email gateway or the ability to use a web service to send a message to a friend after one's mobile phone battery dies.

While there are many public services for sending messages, they also have counterparts in public websites that allow anyone to *receive* a text message online. These systems publish telephone numbers that can receive text messages, and when a text message arrives at that number the web site publicly publishes the text message. These services are completely open—they require no registration or login, and it is clear to all users that any message sent to the gateway is publicly available. Messages that show up in these gateways can be sent through different network technologies (e.g, GSM, UMTS, LTE, or from other ESMEs), as such, we expect to see no difference in message content from any of those technologies. We recognized the research value of these messages for the potential to inform a data-driven analysis, and collected them over a 28 month period from 8 distinct public gateways that facilitate the reception of text messages,[3] listed in Table 1. These gateways have similar names that are potentially confusing, so where appropriate, we reference them by an assigned number 1–8 based on message volume. In Figure 2, we show the overlap in phone numbers provided by gateways. The fact that these gateways share many phone numbers concurrently indicates they are operated by collaborating parties.

To gain more information on the gateways, we did a WHOIS lookup on each of the services. We saw that all services were registered sometime between 2011 and 2014. From the eight gateways, only three had public information available; the rest had their WHOIS information hidden. The three remaining gateways were registered in China, Pakistan, and Brazil. We note that these gateways only provided numbers in countries *outside* the country indicated in their registration.

These different services have essentially the same functionality, but advertise their intended use in different ways. These include avoiding spam, creating phone-verified accounts, and enhancing privacy by not giving out a user's real number. We suspect that the business model of most of these websites relies on advertising revenue, and this is confirmed by at least Gateway 2, which prominently displays "almost all of [our income] comes from our online advertising" in a banner requesting that users disable their ad blocker. However, advertising is not the sole source of revenue for every system: Gateways 3, 4, 5, 6, and 8 sell private numbers for receiving SMS, while Gateways 4 and 5 actually sell verified Google Voice and WhatsApp accounts.

[3]Note that throughout the rest of the article, we use the term "gateway" to refer exclusively to these receive-only SMS gateways.

Ethical Considerations. As researchers, our ultimate goal is to improve the security practices of users and organizations, but we must do so ethically. In particular, we should make every effort to respect the users whose data we use in our studies.

A superficial ethical analysis would conclude that because it is clear that all messages sent to these gateways are public, and their use is strictly "opt-in," users have no reasonable expectation of privacy in the collection and analysis of this data. While we believe this analysis to be true, the situation is more complex and requires further discussion, as there are a number of parties to these messages. In addition to users who knowingly provide a gateway number as their own phone number, other individuals and institutions (companies, charities, etc.) may send information to individuals, not knowing that the messages are delivered to a public gateway. While institutions rightfully have privacy rights and concerns, they differ from those of individuals. As we show in our results, the vast majority of the information that we collect is sent indiscriminately and automatically by organizations to a large number of recipients. This information is unlikely to contain information that would negatively impact the institution if disclosed. Although we study bulk institutional messages, we do not analyze further those messages determined to be of a strictly personal nature. While those messages may have a research value, we *deliberately avoid these messages to prevent further propagating this data.*

Nevertheless, the use of gateways absolutely creates confidentiality and privacy concerns. For example, when personally identifying information (PII) or account credentials are sent to a gateway (whether or not all parties are aware), the compromise of that information is immediate and irrevocable.[4] Because we do not make our data available to others, this study does not change—in severity or duration—the harm done by the existence and use of the gateway. Furthermore, while in Section 5.1 we describe a host of sensitive information found in the dataset, we do not publish, use, or otherwise take advantage of this information. In particular, we especially do not attempt to access accounts owned by gateway users or operators.

We recognize that there are ethical questions raised not just with the collection of this data but also by combining it with other data sources. Our data augmentation is sufficiently coarse-grained that no individual user of a gateway could be identified through our additional data.[5] Geographic information not already disclosed in text messages was limited to country-scale records in the case of gateway users and city-scale in the case of gateway numbers (which in any case do not likely correlate with the location of the gateway operator).

Overall, our hope is this study would raise awareness of the risks of sending sensitive information over insecure media and prevent future harm.

Limitations. To the best of our knowledge, this article presents an analysis of the largest dataset of SMS published to date. However, there are some limitations to this data. First, because the messages are public, many services that use SMS (like mobile banking) are likely underrepresented in our dataset. For this reason, it is likely that our findings about sensitive data appearing in SMS are likely *underestimated.* Second, because gateways change their phone numbers with regularity, it is unlikely that long-term accounts can be successfully created and maintained using these numbers, which may bias the number of services we observe in our dataset. Accordingly, those users are unlikely to enable additional security services like mobile two-factor authentication (2FA) using one-time passwords (OTP), further limiting our visibility to a wider range of services. *These*

[4]Except perhaps by the gateway itself; however, it is clear from our data that gateways are not taking steps to prevent PII exposure.

[5]The one exception to this was an individual whose information was used (likely without his/her knowledge) to register a domain used in a phishing scam. This information was discovered after a routine WHOIS lookup after discovering the phishing domain.

limitations mean that the overall distributions that we report may not generalize to broader populations. Nevertheless, we believe that this work provides useful conclusions for the security community.

3.2 Crawling Public Gateways

We gather messages by crawling gateway pages using the Scrapy (2015) framework. Every 15min, our crawler visited each gateway, obtained new messages, and stored these in a database. We faced two challenges to accurately recording messages: ignoring previously crawled messages and recovering message received times.

Ignoring previously crawled messages was difficult, because gateways display the same messages for a considerable amount of time (days, months, or even years). A consequence of this is that our dataset contains messages that gateways received before our data collection began. To prevent storing the same messages repeatedly (and thus skewing the results), we discard previously crawled messages upon arrival by comparing the hash of the sender and receiver MSISDNs and the message content against hashes already in the database. If a match is found, then the message sent times are compared to ensure that they were the same instance of that message, ensuring that messages that were repeatedly sent are still included in the data.

Message times required finesse to manage, because gateways report a relative time since the message was received (e.g., "3 hours ago") instead of an ideal ISO-8601 timestamp. Parsing these timestamps is fairly simple, but care must be taken when doing comparisons using these times as the precision can vary ("3 minutes" vs. "3 days"). To ensure accuracy, we store and take into account the precision of every timestamp when comparing message timestamps.

3.3 Additional Data Sources and Analyses

3.3.1 Phone Number Analysis. After the scrapers pull the initial data from the gateways, the data is augmented with data from two outside sources. The first service, Twilio (2015), provides a RESTful service that provides mobile, VoIP, and landline number look ups. Twilio resolves the number's country of origin, national number format for that country, and the number's carrier. Carrier information includes the carrier's name, the number's type, and the mobile network and country codes. Twilio is accurate and appropriately handles issues like number porting, which could cause inconsistencies in our data if incorrect.

The second service, OpenCNAM (2015), provides caller identity information for North American numbers. This database contains a mapping of phone numbers and strings; carriers consult this database to provide Caller ID information when connecting a call. Therefore, OpenCNAM is also the most accurate public location to obtain identity information for North American numbers.

We obtained data from both Twilio and OpenCNAM for all gateway numbers as well as a subset of the numbers that contacted the hosted numbers.

3.3.2 URL Analysis. We extracted 51,849 URLs from messages by matching URL regular expressions with each message in the dataset. Overall, there were 1,754 unique second-level domains and 2,390 unique base URLs (fully-qualified domain names and IP addresses) in this set. For each of these domains, we obtained domain registration data. A domain's WHOIS registration data contains useful metadata about the history of a domain, including its creation date. Since this data is distributed among registrars, it is not always available and some fields may be restricted.

Due to the limited length of an SMS message, shortened URLs are often sent in these messages. The short URL is a hop between the user and the destination, allowing URL shortening services to collect data about the users following the links. For each Bitly- and Google-shortened URL, we obtained statistics (e.g., number of clicks) when possible. The SMS gateway services do not

publish data on their users, so this data represents one of the best insights into user demographics in our dataset.

Finally, since these gateways freely accept and publicly post SMS messages, the gateways represent an easy mechanism for delivering malicious messages including phishing or malicious URLs. VirusTotal (2015) can provide valuable insight into the maliciousness of a given URL. We requested scans of each of the URLs via VirusTotal and collected the scan reports.[6] If a URL had a previously-requested scan, then we collected the cached scan and did not rescan the URL. Due to the short lifetimes of some malicious domains, we anticipated earlier scan results would be more accurate. For each product that VirusTotal uses to scan the URL, it reports whether or not the product alerted and if so, the category of detection.

3.3.3 Personally-Identifying Information Analysis. We searched the messages for personally-identifying information (PII) (McCallister et al. 2010) using regular expressions. In particular, we searched for major credit card account numbers (e.g., Visa, Mastercard, American Express, Discover, JCB, and Diners Club). For each match, we further verified these numbers using the Luhn algorithm (Luhn 1960). This algorithm performs a checksum and can detect small input errors in an account number. This checksum is built into all major credit card account numbers and can also assist in distinguishing a 16-digit Visa account number from a 16-digit purchase order number. This check is rudimentary, however, and we manually verified the remaining matches to verify that they contextually appeared to be account numbers (i.e., the messages containing these numbers appeared to reference an account).

Furthermore, we also checked strings of numbers to determine if they were identification numbers such as U.S. Social Security Numbers or national identifiers from Austria, Bulgaria, Canada, China, Croatia, Denmark, Finland, India, Italy, Norway, Romania, South Korea, Sweden, Taiwan, or the United Kingdom. We found no valid matches in our data.

3.4 Message Clustering

A major goal of this study is to determine what types of messages are sent via SMS and how service providers are using this form of communication. To do so, we grouped messages together into unique clusters[7] that are representative of their content. The essence of our clustering algorithm is distance-based clustering (Frey and Dueck 2007). We note that the messages of our interest were virtually identical, apart from known common variable strings like `codes` or email addresses. By replacing these with fixed values, a simple lexical sort would group common messages together. We then identified cluster boundaries by finding where the normalized edit distance was lower than a threshold (0.9) between two consecutive sorted messages. Our threshold was empirically selected to conservatively yield correct clusters with high accuracy.

A more explicit statement of this process follows:

(1) *Load all messages.*
(2) *Preprocess messages by replacing numbers, emails and URLs with fixed strings.*
(3) *Alphabetically sort preprocessed messages.*
(4) *Separate messages into clusters by using an edit distance threshold to find dissimilar consecutive messages.*
(5) *Manually inspect each cluster to label service providers, message types, and so on. In this step, we culled clusters that had* <43 *messages.*[8]

[6] We were not able to obtain reports for 365 URLs due to poor formatting of the SMS messages.
[7] Our use of this term should not be confused with the classic machine learning definition of "clustering."
[8] We initially planned on labeling only clusters with more than 50 messages, but our labelling process resulted in more labeled clusters than expected.

After clustering, we *manually* labeled each cluster, a time-consuming process that allowed us to both verify the correctness of the cluster generation, and guarantees correct labels. We note that this clustering was done on the same dataset (386,327 messages) found in our original work[9] (Reaves et al. 2016a). Instead of reclustering the new messages with this algorithm, we simply made a regular expression for each cluster that was used in new analyses.

3.5 Message Intentions

Due to the lack of standardized terms for the intentions of the authentication and verification values sent via SMS, we divided the various message *intentions* into categories in this section. In this article, we use code to describe the value extracted from any message sent to a user for any of the below intentions. To our knowledge, there is no authoritative source for these intentions, despite their popularity. More than half of the messages contain a code, and the following categories enabled us to more accurately cluster our messages:

- **Account Creation Verification:** The message provides a code to a user from a service provider that requires a SMS verification during a new account creation.
- **Activity Confirmation:** The message provides a code to a user from a service provider asking for authorization for an activity (e.g., payment confirmation).
- **One-Time Password:** The message contains a code for a user login.
- **One-Time Password for Binding Different Devices:** The message is sent to a user to bind an existing account with a new phone number or to enable the corresponding mobile application.
- **Password Reset:** The message contains a code for account password reset.
- **Generic:** We use this category for any codes to which we are unable to assign a more specific intent.

4 DATA CHARACTERIZATION

In this section, we provide high-level information about our collected data. The dataset includes data from 8 gateways over 28 months. Overall, our dataset includes 900,655 messages sent from 625 phone numbers from 65 known carriers in 30 countries. Table 2 shows the message count for gateway phone numbers alongside the total number of gateway numbers by country.

Gateways and Messages. Table 1 shows the eight gateways we scraped, the number of messages from each, and the number of unique phone numbers hosted at each service during the collection time. The number of messages received by each gateway ranged from 12,367 to 180,865. The hosted numbers per service ranged from 19 to 138.

Infrastructure. We obtained detailed data from Twilio about the phone numbers in our dataset, as shown in Table 3. Twilio identified 65 carriers, of which 55 are mobile, 7 are VoIP, and three are labeled as landline carriers. We believe that the numbers seen from these "landline" carriers are mislabeled as landlines by Twilio and are actually mobile numbers, due to all three being carriers that advertise both mobile and landline service. Furthermore, Twilio indicates numbers from Bandwidth as "mobile" numbers (this is not due to porting, as Twilio resolves porting scenarios). Bandwidth is actually a VoIP provider. The numbers in this article are corrected to reflect this.

Geography. Twilio's number data also includes geolocation information for each number, which shows our data is based in 30 countries. The United States has the most gateway controlled numbers with 147 numbers seen receiving 95,138 messages, the most traffic of any country. Conversely,

[9]We also include a more detailed approach to our clustering mechanism in this work.

Table 2. This Table of Gateway Messages and Numbers by Country Shows that Gateways Have an International Presence, with most Message Volume Taking Place in North America and Western Europe

Country	Message Count	Number Count	Country	Message Count	Number Count
United States	246,736	147	Belgium	5,253	3
Canada	139,045	72	India	5,064	2
United Kingdom	93,999	109	Hong Kong	4,597	11
Germany	93,561	75	Israel	4,325	9
Poland	68,777	18	Thailand	4,073	5
Russia	52,014	8	Switzerland	2,610	5
Norway	49,362	13	Austria	1,774	4
Spain	23,227	17	Finland	1,714	13
Sweden	21,485	24	Indonesia	1,201	3
Ukraine	19,929	3	Netherlands	982	1
France	19,218	23	Estonia	976	3
Romania	12,014	17	Ireland	526	4
Italy	10,617	5	Lithuania	520	1
Australia	9,763	20	Denmark	54	1
Mexico	6,880	6	Czech Republic	31	3

The message count represents the number of messages sent to numbers in each country.

Table 3. Using Twilio-provided Data, We Obtained the Carrier Type for Each of the Carriers Associated with Sender and Receiver Numbers on the Gateways

Carrier Type	Total	Percentage of Total
Mobile	397	63.5%
VoIP	207	33.1%
Landline	21	3.4%

Lithuania only had one gateway-controlled number registered to it, the lowest of the countries in our data. The Czech Republic has the fewest messages sent to the gateway-controlled numbers registered to a country, with three numbers receiving only 31 messages.

Twilio data provides only the country of origin, so for all 219 numbers in the United States and Canada we obtain caller ID name (CNAM) data.[10] We found that the vast majority of numbers (57.2%) have no CNAM data at all. Of those messages that have data, the official record in the CNAM database is simply "CONFIDENTIAL," "WIRELESS CALLER," or "Unavailable." Note that "Unavailable" is the actual string that would be displayed to a user, not an indication of no data in the database.

The remainder of the messages are sent to phone numbers that have CNAM data indicating the number is in one of 57 cities or 3 provinces (British Columbia, Ontario, and Quebec) in the United States or Canada. By message volume, the top locations are "Ontario," followed by Boston, MA; Atlanta, GA; Southfield, MI; Harrisburg, PA; Inverness, MS; Denver, CO; Oakland, CA. There are several observations to make from these findings: first, numbers are selected to well beyond what is likely the gateways' main location. Second is that neither gateways nor users feel a need to use

[10]CNAM data only covers the U.S. and Canada.

Table 4. We Separated and Labeled Each Cluster Containing a code the Intent of the Message

Tag	Messages	% Tagged	Tag	Messages	% Tagged
otp-dev	95,685	33.4%	info	2,339	0.8%
code	52,872	18.5%	otp-dev-url	863	0.3%
ver	52,181	18.2%	password	697	0.2%
conf	38,521	13.4%	code-url	676	0.2%
otp	21,919	7.6%	conf-ro	401	0.1%
pw-reset	3,602	1.3%	otp-url	320	0.1%
ver-url	3,139	1.1%	stop	284	0.1%
advertising	2,999	1.0%	username	178	0.06%
pw-reset-url	2,696	0.9%	conf-url	92	0.03%
test	2,612	0.9%			

This table contains each of those labels and the number of messages in each, which total 74.2% of the messages in our dataset.

numbers based in large population centers. With the exception of Centennial, CO, all locations had five or fewer numbers, regardless of population of the location. Gateways 4 and 5 registered the numbers in Centennial.

Clusters. From the clustering algorithm we described in Section 3.4, we found 44,579 clusters in our initial dataset. All messages with more than 43 messages were manually tagged and analyzed giving us 754 tagged clusters. These clusters represent the messages from the most popular services that were found in this dataset. The tagged clusters only represent 1.7% of the total clusters but the tagged clusters cover 286,963 messages (74.2%).

SMS Usage. As shown in Table 4, messages containing a code constitute the majority of our dataset at 67.6% of the total messages, showing that a main usage of SMS in our data is verification and authentication.[11] Account creation and mobile device binding codes are the largest subcategories with 51.6% of the messages. Compared to other messages containing a code, one-time password messages are only 7.6% of messages. The URL variations for these code messages are also rare, constituting only 2.6% of messages. This reflects that most services prefer to plain codes, instead of URLs, which may not work well for older phones.

Password reset messages comprise 1.3% of our dataset. The corresponding URL version comprises another 1.0% of our dataset. Interestingly, these password reset URLs overwhelmingly consist of Facebook reset requests.

Language. We also programmatically determined the language of the messages. We used Google's `langdetect` library to systematically tag each message with a predicted language if the output had a confidence of 60% or greater. For this analysis, we removed any messages for which the language could not be determined. These messages had low confidence value, were too small to predict correctly, or did not contain any characters at all (e.g., ":)," "12345"). After dropping such messages, we found a total of 51 different languages with the top 6 languages (English, Russian, Portuguese, German, French, and Spanish) making up 80% of all messages.

[11] As we note in the previous section, these percentages are reflective of gateway messages, and may not necessarily be representative of broader SMS trends. Additionally, tagging messages is a manual process that takes large amounts of time. As such, the numbers reflected on Table 4 are representative of our previous work (Reaves et al. 2016a) that contained 380k messages.

Not surprisingly, English messages made up 57% of the messages. Detecting the language gives us a reasonable idea of geographical regions where public gateways are mainly used. We briefly discuss how we use this information to track malicious activities later in the article.

We looked deeper into the 27,000 Spanish messages (the sixth top language), because several authors are native or competent speakers. Here we found many VoIP service offering cheap calls to different Latin American countries, and we noticed that most of these services were used to call Cuba.

Finally, a few messages contain partial or complete usernames and passwords. These messages are particularly egregious, because they may lead to account compromise and/or user identification. We discuss this further below.

5 USES OF SMS AS A SECURE CHANNEL

In this section, we discuss what we observed about the security implications if any of the components of the SMS ecosystem are compromised. We found messages capable of exposing financial information, login credentials, and critical alert messages. Additionally, although the usage we discuss in this section is benign, we observe the presence of low entropy in 2FA messages and PII leakage, both of which are dangerous when available to an adversary in this ecosystem.

5.1 PII and Other Sensitive Information

SMS has become a major portion of global telecommunications worldwide, and its use by companies and other organizations comes as unsurprising. However, our dataset contained instances of companies using SMS to distribute payment credentials or other financial information, login credentials, and other personally identifiable information. We also see instances where gateways are used for sensitive services.

Financial Information. We found several distinct instances of credit card numbers being distributed over SMS in our dataset. Two of these appear to be intentional methods of distributing new cards, while another two appear to be the result of commerce. We discovered these using PII regular expressions. We also discovered several instances of CVV2 codes in our data. CVV2 codes are credit card codes meant to verify that the user is in possession of the physical card at the time of purchase.

We found that two services that provide "virtual" credit card numbers to allow access to mobile wallet funds distribute the numbers over SMS. These card numbers are "virtual" in the sense that they are not backed by a credit line, but in fact seem to be persistent. The first service is Paytoo, based in the United States. We recovered three distinct cards from this service, and additional messages containing balance updates, account numbers and transaction identifiers. While password reset was handled over email, identifiers such as email, username, phone number, or account number could all be used for login.

The other service is iCashCard, based in India. They distribute a prepaid credit card account number over SMS; this card is protected by a PIN also distributed over SMS. Additional messages contained a separate PIN that allows for account login with the phone number, meaning that access to SMS reveals access to the entire payment credential and account.

We found an additional credit card number, CVV, and expiration value from an unnamed service whose identity or purpose we could not identify. The message indicated that it was being sent to a user who had purchased a "package" of some sort, and confirmed the purchase using the full credit card number. Incidentally, the purchaser's IP address was listed in the SMS, and that IP address was placed in SANS blocklist for suspected bots and forum spammers.

Our PII regular expressions discovered one final credit card number present in a text message sent to a Mexican phone number. The message contains a reference to a Venezuelan bank, the card

holder's name, and includes the credit card number, the CVV2, and the expiration date. To determine the context for this message, we examined other messages from this sender and found what appeared to be an SMS-based mailing list for purchasing items on the black market in Venezuela; items for sale included U.S. paper products (diapers, tissue), oil, and tires, as well as U.S. dollars at non-official rates (Crooks 2015). Our best hypothesis for the presence of the credit card is that a purchaser of an item mistakenly sent payment information to the list in place of the actual sender. Nevertheless, this highlights that highly sensitive enterprises rely on SMS.

In addition to credit card information, we discovered one unidentified Polish service that includes a CVV2 code in their messages after registering for a prepaid service. Translated (by Google), these messages read:

```
Thank you for registering on the
site prepaid. Your CVV2 code is: 194
```

The financial information in our gateway data is not limited to credit card numbers. We found several instances of messages sent by a prepaid credit card provider in Germany, PayCenter (2015), that distributes bank account numbers (IBANs) in SMS messages. The same provider also sends a verification text to the user with a URL that includes the user's full name.

The messages above indicate that some services unwisely transmit sensitive financial information over SMS.

Usernames and Passwords. In scanning our labeled clusters, we identified several services that would allow user accounts to be compromised if SMS confidentiality is lost. The most prominent example of these is Canadian international calling provider Boss Revolution (2015). Their user passwords are distributed via SMS, and usernames are simply the user's phone number. Thus, an attacker with read access to these messages can compromise an account. Another example was the Frim messaging service (Frim 2015). That service also uses the user's phone number and a password distributed over SMS. Other services distributing usernames and passwords in SMS include eCall.ch (a Swiss VoIP provider) (eCall 2015) and RedOxygen (a bulk SMS provider) (RedOxygen 2015). Fortunately for users, most online services in our data do not distribute password information through SMS.

Password Reset. Several organizations, including Facebook and the investment platform xCFD, distribute password reset information via SMS in addition to or in place of other methods. The most common password request in our data was for Facebook account resets. Upon investigating these messages (using only our own accounts), we found that the messages contained a URL that would allow a password reset with no other identifying information or authentication—not even a name or username. This would allow any adversary with access to the message—either as it transits carrier networks, the receiving device, or any other entity that handles the message—to control the account. If the adversary has the username, then he/she could cause reset messages to be sent for that account, allowing the adversary to take complete control of the account. This highlights the consequences of a compromise of the SMS ecosystem.

Alert Systems and Status Reports. We also found some gateway-provided phone numbers were being used to receive alert messages. These messages notified the phone number of important content that may be time sensitive such as security alarms. To identify the alert messages, we queried for an extensive dictionary of words that have a high likelihood of being present in alert messages or reports (e.g., "warning," "accident," "urgent," etc.). The messages found in this set had various levels of user interactivity ranging from a simple notification update to the request of action from the recipient. For example, we found security alert messages that implied that the lack

of a response from the recipient caused a secondary course of action to be taken by the security company. From the security system's point of view, any confidential information sent to the provided number is assumed to go to the rightful recipient. However, since the gateway numbers are public, the delivery of these messages publicizes the confidential information unbeknownst to the security company.

In addition to the security messages, we also found status messages containing critical measurements from a Biogas plant, which metadata indicated was located in Great Britain, though the source number was partially masked by the gateway. Although the messages we found did not give us enough information to accurately trace the individual plant, the publication of this type of messages have been shown to leak vital information in other power plants (Hilt and Lin 2016). The problem of these leakages can be traced back to the assumption of the phone network being a secure and reliable medium through which the plant could send confidential information. It is important to note that it is not just confidentiality that is a concern; SMS also does not guarantee availability, timely delivery, authentication of either party, or integrity of the data. In fact, previous work (Traynor 2012) has shown that mass alert messages sent through SMS take hours, and in some cases fail, to deliver the critical content. Having that in mind, mission critical messages simply should not be sent through SMS.

Other Personally Identifiable Information. We found numerous examples of PII—including addresses, zip codes, and email addresses. Email addresses are worth noting, because the presence of an email address indicating an association between a phone number an account could be used to associate `codes` or other authenticators sent to that device to the particular account. Our PII regular expressions identified 2,849 messages with emails—most of these were sent by `live.com`, `gmail.com`, `inbox.ru`, or `pop.co` (a hosting provider).

SMS Activity from Sensitive Applications. Finally, we noticed several instances where messages appeared in the gateway from organizations whose very nature is sensitive. The worst among these was the room sharing service Airbnb. One of our messages contained the full address of the shared property (PII obscured):

```
Airbnb reservation reminder:
Jan 25-28 @ <address>.
<name>: <email> or <phone>
```

Although we suspect that the owner of the property listed it in such a way that this data was revealed, the use of SMS gateways for these services is troubling as it could facilitate real-world abuses.

Other examples of sensitive applications include a large set of registrations with other telecommunications services. These include popular phone services like Telegram, Viber, Line, Burner and Frim. The presence of these services in gateway data may indicate the use of these gateways for "number chaining," a practice that allows PVA evaders to acquire a large number of telephone numbers for free (Thomas et al. 2013). In addition, we see registration and activity in the gateway data to a number of bulk SMS services. This may also indicate the use of gateway numbers to obtain access to bulk SMS services for the purposes of sending spam.

5.2 SMS code Entropy

Our message dataset afforded us samples of `codes` sent by many services over SMS. These `codes` provide valuable phone verification capabilities to services that wish to increase the burden of obtaining an account (e.g., to prevent fraudulent account creation), and these `codes` provide a glimpse into the security of the code-generation schemes. We grouped those clusters containing

Table 5. The Results of Our Statistical Analysis of Authentication Codes from Each Service

Service	Effect Size (w)	Effect?	Mean Code
Circle	0.519	large	496,206
Google	0.762	large	559,252
Google	0.779	large	544,276
Google	1.249	large	462,109
Hushmail	0.476	medium	503,494
Instagram	0.802	large	495,483
Instagram	0.787	large	500,052
Instagram	0.655	large	498,085
Jamba	9.053	large	6,801
LINE	0.653	large	5,468
LINE	0.543	large	5,434
Microsoft	3.338	large	334,253
Odnoklassniki	0.585	large	434,079
QQ	0.530	large	498,245
Talk2	1.611	large	5,717
Telegram	0.490	medium	54,906
Viber	3.001	large	388,336
Wechat	0.506	large	4,958
Whatsapp	0.610	large	541,209

(a) `Non-uniformly Random codes` ($p < 0.001$).

Service	p-value	Mean Code
Alibaba	0.931	548,088
Backslash	0.324	556,223
Baidu	0.006	502,144
Beetalk	0.731	541,852
Gett	0.391	5,503
LINE	0.459	5,525
LINE	0.611	5,567
Origin	0.969	501,422
Origin	0.377	502,599
Runabove	0.332	494,698
Skout	0.342	5,049
Smsglobal	0.802	5,502
Tuenti	0.939	5,062
Weibo	0.008	504,205

(b) `Uniformly Random codes`.

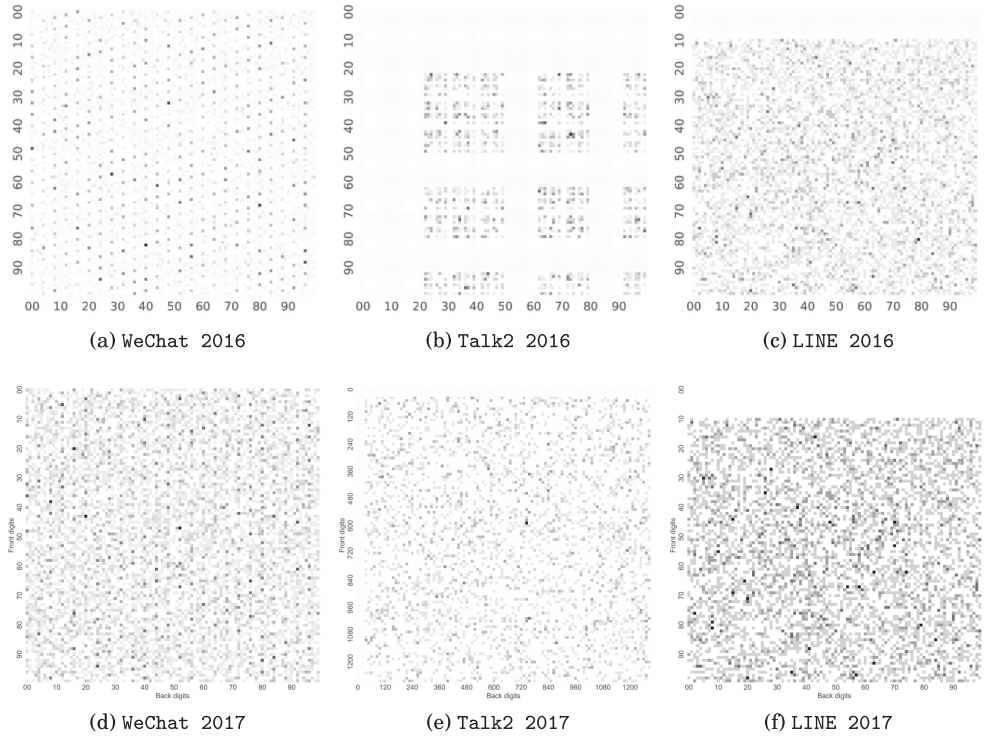
Some services appear more than once in the data because their messages were split into multiple clusters (e.g., one for password resets and one for logins).

codes by service and extracted the numeric code from each message. Overall, we extracted codes from 33 clusters containing 232,999 authentication codes across 25 services, as shown in Table 5.

We first tested the entropy of each set of codes using a chi-square test. The chi-square test is a null hypothesis significance test, and in our use case indicates if the codes are uniformly generated between the lowest and highest value. The p-value less than 0.001 means that there is a statistically significant difference between the observed data and an ideal uniform distribution. Only 13 of 33 clusters (39%) had p-values > 0.001. We also measure the effect size for each test, finding that most effect sizes were large ($w > 0.5$) with only one medium ($w > 0.3$), indicating our statistically significant differences were in fact meaningful. Finally, we confirmed that all tests performed had a statistical power of 0.99 or higher, indicating that our test had a high likelihood of observing any effect present.

Of the clusters, those belonging to the WeChat and Talk2 services had the least entropy of the authentication codes we analyzed. Not only did both services have p-values of 0.0 in the above chi-square test, the service's codes each generate a specific pattern. We mapped the first two digits of each code with the back two digits and show these two services' codes in Figure 3.

WeChat. Until April 2015, WeChat's authentication codes followed a pattern of $rand() * 16 \bmod 10{,}000$, which caused the grid-shaped heatmap in Figure 3(d). The pattern could be explained by a random number generator with low entropy in the four least significant bits. This effectively reduced the possible space of 4-digit codes to 625. In April 2015, WeChat changed its code generation algorithm. We excluded the the codes gathered before April and recomputed the chi-square test. Although the new map shows like they use a better random generator, we still see the service fail the chi-square test.



(a) WeChat 2016 (b) Talk2 2016 (c) LINE 2016

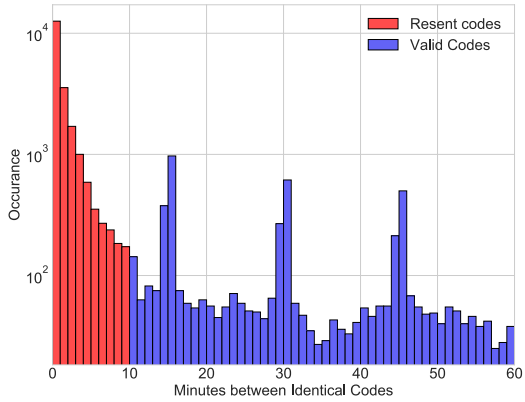
(d) WeChat 2017 (e) Talk2 2017 (f) LINE 2017

Fig. 3. These figures present heatmaps of codes where the first two digits are represented on the y-axis and the last two digits are represented on the x-axis. Darker values represent higher frequencies of a code in our data. In the top row, which visualizes data from our first study, we see that WeChat and Talk2 present an egregious lack of entropy in their authentication codes, while LINE generates random codes without leading zeros. In the bottom row, which visualizes our newer data, we can see that WeChat changed the code generation to include more possible values and Talk2 changed to generating alpha numeric codes. Our statistical analysis showed that these changes did not result in uniformly randomly generated codes.

Talk2. This service has an extreme lack of entropy in its code-generation algorithm, as seen in Figure 3(e). In particular, it appears to avoid digits 0, 1, 2, 5, and 8 in positions 1 and 3 of a 4-digit code. We made several attempts to reproduce this entropy pattern, but we were unable to produce a reasonable explanation for the reduction in entropy.

After March 2016, Talk2 changed their code generation from a four digit code to a four character alphanumeric string (e.g.,"4KXT" instead of "1234"). We were able to get a sufficient sample size to map these new codes to base 36 (10 digits plus 26 letters) and rerun the chi-square analysis test separately from the previous codes. As before, this new method implemented by Talk2 did not pass the random uniformity test. We generated a heatmap for these codes, but were not able to explain the non-uniformity.

Google. While the Google codes we harvested did not appear to be uniformly-random in our experiments, this appears to be caused by duplicate codes. When requesting that a code be resent, Google will send the same code again. This practice is potentially problematic, because it indicates that the Google codes have a long lifetime. Since messages on gateways may be accessible for weeks or months, it may be possible for an adversary that can identify the associated account to use an unclaimed code. Without access to the associated accounts, however, we were unable to determine the exact lifetime of Google's codes.



(a) Time Difference Between Identical Codes

Fig. 4. Users often request a code multiple times for the same use (e.g., login), so many services cache this code and resend instead of generating a fresh code. This figure shows that overall the likelihood of sending an identical code falls sharply after 10min, so we exclude identical codes sent within 10min from our randomness measurement.

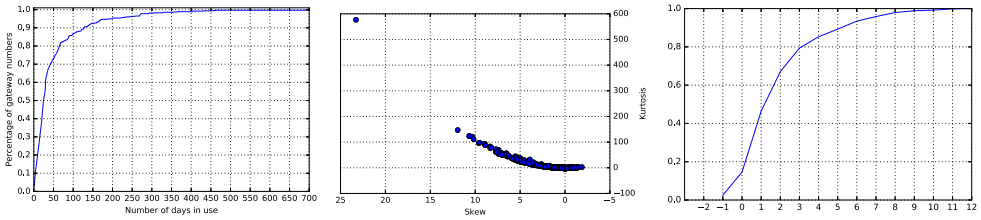
LINE. Although our experiments show LINE's codes are likely uniformly generated, the service does not generate codes with a leading zero, reducing the overall space of codes by 10%. This practice is common among our clusters, with 13 total clusters exhibiting this behavior. For comparison, we display LINE's codes in Figure 3(f).

During our collection period, we frequently found identical codes appearing during the same time frame. This was due to services sending duplicate messages rather than generating new codes from scratch. In our previous study (Reaves et al. 2016a), these repeated codes had the potential to bias our statistics in favor of considering a code distribution to be non-uniform. In this study, we excluded duplicated codes that have been collected within 10min of each other because of the high likelihood these codes were simply duplicate messages. Figure 4 shows that the bulk of duplicate codes were sent within 10min of each other (hence our choice of time limit for that value). While this revised analysis computed slightly different results for our data, we still found that many online services still use a poor random number generator.

In addition to the entropy analysis, we also checked if the services had complied with NIST's deprecation of SMS-based authentication. To do so, we checked the date of the last message received from each cluster. Our dataset stopped finding messages for 15 clusters during our collection period (9 from before the recommendations were made public and 6 after the release). We manually checked through different messages and security policies and found that all but three services still support SMS two-factor authentication. Of the three services, we were not able to find information that indicated whether or not the services stopped supporting SMS authentication. Accordingly, we see that the NIST recommendations are *not* being followed and our data provides no reason to believe this will happen in the near future.

5.3 Takeaways

In this section, we explored the data that is exposed in the SMS channel for benign purposes. This is problematic if an adversary has access to SMS messages, as is the case with the gateways. We observed services that expose sensitive user data via SMS including financial data, account information, password reset URLs, and personal information such as physical and e-mail addresses. We then found that 65% of services that use SMS to deliver codes generate low-entropy codes,



(a) CDF: Number Lifetime (b) Daily Activity Skewness and Kurtosis (c) CDF: Daily Activity Skewness

Fig. 5. (a) Only 37% of gateway-controlled numbers are used after one month. The median number lifetime is only 24 days. (b) The skewness and kurtosis of number lifetime indicates that 60% of messages have a significant skew towards heavier use at the beginning of the lifetime, while the kurtosis indicates that these numbers see a sharp increase in activity followed by steep decline. (c) 60% of numbers used show a strong tendency for heavy use in the early lifetime of the number.

which may be predictable and grant unauthorized access to accounts. The design of such services is guided by an assumption that the SMS channel is secure from external observation, and our observations show that this results in poor security design in those applications. Last, during our collection period, NIST deprecated the use of SMS as an authentication method (Grassi et al. 2016). However, we continued to see messages containing authentication codes from various services.

6 ABUSES OF SMS

Having explored how services attempt to use SMS as a secure channel, we now discuss what we observed about the security implications and evidence of abuse related to gateway activity. This includes phone verified account evasion, failed attempts at location anonymity, whether similar gateway numbers can be detected, spam, and the global reach of malicious behavior. Additionally, we apply predictive analytics to the the various SMS activities we have discussed throughout the article.

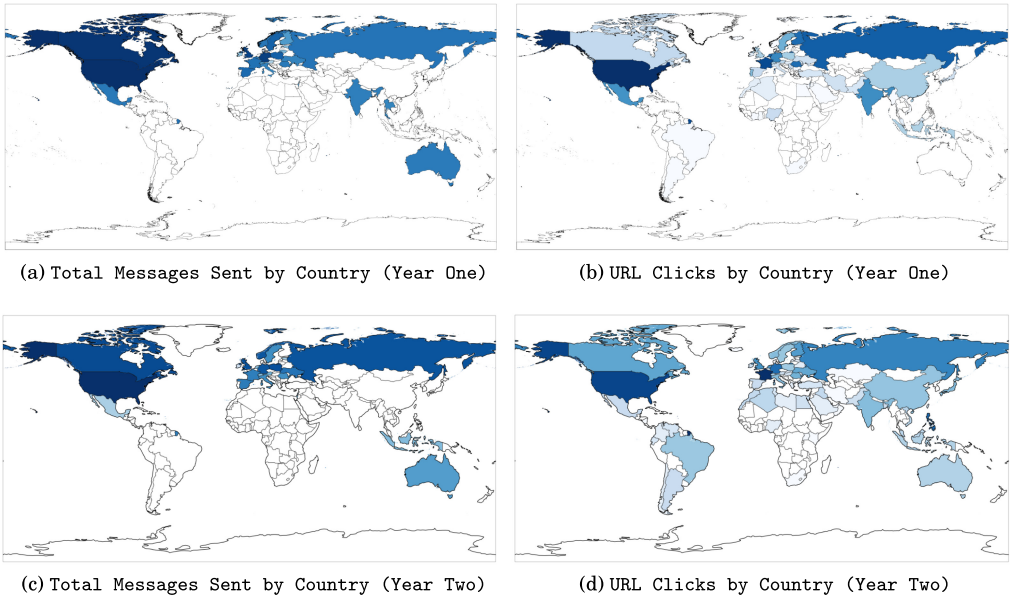
6.1 Gateways and PVA

In this subsection, we discuss the relevance of our data to phone-verified accounts (PVA). In particular, we present evidence that the primary activity of the gateways we observe is evading PVA restrictions, and that existing countermeasures are ineffective.

Message Activity Statistics. In Section 4, we noted that more than half of the messages received by gateways are related to account verification. This vastly outweighed any other purpose of sending SMS. Beyond this information, message activity statistics also support this claim. The median number lifetime (the time from first message to last) in our dataset is 24 days, and the CDF of number lifetime is shown in Figure 5(a). This lifetime is fairly short, and in fact 62.3% of numbers do not even last a full billing cycle (31 days).

There are two likely explanations for the short lifetime: one is that services that facilitate PVA need to replace their numbers often as they exhaust their usefulness to create new accounts. The second is that many of these numbers are in carriers (especially mobile carriers) that shut off numbers for anomalous message volume. These explanations are not necessarily mutually exclusive.

To gain insight onto this question, we computed the daily volume of messages for each phone number used by a gateway, and we call this series the "daily activity" of the number. If these numbers were being primarily for personal messages or informational activities (like signing up for advertising alerts), then we would expect the daily activity of the number to be fairly constant



(a) Total Messages Sent by Country (Year One)
(b) URL Clicks by Country (Year One)
(c) Total Messages Sent by Country (Year Two)
(d) URL Clicks by Country (Year Two)

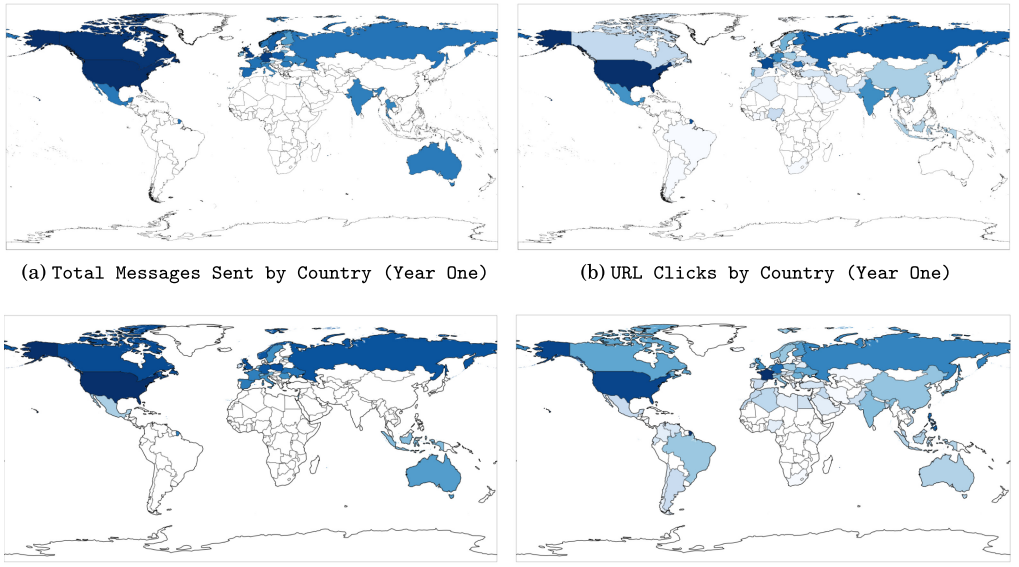
Fig. 6. These maps visualize the sender phone number locations of all messages (a and c) sent to the gateways and the locations of users that have clicked shortened URLs (b and d) for the first and second year. Overall, the locations of the gateways' users significantly differ from the services sending messages, implying the primary purpose of these gateways is PVA fraud.

across the lifetime of the number, or for there to be a "ramp up" period as new users discover the new line.

Instead, we see almost the exact opposite behavior. To concisely express this, we computed skewness and kurtosis statistics of the daily activity of every number. Simply, kurtosis is a statistic that indicates if a series is "flat" or "peaky," while skewness indicates whether a peak falls closer to the middle, beginning, or end of a series. A skewness of between $(-1, 1)$ indicates the peak falls in the middle of the series, while a positive skewness indicates a peak that arrives "earlier" in the series. We plot the skewness and kurtosis for every number in Figure 5(b). Note that we reverse the x-axis, so that the further left in the plot a number falls, the "earlier" its peak.

Figure 5(c) shows the CDF of the daily activity skewness, and we observe that approximately 60% of numbers have a skewness towards early activity. This implies that most numbers have a high message volume early in the lifetime, and consequently, most of the activity of the number has been completed by the time it is shut down. If carriers are disabling numbers (for exceeding a message rate cap, for example), then they are doing so well after most numbers have seen their peak use. Likewise, if online services are considering a number invalid for phone verification (e.g., multiple accounts with the same phone number), they are still permitting a high-volume of registration requests for a number (in aggregate) before blacklisting.

User Location Leakage. Some gateways advertise their services for users that are seeking privacy or anonymity. Although SMS does not provide these properties, the use of a gateway may provide a sense of anonymity for a user registering for a service. Shortened URLs (often provided in space-constrained SMS messages) leak information about the user clicking the link to the URL-shortening service. With the statistics we collected from these services, we have identified both the source and destination countries for each message, we also found that the *users* of these services are located



(a) Total Messages Sent by Country (Year One)

(b) URL Clicks by Country (Year One)

(c) Total Messages Sent by Country (Year Two)

(d) URL Clicks by Country (Year Two)

Fig. 7. These maps visualize the sender phone number locations of all messages (a and c) sent to the gateways and the locations of users that have clicked shortened URLs (b and d) for the first and second year. Overall, the locations of the gateways' users significantly differ from the services sending messages, implying the primary purpose of these gateways is PVA fraud.

in significantly different locations. *We do not attempt to deanonymize, track, or identify any users.* Our data consists solely of publicly-available aggregate click statistics.

The number of clicks recorded ranged from all shortened URLs found range from 0–4,995,788 with a median of 8. This data represents *any* click to these URLs, not just those from the gateway pages. As a result, to prevent skewing our data with popular URLs and spam messages, we focused on URLs with ≤10 clicks, since many incoming links expected by users of SMS gateways are likely clicked a small number of times. We collected the countries associated with each of the remaining clicks and aggregated the results. Figure 7 shows the total clicks for each country across all shortened URLs.

We split our dataset into the messages included in our previous study (Reaves et al. 2016a) and the new messages gathered after that study to measure changes in long-term location trends. As before, we collected both the location of the source phone number and the click statistics of the links for the new messages. To ensure that the click metrics are reflecting the period of when they were gathered, and not the clicks of the previous study, we computed the Jaccard similarity of both sets and only found a 0.12% similarity, indicating that these two sets of URLs are largely disjoint.

In Figures 7(b) and 7(d), we can see that the click metrics obtained from URL shorteners indicate that gateway users are coming from more countries than before. We note that the color intensity of both maps are relative to the individual dataset graphed, and as such, comparing color intensity does not fully capture the traffic changes. We analyzed the raw figures and found that on year two, there were 29 new countries while 12 countries were no longer available. Additionally, of the countries that were available in both years, 9 showed a decrease in traffic by over 50%. The churn in countries shows us a volatile user base of the recipients of gateway messages. While the gateway user base comes from a more diverse set of countries, if we compare the location of the

sender's phone number of each message found in both years (Figures 7(a) and 7(c)), we actually see fewer countries sending messages, especially in Southeast Asia. Although the locations of messages and users have changed, the central finding in our initial study that messages and the users that consume them are in different countries remains true.

6.2 Detecting Gateways

As we have discussed above, these gateways facilitate PVA evasion and the demographic data we can obtain about the users of these services clearly shows usage patterns consistent with PVA fraud. It is clear that in most cases even reputable well-funded online services are not successfully defending against these gateways. Although number lifetimes are short, the sheer volume of verification messages in our data indicates that evasion is still an effective driver of profit for gateways.

PVA evasion is not new to online services. In particular, Google is acutely aware of this problem, having published a paper on the topic (Thomas et al. 2014). In that paper, Thomas et al. propose several strategies to detect PVA evasion. They include blocking irreputable carriers, restricting how quickly numbers can verify accounts, and phone re-verification. In this section we explore the recommendations in that work and discuss how our data shows that *these recommendations are unlikely to be effective*:

Carrier Reputation. While we only see one of the carriers identified as abuse-prone in Thomas et al. (2014) (Bandwidth), blacklisting blocks of numbers by carrier would not stop all PVA evasion. Carrier-based blocking is prohibitively expensive for all but the largest of organizations. We obtained Twilio data for each number in our data set and although the cost was relatively small ($0.005/lookup), scaling this (and additional number metadata such as CNAM and HLR data) to cover all of a business' customers represents a substantial cost. Furthermore, this kind of bulk blacklisting is difficult to enforce in the face of gateway services that maintain a large pool of numbers over many carriers. Online services that attempt to restrict the speed at which numbers can be reused for new accounts face an arms race against gateways.

Phone Reputation. One option suggested in Thomas et al. (2014) for determining phone reputation is to create a service that shares abuse data between service providers. Although there is little information about how such a service could be created, we considered that it might be possible to blacklist abusive numbers if they are similar to each other.

We conducted a self-similarity analysis against the phone numbers in our dataset to determine how numbers are purchased. If they are purchased in bulk, then it may be possible to detect them. We analyzed all of the gateways' numbers to determine similar numbers using Hamming distance. We found that most carriers use similar numbers (i.e., those with a Hamming distance of two or less), and the results are shown in Table 6. Over 40% of all of a gateway's numbers were similar in 6 of 8 gateways, however we found that most of these repeated numbers are in *mobile* carriers, not VoIP, as shown in Table 7. The data shows that the gateway numbers are in the carriers that are most likely to serve legitimate users, so attempting to block these numbers may result in a high false positive rate. Furthermore, as shown in Table 6, we reexamined this analysis a year later to see if previous trends still hold true. We find only a marginal decrease in average number similarity (from 43.7% to 41.3%), so blocking similar numbers would still result in a high false positive rate.

Phone Re-verification. Phone number re-verification would fail if the number were checked again outside the expected lifetime of a gateway number. Thomas et al. (2014) saw a median number lifetime of one hour, a reasonable point to perform a re-verification. In our dataset, however, we have seen that half of all gateway numbers last *up to 24 days*. Therefore, re-verification at any interval is unlikely to be universally effective, since phone number longevity is not guaranteed.

Table 6. We Analyzed the Numbers from Each Gateway for Similarity in Both Studies

	Before		After	
Site	Similar/Total	Percentage	Similar/Total	Percentage
[1] receive-sms-online.info	15/59	25.4%	33/129	25.6%
[2] receivesmsonline.net	16/38	42.1%	18/55	32.7%
[3] e-receivesms.com	7/14	50.0%	7/19	36.8%
[4] hs3x.com	28/57	49.1%	63/110	57.3%
[5] receivefreesms.com	52/93	55.9%	62/134	46.2%
[6] receivesmsonline.com	38/93	40.9%	59/138	42.8%
[7] receive-sms-online.com	8/19	42.1%	12/32	37.5%
[8] receive-sms-now.com	20/48	48.0%	26/61	42.6%
Overall	184/421	43.7%	280/678	41.3%

Finding that on overage, over 41% of numbers are similar in both years.

Table 7. An Analysis of the Similarity of Gateway Numbers Shows that the Majority of Numbers are in Mobile Carrier Number Blocks, not VoIP as We Expected

	Before		After	
Carrier Type	Similar/Total	Percentage	Similar/Total	Percentage
Mobile	159/184	86.4%	213/250	85.2%
Landline	5/184	2.7%	12/250	4.8%
VoIP	20/184	10.9%	25/250	10.0%

As a result, attempting to block these number blocks may result in high false positives. After collecting more data over time, we still see a similarity in mobile phone numbers.

6.3 Abuse Campaigns in SMS

Since gateways accept unsolicited messages, often do not filter messages, and are subject to users providing these numbers to various services, our data contains SMS from SPAM campaigns, phishing campaigns, and even one black market, as discussed in Section 5.1. In this section, we will discuss these campaigns.

6.3.1 Spam Campaigns. Our spam analysis used two approaches to identify spam. The first approach uses the clusters identified as spam that we previously generated and manually classified in our original study (Reaves et al. 2016a). The second approach is based on developing a machine learning-based automatic spam classifier (Reaves et al. 2016b).

Cluster Analysis. With this method, we found 1.0% of tagged messages across 32 clusters related to advertising. Upon manual inspection none of these appeared to be solicited messages, so we consider these to be spam messages. Of the advertising clusters we identified, 15 are UK-based financial services (e.g., payday loans, credit lines) from 14 numbers. Five are for distinct bulk messaging services. These services advertise gateways and the ability to avoid phone verification: "Using our service to create and verify accounts without your own phone number."

Another six clusters are from a specific job staffing site and appear to be bulk messages related to a job search. Curiously, these messages contain a name and zip code. We expanded the search beyond the labeled clusters and found 282 messages in 107 clusters. These messages may be related to this organization testing their bulk SMS API. All of these messages were sent to a single gateway number within a seven-hour timespan, which is unusual when compared to other bulk message campaigns in our dataset.

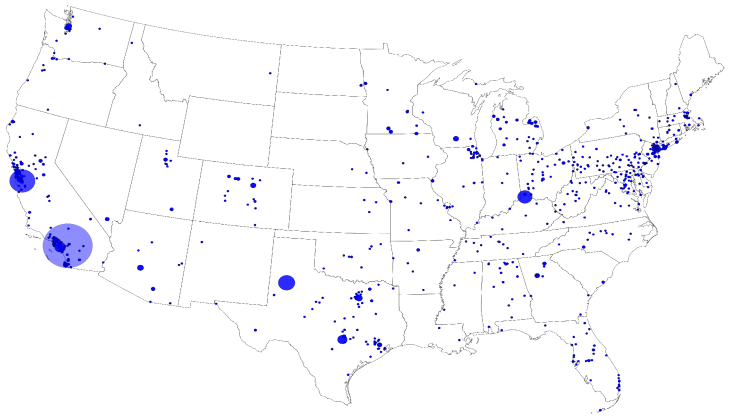


Fig. 8. This map shows the density of spam messages transiting through public gateways based on the source phone numbers.

We were surprised at the low spam volume observed in public gateways, as they market themselves as a service for avoiding spam. This has been a major topic of research, but the volume of spam traffic in our dataset is lower than previously measured (Delany et al. 2012; Skudlark 2014).

Spam Classifier. To get a better representation of the actual spam present in our dataset, we used the method proposed by Reaves et al. (2016b) to classify spam. This method works by using a support vector machine (SVM). To extract features, it uses a simple binary vector to indicate the presence of certain words that may relate to spam. This method can classify spam with a precision and recall rate of 100% and 96.6%, respectively. This method tagged 2.4% of all the messages as spam. We use these more comprehensive results for the following subsection.

6.3.2 Spam Transiting Through Public Gateways. In total, we identified over 22,000 spam messages in our dataset, and an important question is where these messages come from. While we used Twilio's Lookup service for our previous analysis of gateway-owned phone numbers, the sheer volume of messages and Twilio's per-number fee structure made using that service impractical. Instead, to better understand the geography of these messages, we made use of the fact that in the United States it is possible to identify a source city directly from the phone number itself. Specifically, the first six digits of a phone number identify both the area code and the central office that serves a call. For example, numbers of the form 305-200-XXXX are based in Miami. The mapping of phone number prefix to location is known as the North American Numbering Plan (NANP) and is maintained in a public database by the North American Numbering Plan Administration (NANPA). We note that this method is not perfect; numbers that send spam can be ported or spoofed entirely. We believe this issue to be minimal, because spam in many cases has an incentive to not spoof (e.g., to receive replies) or to spoof a number local to the target. As a result, this analysis still provides useful insights into spam behavior.

In Figure 8, we show how spam present in public gateways transit through the United States. As expected, spam is mostly proportional to the population density of large cities and metropolitan areas. The central office with the most spam transiting through it is located in Riverside, California. Most of the messages transiting through this central office were attempting to advertise available jobs.

Case Study: New York City. Large cities are usually managed by multiple central offices to accommodate the large subscriber density. When we looked at the source phone numbers of the

Table 8. The Amount of Malicious URLs Marked Positive by VirusTotal for All Languages that Had More than 1,000 Unique URLs

Language	Malicious URLs	Unique URLs	Percentage
Russian	56	1,011	5.4%
English	782	36,614	2.1%
Italian	14	1,051	1.3%
French	18	1,513	1.1%
Norwegian	15	1,249	1.1%
German	27	3,098	0.1%

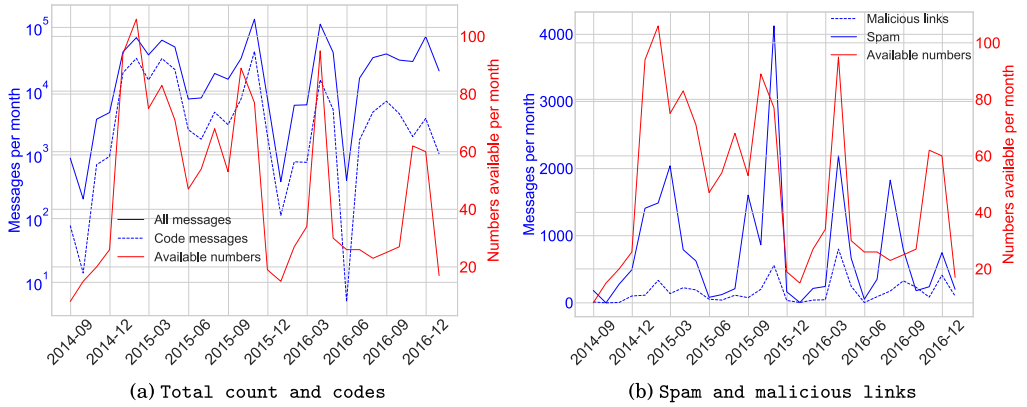
spam messages in these cities, we saw that New York City had 58 unique numbers sending spam. As a comparison, most cities saw fewer than 20 unique numbers that sent spam. Out of those 58 unique numbers sending spam from New York City, 24 were handled by a single central office. The messages from the numbers in this central office were diverse in content indicating multiple spam campaigns. We saw that, on average, other central offices hosted one or two numbers that contained messages from spam campaigns. Based on the spam messages transiting through public gateways, New York City has the most diverse targets and diverse content distribution of spam out of the metropolitan areas in the United States.

We also looked at the service providers of these numbers to see if we can use this information to better detect spam flow. The spam phone numbers sending spam were distributed among multiple service providers. However, the service provider with the most phone numbers used for spam was Level 3 Communications with around 14% of all the numbers. As a comparison, the next highest service provider only accounted for half of that amount (7%).

As mentioned earlier, 2.4% of all messages in our dataset were classified as spam. This figure does not tell us anything about the changes that spam messaging may have in the 28-month span of our collection period. Therefore, we compared the volume of spam found in the original study and in the messages that followed. We found 11,193 spam messages (2.88%) in the first study and 11,010 spam messages (2.15%) in the data collected in the 14 months following the first study. This very small change in spam volume indicates that our original findings of spam volume have been stable and show no strong indicators of broader change during our study period.

6.3.3 Link Analysis and Malicious Behaviour. Another empirical measure of the maliciousness of the URLs is scanning these URLs with security products. VirusTotal provides one such measure by requesting scans from multiple products. The results from VirusTotal returned 1393 URLs with at least one detection. Only 6 URLs had 5 detections, and no URL had more than 5 detections. Of these detections, 958 were detected as "malicious site," 738 as "malware site," and 64 as "phishing site."

We further filtered the URLs to analyze how susceptible a language community is to malicious activities based on infected URLs found in the messages. To do this, we clustered the messages by language and extracted all of the unique links from each language. If a link was present in more than one language, then we included that link in the statistics for each language where it was present. A link was recognized as malicious if at least one scanner from VirusTotal returned a positive result. Although we collected results for all the languages found in our dataset, in Table 8, we only show the results for languages that had over 1,000 unique URLs. We do this because we do not have enough unique samples to make strong claims for the other languages. For example, we found that Swahili had the highest malicious activity rate at 15.3%, but we cannot claim this to



(a) Total count and codes (b) Spam and malicious links

Fig. 9. SMS behaviors (black and blue) in public gateways are stationary and have a strong correlation to the amount of available numbers in a given month (red).

be representative of the actual language community, because the sample size was only 13 unique links.

From the languages present in Table 8, English had an order of magnitude more infected links that any other language. However, due to English being used as the primary language for many services, we see that malicious links only account for 2.1% of the unique URLs for this language. We noticed that many of the links found here were used for one-time confirmation, verification, and password resets. Furthermore, we found URL messages sent in the Russian language had the highest probability of containing malicious content: 5.4% of URLs in that language. We found that at least one of these seemed to be trying to spread an Android APK file while impersonating IBMs Security Trusteer Rapport, software that *"helps prevent malware and phishing attacks that are the root cause of most financial fraud"* (IBM Security Trusteer Rapport 2016).

Overall, abusive messages (spam, phishing, and malware) comprised only a small portion of our dataset, despite being billed as a major problem in popular press. This is especially strange given that evasion of spam is something many of the gateways advertise, as we discussed in Section 3. Given previous reports on the pervasiveness of SMS spam, we believe that some entity in the SMS ecosystem is performing adequate spam filtering and that this problem may no longer be as severe as it once was.

6.4 Predictive Analyses

Throughout the article, we measured various behaviors of the SMS ecosystem. In this section, we look into predicting future trends based on the monthly time series of messages gathered. These trends include available phone number volume, message volume (i.e., total count and code messages), and abuse volume (i.e., spam and malicious links). We find that while all trends examined are likely stationary (stable) over our observation time period, short term predictions based on prior activity are not accurate.

In Figure 9, we show the aggregate monthly count of each of the time series. Qualitatively, we see that each studied behavior seems to be influenced by the amount of available phone numbers for each time period. As such, predicting the availability of phone numbers can provide information about the other behaviors of interest. To verify, we set $\alpha = 0.05$ and evaluated the Spearman rank correlation coefficient of the available phone numbers to all measured SMS behaviors[12] and found

[12]Pearson correlation is not applicable, because none of the time series mentioned above follow a normal distribution.

a strong positive monotonic correlation ($\rho > 0.60$) for all compared series ($p < 0.001$). This insight means that models that predict available phone numbers should also work for predicting these other behaviors.

Before making predictive models for such behaviors, we first need to determine if the monthly available phone numbers series is stationary or non-stationary. As such, we used the Augmented Dickey-Fuller (ADF) test with $\alpha = 0.05$ and determined that the available numbers series is stationary ($p = 0.028$). This tells us that the publication of phone numbers from the public gateways do not follow a certain trend nor dependency on seasonality.[13]

Finally, since the time series was stationary, we looked into future short term predictions using ARMA models for the available phone numbers. Unfortunately, there are many variables in the publication of phone numbers in the public gateways that are unknown to us, making the ARMA model a bad month-to-month predictor.

While we can not predict month-to-month behavior of these time series, the fact that these series are stationary implies that we should expect gateways to operate at the same volume of numbers, messages, and associated abuse for the foreseeable future.

6.5 Takeaways

In this section, we explored malicious uses of the SMS channel. First, we discussed how our data shows the prevalence of PVA evasion due to the stark contrast between gateway number locations and locations of users interacting with the gateways. We then discussed the difficulty of detecting gateways with carrier blocking due to cost and number lifetimes. Next, we explored abuse campaigns via SMS and found that spam, and suspicious URLs are infrequent, which may indicate that SMS filtering at the gateways and in the network are sufficient. Finally, we found minimal changes were made to the SMS ecosystem during the 28 months of our data collections. This last finding indicates that previous problems are still present in the SMS ecosystem.

7 RELATED WORK

Prior measurement work has studied the underground economies (Thomas et al. 2015) that drive spam (Kanich et al. 2008, 2011; Thomas et al. 2013), malware (Stone-Gross et al. 2009; Cho et al. 2010; Grier et al. 2012) and mobile malware (Felt et al. 2011; Zhou and Jiang 2012; Lever et al. 2013), and other malicious behavior. While others have investigated SMS content and metadata in the context of SMS spam (Murynets and Piqueras Jover 2012; Tan et al. 2012; Jiang et al. 2013; Narayan and Saxena 2013), this work is the first to expansively measure how SMS is used for security purposes by legitimate services. We note that much of the research in this area has been forced to rely on small datasets (some less than 2,000 messages (Narayan and Saxena 2013)). Mobile two-factor authentication is increasing in popularity, with some eagerly heralding its arrival (Atwood 2012) and others cautioning that it may only provide a limited increase in security (Schneier 2005). Much of the data we collected contained mobile two-factor authentication tokens sent over SMS. While SMS tokens are popular in many contexts, including mobile banking and finance (Reaves et al. 2015), other approaches have been implemented in a variety of forms including keychain fobs (SecurID 2015; IdentityGuard 2015), one-time pads (Leyden 2008; SiPix Imagining, Inc. 2006), biometric scanners (Stensgaard 2006; CardTechnology 2007), and mobile phones (Aloul et al. 2009; DeFigueiredo 2011; Duo Mobile 2015). Analysis of individual systems has led to the discovery of a number of weaknesses, including usability concerns (Adham et al. 2013) and susceptibility to desktop (Konoth et al. 2016) or mobile malware (Castillo 2011; Koot 2012; Mulliner et al. 2013;

[13]We note that we also used the ADF test with $\alpha = 0.05$ for all other time series and determined that each series was also stationary ($p < 0.01$).

Koenig et al. 2013; Dmitrienko et al. 2014; Eide 2015). SMS-based tokens are especially vulnerable to link-layer attacks against the cellular network. These networks use vulnerable channel encryption (Biryukov et al. 2001; Barkan et al. 2007; Dunkelman et al. 2010), allow end devices to connect to illicit base stations (Ahmadian et al. 2009; Golde et al. 2012; Dabrowski et al. 2014), and are vulnerable to low-rate denial of service attacks (Traynor et al. 2007, 2008, 2009). However, the majority of the infrastructure behind many two-factor authentication systems—the portions of the system outside the cellular network—has not been previously explored from a security perspective. Additionally, while end-to-end SMS encryption schemes exist to prevent content exposure (Saxena and Chaudhari 2014; Saxena et al. 2018; De Santis et al. 2010), such precautions become ineffective when one endpoint is either compromised or is willing to publicize the content of the messages (e.g., public gateways).

Dmitrienko et al. were among the first to examine SMS messages to study security of two-factor authentication schemes (Dmitrienko et al. 2014). We greatly exceed the scope of their work in five important ways. First, our work presents a comprehensive examination of the entire SMS infrastructure—from online services to end devices. Second, we focus on how online services use SMS well beyond two-factor authentication. Third, our data includes two orders of magnitude more services and we identify and classify the intent of each message. Fourth, we provide a more detailed classification of two-factor authentication systems. Finally, our more rigorous entropy analysis of two-factor authentication PINs allow us to make strong claims for more than 30 services (instead of just 3), helping us to find egregious entropy problems in the popular WeChat and Talk2 services.

Our emphasis on phone verified accounts provides a separate contribution. Thomas et al. study the effects of phone verified accounts at Google (Thomas et al. 2014). While they use datasets of purchased or disabled PVAs, we provide insight into PVA fraud from enabling services. While we confirm some of their observations, our data indicated their recommendations may prove ineffective at defeating PVA evasion.

8 CONCLUSIONS

Text messaging has become an important part of the security infrastructure. However, this ecosystem has evolved significantly since its inception, and now includes a wide range of devices and participants external to traditional cellular providers. Public SMS gateways directly embody this change and allow us to not only observe at scale how a range of providers are implementing security solutions via text messages but also provide us evidence of how assumptions about SMS are being circumvented in the wild. While our data may not fully encompass all communications sent over SMS, our measurements identify a range of popular services whose one-time messaging mechanisms should be improved, and additional entities who may be creating new opportunities for compromise by sending highly sensitive data (e.g., credit card numbers) via these channels. On the abuse side, we see the ease with which these gateways are being used to circumvent authentication mechanisms, and show that previously proposed mitigations to PVA fraud such as block banning are unlikely to be successful in practice. These measurements indicate that all providers relying on SMS as an out of band channel for authentication with strong ties to a user's identity should reevaluate their current solutions for this evolving space. However, we find no discernible changes were made to the SMS ecosystem during our collection period; this suggests previous problems are still persistent.

ACKNOWLEDGMENTS

The authors are grateful to our anonymous reviewers for their helpful guidance. The authors thank Twilio and VirusTotal for their generous access to their data and Benjamin Mood for providing considerable assistance formatting our tables and figures. Any opinions, findings, and conclusions

or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Manal Adham, Amir Azodi, Yvo Desmedt, and Ioannis Karaolis. 2013. How to attack two-factor authentication internet banking. In *Financial Cryptography and Data Security*. No. 7859 in Lecture Notes in Computer Science. Springer, Berlin, 322–328.

Z. Ahmadian, S. Salimi, and A. Salahi. 2009. New attacks on UMTS network access. In *Proceedings of the Wireless Telecommunications Symposium (WTS'09)*. 1–6.

F. Aloul, S. Zahidi, and W. El-Hajj. 2009. Two factor authentication using mobile phones. In *Proceedings of the IEEE/ACS International Conference on Computer Systems and Applications (AICCSA'09)*. 641–644.

Apple Continuity 2015. Apple continuity. Retrieved from https://support.apple.com/en-us/HT204681.

Jeff Atwood. 2012. Make your email hacker proof. *Coding Horror*. Retrieved from http://blog.codinghorror.com/make-your-email-hacker-proof/.

Elad Barkan, Eli Biham, and Nathan Keller. 2007. Instant ciphertext-Only cryptanalysis of GSM encrypted communication. *J. Cryptol.* 21, 3 (Sep. 2007), 392–429.

Alex Biryukov, Adi Shamir, and David Wagner. 2001. Real time cryptanalysis of A5/1 on a PC. In *Proceedings of the 7th International Workshop on Fast Software Encryption (FSE'00)*. Springer-Verlag, London, 1–18.

Boss Revolution 2015. Boss revolution. Retrieved from https://www.bossrevolution.ca.

Burner 2015. Burner app. Retrieved from http://www.burnerapp.com.

Kelsey Campbell-Dollaghan. 2014. How hackers reportedly side-Stepped Google's two-factor authentication. *Gizmodo*. Retrieved from http://gizmodo.com/how-hackers-reportedly-side-stepped-gmails-two-factor-a-1653631338.

CardTechnology. 2007. UAE ID card to support iris biometrics. Retrieved from http://www.cardtechnology.com/article.html?id=20070423V0XCZ91L.

Carlos Castillo. 2011. Spitmo vs zitmo: Banking trojans target android. *McAfee Labs Blog*. Retrieved from http://blogs.mcafee.com/mcafee-labs/spitmo-vs-zitmo-banking-trojans-target-android.

Chia Yuan Cho, Juan Caballero, Chris Grier, Vern Paxson, and Dawn Song. 2010. Insights from the inside: A view of botnet management from infiltration. In *Proceedings of the USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET'10)*.

Nathan Crooks. 2015. Venezuela, the country with four exchange rates. *Bloomberg Business*. Retrieved from http://www.bloomberg.com/news/articles/2015-02-19/venezuela-the-country-with-four-exchange-rates.

Adrian Dabrowski, Nicola Pianta, Thomas Klepp, Martin Mulazzani, and Edgar Weippl. 2014. IMSI-Catch me if you can. In *Proceedings of the 30th Annual Computer Security Applications Conference*.

Alfredo De Santis, Aniello Castiglione, Giuseppe Cattaneo, Maurizio Cembalo, Fabio Petagna, and Umberto Ferraro Petrillo. 2010. An extensible framework for efficient secure SMS. In *Proceedings of the International Conference on Complex, Intelligent and Software Intensive Systems (CISIS'10)*. IEEE, 843–850.

D. DeFigueiredo. 2011. The case for mobile two-Factor authentication. *IEEE Secur. Privacy Mag.* 9, 5 (Sep. 2011), 81–85.

Sarah Jane Delany, Mark Buckley, and Derek Greene. 2012. SMS spam filtering: Methods and data. *Expert Syst. Appl.* 39, 10 (2012), 9899–9908.

Alexandra Dmitrienko, Christopher Liebchen, Christian Rossow, and Ahmad-Reza Sadeghi. 2014. On the (in)security of mobile two-factor authentication. In *Proceedings of the Conference on Financial Cryptography and Data Security (FC'14)*. Springer.

Orr Dunkelman, Nathan Keller, and Adi Shamir. 2010. A Practical-time Related-key attack on the KASUMI cryptosystem used in GSM and 3G telephony. In *Proceedings of the 30th Annual Conference on Advances in Cryptology (CRYPTO'10)*. Springer-Verlag, Berlin, 393–410.

Duo Mobile 2015. Mobile authentication. *Duo Mobile*. Retrieved from https://www.duosecurity.com/product/methods/duo-mobile.

eCall 2015. eCall. Retrieved from http://www.ecall.ch.

Jan-Erik Lothe Eide. 2015. *SMS One-Time Passwords: Security in Two-Factor Authenication*. Master's Thesis. University of Bergen.

Adrienne Porter Felt, Matthew Finifter, Erika Chin, Steve Hanna, and David Wagner. 2011. A survey of mobile malware in the wild. In *Proceedings of the ACM Workshop on Security and Privacy in Mobile Devices*.

Brendan J. Frey and Delbert Dueck. 2007. Clustering by passing messages between data points. *Science* 315, 5814 (2007), 972–976.

Frim 2015. Frim. Retrieved from http://fr.im.

Nico Golde, Kevin Redon, and Ravishankar Borgaonkar. 2012. Weaponizing femtocells: The effect of rogue devices on mobile telecommunications. In *Proceedings of the 19th Network and Distributed System Security Symposium*.

Google Voice 2015. Google voice. Retrieved from http://www.google.com/voice.

Paul A. Grassi, James L. Fenton, Elaine M. Newton, Ray A. Perlner, Andrew R. Regenscheid, William E. Burr, Justin P. Richer, Naomi B. Lefkovitz, Jamie M. Danker, YeeYin Choong et al. 2016. DRAFT NIST special publication 800 63B digital identity guidelines. Retrieved from https://pages.nist.gov/800-63-3/sp800-63b.html.

Chris Grier, Lucas Ballard, Juan Caballero, Neha Chachra, Christian J. Dietrich, Kirill Levchenko, Panayiotis Mavrommatis, Damon McCoy, Antonio Nappa, Andreas Pitsillidis, Niels Provos, M. Zubair Rafique, Moheeb Abu Rajab, Christian Rossow, Kurt Thomas, Vern Paxson, Stefan Savage, and Geoffrey M. Voelker. 2012. Manufacturing compromise: The emergence of exploit-as-a-service. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security (CCS'12)*. ACM, New York, NY, 821–832.

Stephen Hilt and Philippe Lin. 2016. Leaking beeps: Unencrypted pager messages in the healthcare industry. In *TrendLabs Research*. Retrieved from https://documents.trendmicro.com/assets/threat-reports/wp-leaking-beeps-healthcare.pdf.

IBM Security Trusteer Rapport 2016. IBM security trusteer rapport. Retrieved from http://www-03.ibm.com/software/products/en/trusteer-rapport.

IdentityGuard 2015. IdentityGuard identity authentication platform. *Entrust, Inc.* Retrieved from https://www.entrust.com/products/entrust-identityguard/.

Nan Jiang, Yu Jin, Ann Skudlark, and Zhi-Li Zhang. 2013. Greystar: Fast and accurate detection of SMS spam numbers in large cellular networks using grey phone space. In *Proceedings of the 22nd USENIX Security Symposium*. USENIX Association, Washington DC.

Chris Kanich, Christian Kreibich, Kirill Levchenko, Brandon Enright, Geoffrey M. Voelker, Vern Paxson, and Stefan Savage. 2008. Spamalytics: An empirical analysis of spam marketing conversion. In *Proceedings of the 15th ACM Conference on Computer and Communications Security*. ACM, 3–14.

Chris Kanich, Nicholas Weaver, Damon McCoy, Tristan Halvorson, Christian Kreibich, Kirill Levchenko, Vern Paxson, Geoffrey M. Voelker, and Stefan Savage. 2011. Show me the money: Characterizing spam-advertised revenue. In *Proceedings of the 2015 USENIX Security Symposium*.

Karsten Nohl. 2016. SS7 attack update and phone phreaking. In *GeekFest Berlin*. Retrieved from https://www.youtube.com/watch?v=BbPLscWQ1Bw.

Reto E. Koenig, Philipp Locher, and Rolf Haenni. 2013. Attacking the verification code mechanism in the norwegian internet voting system. In *E-Voting and Identity*, James Heather, Steve Schneider, and Vanessa Teague (Eds.). Springer, Berlin, 76–92.

Radhesh Krishnan Konoth, Victor van der Veen, and Herbert Bos. 2016. How anywhere computing just killed your phone-based two-factor authentication. In *Proceedings of the 20th International Conference on Financial Cryptography and Data Security*.

Laurens Koot. 2012. *Security of Mobile TAN on Smartphones*. Master's Thesis. Radboud University Nijmegen, Nijmegen.

Brian Krebs. 2014. Banks: Credit card breach at Home Depot. *Krebs on Security*. Retrieved from http://krebsonsecurity.com/2014/09/banks-credit-card-breach-at-home-depot/.

Brian Krebs. 2015a. Experian breach affects 15 million consumers. *Krebs on Security*. Retrieved from http://krebsonsecurity.com/2015/10/experian-breach-affects-15-million-consumers/.

Brian Krebs. 2015b. Online cheating site ashleymadison hacked. *Krebs on Security*. Retrieved from http://krebsonsecurity.com/2015/07/online-cheating-site-ashleymadison-hacked/.

Charles Lever, Manos Antonakakis, Bradley Reaves, Patrick Traynor, and Wenke Lee. 2013. The core of the matter: Analyzing malicious traffic in cellular carriers. In *Proceedings of the 20th Network and Distributed System Security Symposium*.

John Leyden. 2008. Visa trials PIN payment card to fight online fraud. Retrieved from http://www.theregister.co.uk/2008/11/10/visa_one_time_code_card/.

Luca Melette. 2016. Effective SS7 protection. In *Proceedings of the ITU Workshop on SS7 Security*.

Hans Peter Luhn. 1960. Computer for verifying numbers. Retrieved from https://www.google.com/patents/US2950048 U.S. Patent 2,950,048.

Erika McCallister, Tim Grance, and Karen Scarfone. 2010. Guide to protecting the confidentiality of personally identifiable information (PII). *Recommendations of the National Institute of Standards and Technology*. Retrieved from http://csrc.nist.gov/publications/nistpubs/800-122/sp800-122.pdf.

MightyText 2015. MightyText. Retrieved from http://mightytext.net.

Collin Mulliner, Ravishankar Borgaonkar, Patrick Stewin, and Jean-Pierre Seifert. 2013. SMS-based one-time passwords: Attacks and defense. In *Detection of Intrusions and Malware, and Vulnerability Assessment*. Springer, 150–159.

Ilona Murynets and Roger Piqueras Jover. 2012. Crime scene investigation: SMS spam data analysis. In *Proceedings of the 2012 ACM Conference on Internet Measurement Conference (IMC'12)*. ACM, New York, NY, 441–452.

Akshay Narayan and Prateek Saxena. 2013. The curse of 140 characters: Evaluating the efficacy of SMS spam detection on android. In *Proceedings of the Third ACM Workshop on Security and Privacy in Smartphones & Mobile Devices (SPSM'13)*. ACM, New York, NY, 33–42.

Nexmo 2015. Nexmo. Retrieved from https://www.nexmo.com/.

OpenCNAM 2015. OpenCNAM. Retrieved from https://www.opencnam.com.

PayCenter 2015. PayCenter. Retrieved from https://www.paycenter.de.

Christian Peeters, Hadi Abdullah, Nolen Scaife, Jasmine Bowers, Patrick Traynor, Bradley Reaves, and Kevin R. B. Butler. 2018. Sonar: Detecting SS7 redirection attacks with audio-based distance bounding. In *Proceedings of the IEEE Symposium on Security and Privacy (SP'18)*. 86–101.

Pinger 2015. Pinger. Retrieved from http://www.pinger.com.

Plivo 2015. Plivo. Retrieved from https://www.plivo.com/.

Pushbullet 2015. Pushbullet. Retrieved from http://pushbullet.com.

Bradley Reaves, Logan Blue, Dave Tian, Patrick Traynor, and Kevin R. B. Butler. 2016b. Detecting SMS spam in the age of legitimate bulk messaging. In *Proceedings of the 9th ACM Conference on Security & Privacy in Wireless and Mobile Networks*. ACM, 165–170.

Bradley Reaves, Nolen Scaife, Adam Bates, Patrick Traynor, and Kevin Butler. 2015. Mo(bile) money, mo(bile) problems: Analysis of branchless banking applications in the developing world. In *Proceedings of the USENIX Security Symposium (SECURITY'15)*.

Bradley Reaves, Nolen Scaife, Dave Tian, Logan Blue, Patrick Traynor, and Kevin R. B. Butler. 2016a. Sending out an SMS: Characterizing the security of the SMS ecosystem with public gateways. In *Proceedings of the IEEE Symposium on Security and Privacy (SP'16)*.

RedOxygen 2015. RedOxygen. Retrieved from http://www.redoxygen.com.

N. Saxena and N. S. Chaudhari. 2014. EasySMS: A protocol for end-to-end secure transmission of SMS. *IEEE Trans. Info. Forens. Secur.* 9, 7 (July 2014), 1157–1168. DOI: https://doi.org/10.1109/TIFS.2014.2320579

N. Saxena, H. Shen, N. Komninos, K. K. R. Choo, and N. S. Chaudhari. 2018. BVPSMS: A batch verification protocol for end-to-end secure SMS for mobile users. *IEEE Trans. Depend. Secure Comput.* (2018), 1. DOI: https://doi.org/10.1109/TDSC.2018.2799223

Bruce Schneier. 2005. Two-factor authentication: Too little, too late. *Commun. ACM* 48, 4 (Apr. 2005).

Scrapy 2015. Scrapy. Retrieved from http://scrapy.org.

SecurID 2015. RSA SecurID hardware tokens. *EMC Security*. Retrieved from http://www.emc.com/security/rsa-securid/rsa-securid-hardware-tokens.htm.

SiPix Imagining, Inc. 2006. World's first ISO compliant payment displaycard using SiPix and SmartDisplayer's flexible display panel. Retrieved from http://www.businesswire.com/portal/site/google/index.jsp?ndmViewId=news_view&newsId=20060510006193&newsLang=en.

Ann Skudlark. 2014. Characterizing SMS Spam in a Large Cellular Network via Mining Victim Spam Reports. Retrieved from http://web2-clone.research.att.com/export/sites/att_labs/techdocs/TD_101435.pdf.

SMS Forum. 2003. Short message peer to peer protocol specification 5.0. Retrieved from http://opensmpp.org/specs/smppv50.pdf.

Anne-Birte Stensgaard. 2006. Biometric breakthrough—Credit cards secured with fingerprint recognition made feasible. Retrieved from http://www.ameinfo.com/58236.html.

Brett Stone-Gross, Marco Cova, Lorenzo Cavallaro, Bob Gilbert, Martin Szydlowski, Richard Kemmerer, Christopher Kruegel, and Giovanni Vigna. 2009. Your botnet is my botnet: Analysis of a botnet takeover. In *Proceedings of the 16th ACM Conference on Computer and Communications Security (CCS'09)*. ACM, New York, NY, 635–647.

Henry Tan, Nazli Goharian, and Micah Sherr. 2012. $100,000 prize jackpot. call now!: Identifying the pertinent features of SMS spam. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, 1175–1176.

The Open University. 2014. Text messaging usage statistics. Retrieved from http://www.openuniversity.edu/news/news/2014-text-messaging-usage-statistics.

Kurt Thomas, Danny Huang, David Wang, Elie Bursztein, Chris Grier, Thomas J. Holt, Christopher Kruegel, Damon McCoy, Stefan Savage, and Giovanni Vigna. 2015. Framing dependencies introduced by underground commoditization. In *Proceedings of the 14th Annual Workshop on the Economics of Information Security*.

Kurt Thomas, Dmytro Iatskiv, Elie Bursztein, Tadek Pietraszek, Chris Grier, and Damon McCoy. 2014. Dialing back abuse on phone verified accounts. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. ACM, New York, NY, 465–476.

Kurt Thomas, Damon McCoy, Chris Grier, Alek Kolcz, and Vern Paxson. 2013. Trafficking fraudulent accounts: The role of the underground market in Twitter spam and abuse. In *Proceedings of the 2015 USENIX Security Symposium*. 195–210.

Anna Tims. 2015. "SIM swap" gives fraudsters access-all-areas via your mobile phone. *The Guardian* (Sept. 2015). Retrieved from http://www.theguardian.com/money/2015/sep/26/sim-swap-fraud-mobile-phone-vodafone-customer.

Patrick Traynor. 2012. Characterizing the security implications of third-party EAS over cellular text messaging services. *IEEE Trans. Mobile Comput.* 11, 6 (2012), 983–994.

Patrick Traynor, William Enck, Patrick McDaniel, and Thomas La Porta. 2008. Exploiting open functionality in SMS-capable cellular networks. *J. Comput. Secur.* 16, 6 (2008), 713–742.

Patrick Traynor, William Enck, Patrick McDaniel, and Thomas La Porta. 2009. Mitigating attacks on open functionality in SMS-capable cellular networks. *IEEE/ACM Trans. Netw.* 17, 1 (2009), 40–53.

Patrick Traynor, Patrick McDaniel, and Thomas La Porta. 2007. On attack causality in internet-connected cellular networks. In *Proceedings of the USENIX Security Symposium (SECURITY'07)*.

Patrick Traynor, Patrick McDaniel, and Thomas La Porta. 2008. *Security for Telecommunications Networks*. No. 978-0-387-72441-6 in Advances in Information Security Series. Springer.

Twilio 2015. Twilio. Retrieved from http://www.twilio.com.

U.S. Office of Personnel Management. 2015. Cybersecurity incidents. Retrieved from https://www.opm.gov/cybersecurity/cybersecurity-incidents/.

VirusTotal. 2015. VirusTotal—Free online virus, malware and URL scanner. Retrieved from https://www.virustotal.com/.

Yajin Zhou and Xuxian Jiang. 2012. Dissecting android malware: Characterization and evolution. In *Proceedings of the IEEE Symposium on Security and Privacy (SP'12)*. 95 –109.

Received April 2017; revised June 2018; accepted August 2018