usenix
THE ADVANCED
COMPUTING SYSTEMS
ASSOCIATION

Greystar: Fast and Accurate Detection of SMS Spam Numbers in Large Cellular Networks using Grey Phone Space

Nan Jiang, *University of Minnesota;* Yu Jin and Ann Skudlark, *AT&T Labs;*
Zhi-Li Zhang, *University of Minnesota*

This paper is included in the Proceedings of the
22nd USENIX Security Symposium.
August 14–16, 2013 • Washington, D.C., USA
ISBN 978-1-931971-03-4

Greystar: Fast and Accurate Detection of SMS Spam Numbers in Large Cellular Networks using Grey Phone Space

Nan Jiang
University of Minnesota

Yu Jin
AT&T Labs

Ann Skudlark
AT&T Labs

Zhi-Li Zhang
University of Minnesota

Abstract

In this paper, we present the design of Greystar, an innovative defense system for combating the growing SMS spam traffic in cellular networks. By exploiting the fact that most SMS spammers select targets randomly from the finite phone number space, Greystar monitors phone numbers from the grey phone space (which are associated with data only devices like laptop data cards and machine-to-machine communication devices like electricity meters) and employs a novel statistical model to detect spam numbers based on their footprints on the grey phone space. Evaluation using five month SMS call detail records from a large US cellular carrier shows that Greystar can detect thousands of spam numbers each month with very few false alarms and 15% of the detected spam numbers have never been reported by spam recipients. Moreover, Greystar is much faster in detecting SMS spam than existing victim spam reports, reducing spam traffic by 75% during peak hours.

1 Introduction

The explosion of mobile devices in the past decade has brought with it an onslaught of unwanted SMS (Short Message Service) spam [1]. It has been reported that the number of spam messages in the US has risen 45% in 2011 to 4.5 billion messages [2]. In 2012, there were 350K variants of SMS spam messages accounted for globally [3] and more than 69% of the mobile users claimed to have received text spam [4]. The sheer volume of spam messages not only inflicts an annoying user experience, but also incur significant costs to both cellular service providers and customers alike. In contrast to email spam where the number of possible email addresses is unlimited - and therefore the spammer generally needs a seed list beforehand, SMS spammers can more easily reach victims by, e.g., simply enumerating all numbers from the *finite* phone number space. This, combined with wide adoption of mobile phones, makes SMS a medium of choice among spammers. Furthermore, the increasingly rich functionality provided by smart mobile devices also enables spammers to carry out more sophisticated attacks through both voice and data channels, for example, using SMS spam to entice users to visit certain websites for product advertisement or other illicit activities.

Because SMS spam inflicts financial loss to mobile users and adverse impact to cellular network performance, the objective of defense techniques is to restrict spam numbers quickly before they reach too many victims. To this end, instead of applying popular solutions in controlling email spam (e.g., filtering based on sending patterns), which can cause a high false alarm rate, cellular carriers often seek help from their customers to alert them of emerging spamming activities. More specifically, cellular carriers deploy reporting mechanism for spam victims to report received spam messages and then examine and restrict these reported spam numbers accordingly. Such spam detection techniques using victim spam reports are very accurate, thanks to the human intelligence added while submitting these reports. However, these methods can suffer from significant delay due to the low report rate and slow user responses, rendering them inefficient in controlling SMS spam.

To address the issues in existing solutions, in this paper, we carry out extensive analysis of SMS spamming activities using five months of SMS call detail records collected from a large cellular network in the US and the SMS spam messages reported from the spam recipients to that cellular carrier. We find that a majority of spammers choose targets randomly from a few area codes or the entire phone number space, and initiate spam traffic at high rates. To detect such aggressive random spammers, we advance a novel notion of *grey phone space*. Grey phone space comprises a collection of *grey phone numbers* (or grey numbers in short). Grey numbers are associated with two types of mobile devices: data only devices (e.g., many laptop data cards and data modems, etc.) and machine-to-machine (M2M) communication devices (e.g., utility meters and medical devices, etc.). These grey numbers usually do not participate actively in SMS communication as other mobile numbers do (e.g., those associated with smartphones), they thereby form a grey territory that legitimate mobile users rarely enter. In the mean time, the wide dispersion of grey numbers makes them hard to be evaded by spammers who choose targets randomly.

On top of grey phone space, we propose the design of *Greystar*. Greystar employs a novel statistical model to detect spam numbers based on their interactions with grey numbers and other non-grey phone numbers. We evaluate Greystar using five months of SMS call records. Experimental results indicate that Greystar is superior to the existing SMS spam detection algorithms, which rely heavily on victim spam reports, in terms of both accuracy and detection speed. In particular, Greystar detected over 34K spam numbers in five months while only generating two false positives. In addition, more than 15% of the detected spam numbers have never been reported by mobile users. Moreover, Greystar reacts fast to emerging spamming activities, with a median detection time of 1.2 hours after spamming activities occur. In 50% of the cases, Greystar is at least 1 day ahead of victim spam reports. The high accuracy and fast response time allow us to restrict more spam numbers soon after spamming activities emerge, and hence to reduce a majority of the spam messages in the network. We demonstrate through simulation on real network data that, after deploying Greystar, we can reduce 75% of the spam messages during peak hours. In this way, Greystar can greatly benefit the cellular carriers by alleviating the load from aggressive SMS spam messages on network resources as well as limiting their adverse impact on legitimate mobile users.

The remainder of this paper is organized as follows. We introduce the SMS architecture and the datasets used in our study in Section 2. We then motivate the design of Greystar in Section 3. In Section 4 we study the SMS activities of spammers and legitimate users. The definition of grey numbers is presented in Section 5. In Section 6, we explain in detail the design of Greystar. Evaluation results are presented in Section 7. Section 8 discusses the related work and Section 9 concludes the paper.

2 Background and Datasets

In this section, we briefly describe the cellular network focused in our study. We then introduce the datasets and our ground truth for identifying spam phone numbers.

2.1 SMS Architecture in UMTS

The cellular network under study utilizes primarily UMTS (Universal Mobile Telecommunication System), a popular 3G mobile communication technology adopted by many mobile carriers across the globe. Here we introduce the architecture for delivering SMS messages inside UMTS networks (for other aspects regarding UMTS networks, e.g., mobile data channels, see [5]). Fig. 1 depicts a schematic view of the architecture. When sending an SMS message, an end user equipment (UE_A) directly communicates with a cell tower (or node-B), which forwards the message to a Radio Network Controller (RNC). The RNC then delivers the message to a Mobile Switching Center (MSC) server, where the message enters the Signaling System 7 (SS7) network and is stored temporarily at a Short Message Service Center (SMSC). From the SMSC, the message will be routed to the serving MSC of the recipient (UE_B), then to the serving RNC and Node-B, and finally reaches UE_B. Similarly, messages originated from other carrier networks (e.g., from UE_C) will also traverse the SS7 network and bypass the serving MSC before arriving at UE_B[1].

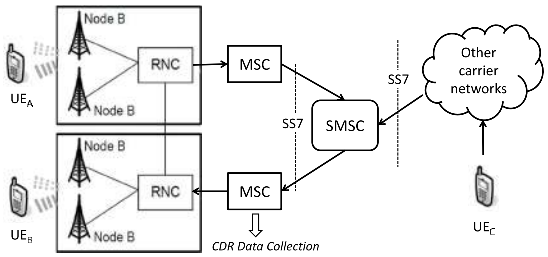

Figure 1: SMS architecture in UMTS networks.

2.2 Datasets

In this paper, we use two different datasets for our study.

SMS Call Detail Records (CDRs) are used for understanding SMS user/spammer activities and evaluating the performance of the proposed Greystar system. These records were collected at the serving MSC's of SMS recipients (see Fig. 1). This means that CDR records represent SMS messages targeting registered mobile customers of the UMTS network under study[2] and have been

[1] Note that similar SMS architecture is also adopted in other types of 3G/4G cellular networks. Additionally, in this paper, we only focus on SMS through the voice control channel. Short message services through mobile data channels, such as iMessage, Tweets and MMS, etc., are out of the scope of this paper (though defenses for fighting email spam can be applied to detect short message spam through data channels, which we shall discuss in Section 8).

[2] SMS messages targeting mobile users in other carrier networks and landline numbers are not seen at the serving MSCs and hence are not

successfully routed through the SS7 network. The CDR dataset spans 5 months from Jan 2012 to May 2012. Each record contains the SMS receiving time, the originating number, the terminating number and the International Mobile Equipment Identity (IMEI) for the device associated with the terminating number[3]. We note that CDR records do not contain text content of the original SMS messages.

Victim spam reports contain spam messages reported by spam recipients to the carrier. The said cellular carrier deploys an SMS spam reporting service for its users: when a user receives an SMS text and deems it as a spam message, s/he can forward the message to a *spam report number* designated by the cellular service provider. Once the spam is forwarded, an acknowledgment message is returned, which asks the user to reply with the spammer's phone number (referred to as the *spam number* hereafter). Once the above two-stage process is completed within a predefined time interval, a spam report is created, which includes the reporter's phone number, the spam number, the reporting time and the text content of the reported spam message. We employ six months of spam reports from Jan 2012 to June 2012 in order to cover spam numbers observed between Jan and May but are reported after May due to the delay of the spam reports (see Section 3.2).

We emphasize that no customer personal information was collected or used for our study. All customer identifies were *anonymized* before any analysis was conducted. In particular, for phone numbers, only the area code (i.e., the first 3 digits of the 10 digit North American numbers) was used and the remaining digits were hashed. Similarly we only retain the first 8-digit Type Allocation Code (TAC) of the IMEI to identify device types and anonymize the remaining 8-digit to preserve customers' privacy. In addition, to adhere to the confidentiality under which we have access to the data, in places we only present normalized views of our results while retaining the scientifically relevant magnitudes.

2.3 Obtaining Ground Truth

Although victim spam reports provide us with ground truth for some spam numbers, they are by no means comprehensive and can be noisy (see Section 3.2). Therefore, in this paper, we employ a more reliable source of ground truth. In particular, we request the fraud agents from the said UMTS carrier to manually verify spam number candidates detected by us. These fraud agents are exposed to much richer (and more expensive) sources of information. For example, fraud agents can investigate the ownership and the price plan information of the candidates, examine their SMS sending patterns and correlate them with known spam numbers in terms of their network locations and active times, etc. The final decision is made conservatively by corroborating different evidence.

Admittedly, fraud agents can make mistakes during their investigation. Meanwhile, their breadth may be limited by not being able to inspect all mobile numbers in the network. Nevertheless, fraud agents provide us with the most authoritative ground truth available for our study. It is worth mentioning that such investigation by fraud agents has been deployed independently for SMS spam number detection and restriction for more than one year and no false alarm has yet been observed (e.g., no user complaint is observed so far regarding incorrectly restricted phone numbers). Therefore, in our study, we will treat fraud agents as a black box authority, i.e., we submit a list of spam number candidates to fraud agents and they return a list of confirmed spam numbers.

3 Objectives and Existing Solutions

In this section, we discuss the objectives of developing an effective defense against SMS spam by comparing the difference between SMS spam and traditional email spam. We then review the most widely adopted SMS spam detection method based on crowdsourcing victim spam reports and point out its inefficacy. In the end, we present the rationale of the proposed Greystar system.

3.1 SMS Spam Defense Objectives

In a conventional SMS spamming scenario, an SMS spammer (note that we refer to an SMS *spammer* as the person who employs a set of spam numbers to launch SMS spam campaigns) first invests in a set of phone numbers and special high-speed devices, such as 3G modems and SIM boxes [6]. Using these devices, s/he then initiates unsolicited SMS messages to a large number of mobile phone numbers. Akin to traditional email spam, the objective of SMS spam is to advertise certain information to entice further actions from the message recipients, e.g., calling a fraud number or clicking on a URL link embedded in the message which points to a malicious site. However, SMS spamming activities exhibit unique characteristics which shift the focus of the defense mechanisms and hence render inapplicable or

included in CDR records.

[3] IMEI's are stored at MSC's and are updated every time users connect to the network. Although we have observed that spammers sometimes modify the IMEIs of their spamming devices (e.g., through special equipment like SIM boxes), IMEI spoofing among legitimate users is rare. Therefore we can reliably identify the types of user devices based on their corresponding IMEIs. Meanwhile, since all the CDRs are collected at MSCs, we can identify the original phone numbers that initiate the SMS messages. Hence our approach is not affected even when spammers employ spoofing techniques to change their caller IDs.

inefficient existing solutions for defending against traditional email spam.

Email service providers usually detect and filter email spam at their mail servers, to which they have full access. There they can build accurate spam filters by exploiting rich features in emails including the text content. Spam filters at end user devices are also a common choice, where email clients (apps) filter spam while retrieving emails from remote mail servers. Though blacklist of email spammers are sometimes used to assist spam classification [7–9], restricting email spam senders is usually not the main focus of the defense, since it requires close collaboration between email providers and network carriers. Moreover, it is observed that many spam emails are originated from legitimate hosts due to botnet activities [10], which makes restricting spam originators an inapplicable solution.

In comparison to emails which are generally stored on servers and wait for users to retrieve them, SMS messages are delivered instantly to the recipients through the SS7 network. Along the path, SMS messages are only cached temporarily at SMSC (only when the recipients are offline), leaving little time for cellular carriers to react to them. The task becomes even more challenging especially when the SMS traffic volume peaks during busy hours. Filtering SMS spam at end user devices (e.g., using mobile apps) is also not applicable given many SMS capable devices (e.g., feature phones) do not support running such apps. In addition, for a user with a pay-per-use SMS plan, she is already charged for the spam message once it arrives at her device. More importantly, even when SMS spam filters are deployed at SMSC's and end user devices, SMS spammers can still inflict significant loss to the carrier and other mobile users. This is because the huge number of spam messages can lead to a significant increase in the SMS traffic volume at the cell towers serving the spam senders, possibly causing congestion and hence deteriorating voice/data usage experience of nearby users. For example, we have found the SMS traffic volume at cell towers can easily get multiplied by more than 10 times due to the activities of spammers. Therefore, the focus of the SMS spam defense is to *control spam numbers as soon as possible before they reach a large number of victims*.

An efficient SMS spam detection algorithm is hence expected to react quickly to emerging spamming activities. Meanwhile, the focus on restricting spam numbers places a strong emphasis on the accuracy of the algorithm. First, it requires a spam detection algorithm to limit false alarms, because false alarms can lead to incorrect restriction of legitimate users from accessing SMS services. Second, it demands the algorithm detect as many spam numbers as possible so as to minimize the impact of SMS spam activities on the network. Such high accuracy requirements are hard to achieve solely based on the SMS sending patterns of the spammers. For example, it is difficult to separate spam campaigns from legitimate SMS campaigns, such as a school sending messages to its students to alert adverse weather conditions. These legitimate senders can exhibit characteristics that are common to SMS spammers[4]. Spammers may also alter their sending patterns to mimic legitimate users to avoid detection. As a result, cellular carriers often seek the assistance from their customers to alert them of emerging SMS spam activities.

3.2 Spam Detection by Crowdsourcing Victim Spam Reports

The emphasis on high accuracy gives rise to the wide adoption of spam detection methods based on victim spam reports which were introduced in Section 2. Victim spam reports represent a more reliable and cleaner source of SMS spam samples, as all the spam messages contained in the reports have been vetted and classified by mobile users (using human intelligence). To further mitigate the possible errors caused during the two-step reporting process, cellular carriers often crowdsource spam reports from different users. For example, a simple yet effective strategy is to identify a spam number after receiving reports from K distinct users. Meanwhile, defense mechanisms based on victim spam reports are also of low cost, because only numbers reported by users need to be further analyzed. Due to this reason, spam reports are usually a trigger for more sophisticated investigation on the senders, such as their sending patterns, service plans, etc..

Despite the high accuracy and low cost, detecting SMS spam based on spam reports is analogous to performing spam filtering at user devices. The major drawback is detection delay, which we illustrate in Fig. 2 based on the CDR data from January 2012. The red solid curve in Fig. 2 measures how long it takes for a spam number to be reported after spam starts (a.k.a. *report delay*). We consider a spam number starts spamming when it first reaches at least 50 victims in an hour (see Section 4 for discussion on spamming rates). From Fig. 2, we observe that less than 3% of the spam numbers are reported within 1 hour after spamming starts. More than 50% of the spam numbers are reported 1 day after. The report delay is mainly due to the extremely low report rate from users. In fact, less than 1 in 10,000 spam messages were

[4] Maintaining a whitelist of such legitimate intensive SMS users can be challenging. First, we have little information to identify the white list if the users are outside the network. Second, even for the users inside the network, the whitelist can still be dynamic, with new businesses/organizations initiating/stopping SMS broadcasting services every day. More importantly, users are not obliged to report to the carrier when they intend to start such services.

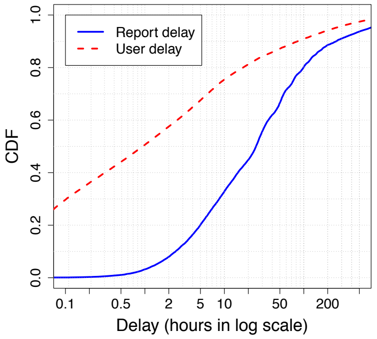

Figure 2: Lags of user reports.

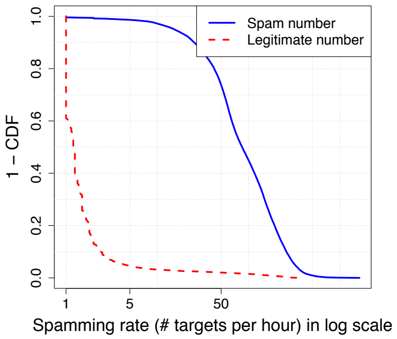

Figure 3: Spamming rate.

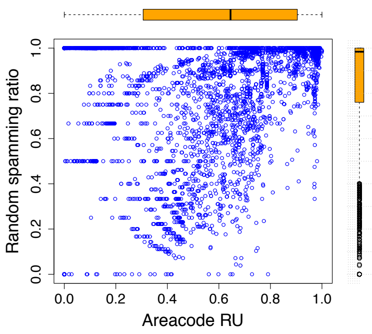

Figure 4: Target selection strategies.

reported during the five month observation period. Aside from causing a long detection delay, the low report rate also leads to many missed detections (see Section 7).

In addition, even when a victim reports a spam message, how long it takes him/her is at the reporter's discretion. The blue dotted curve in Fig. 2 shows how fast a user reports a spam message after receiving it (*user delay*). Note that each user can receive multiple spam messages (possibly with different text content) from the same sender and hence can report the same sender multiple times. Thus, we define *user delay* as the time difference between when a user reports a spam message and the *last* time that the user receives spam from that particular spam number before the report. We observe in Fig. 2 , among users who report spam, half of the spam messages are reported more than 1 hour after they receive the spam messages. Around 20% spam are reported even after a day. Due to such a long delay, spammers have already inflicted significant loss to the network and its customers.

In addition to the problem of detection delay, the current two-stage reporting method is error-prone. We find around 10% reporters fail to provide a valid spam number at the second stage. Moreover, spam report based methods are vulnerable to attacks, as attackers can easily game with the detection system by sending bogus reports to Denial-of-Service (DoS) legitimate numbers. All these drawbacks render spam detection using victim spam reports an insufficient solution.

3.3 Overview of Greystar

Recognizing the drawbacks of existing victim report based solutions, we introduce the rationale behind Greystar. The objective of Greystar is to accurately detect SMS spam while at the same time being able to control spam numbers as soon as possible before they reach too many victims. To this end, we advance a novel notion of grey phone numbers. These grey numbers usually do not communicate with other mobile numbers using SMS, they thereby form a grey territory that legitimate mobile users rarely enter. On the other hand, as we shall see in Section 4, it is difficult for spammers to avoid touching these grey numbers due to the random target selection strategies that they usually adopt. Greystar then passively monitors the footprints of SMS senders on these grey numbers to detect impending spam activities targeting a large number of mobile users.

Greystar addresses the problems in existing spam report based solutions as follows. First, the population of grey numbers is much larger and widely distributed (see Section 5), providing us with more “spam alerts” to capture more spam numbers more quickly. Second, by passively monitoring SMS communication with grey numbers, we avoid the user delay and errors introduced when submitting spam reports. Last, Greystar detects spammers based on their interactions with grey phone space. This prevents malicious users from gaming the Greystar detection system and launching DoS attacks against other legitimate users.

In the following, we first discuss related work in Section 8. We then study the difference of spamming and legitimate SMS activities in Section 4, which lays the foundation of the Greystar system. In Section 5 we introduce our methodology for identifying grey numbers. We then present the design of Greystar in Section 6 and evaluate it in Section 7.

4 Analyzing SMS Activities of Spammers and Legitimate Users

We first formally define SMS spamming activities. During a spamming process, a spammer selects (following a certain strategy) a sequence of *target phone numbers*, $X := \{x_1, x_2, \cdots, x_i, \cdots\}$ $(1 \leq i \leq n)$, to send SMS messages to over a time window T. Each target phone number is a concatenation of two components, the 3-digit area code x_i^a, which is location specific, and the 7-digit subscriber number x_i^s. Note that we only examine US phone numbers (which have 10 digits excluding the leading country code “1”). Phone numbers of SMS senders from other countries which follow the same North American Numbering Plan (NANP) are removed before the study. All the statistics in this section are calculated based on a whole month data from January 2012. To compare the activities of spam numbers and legitimate numbers, we obtain an equal amount of samples from both groups. In particular, the spam numbers are identified from victim spam reports and the legitimate numbers are randomly sampled from the remaining SMS senders appearing in the month-long CDR data set. Both samples of phone numbers are checked by fraud agents before the analysis to remove false positives and false negatives.

4.1 SMS Sending Rates

We first compare the SMS sending rates of known spam numbers and legitimate numbers.We measure the sending rate at the granularity of hours, i.e., the average number of unique recipients a phone number communicates with hourly.The CCDF curves of the sending rates are shown in Fig. 3.

From Fig. 3, spam numbers have a much higher SMS sending rate than legitimate numbers. This is not surprising given the purpose of spamming is to reach as many victims as possible within a short time period. In particular, more than 95% of spam numbers have a sending rate above 5 and more than 70% spam numbers exhibit a sending rate above 50. In contrast, more than 97% of the legtimate numbers have a sending rate below 5. As we can see in Section 6, by enforcing a threshold on the sending rate, we can filter out most of the legitimate numbers without missing many spam numbers.

Due to their high spamming rates, at the node-Bs that spam numbers are connected to, we find that the sheer volume of spamming traffic is astonishing. Spamming traffic can exceed normal SMS traffic by more than 10 times. Even at RNCs, which serve multiple node-Bs, traffic from spamming can account for 80% to 90% of total SMS traffic at times. Such a high traffic volume from spammers can exert excessive loads on the network, affecting legitimate SMS traffic. Furthermore, since SMS messages are carried over the voice control channel, excessive SMS traffic can deplete the network resource, and thus can potentially cause dropped calls and other network performance degradation. Meanwhile, the increasing malware app instances that propagate through the SMS channel also emphasize the importance of restricting SMS spam activities in cellular networks.

We note that, although most legitimate numbers send SMS at low rates (e.g., below 50), due to the large population size of the legitimate numbers, there are still many of them with high sending rates indistinguishable from those of spam numbers. Investigation shows that they belong to organizations which use the SMS service to disseminate information to their stakeholders, e.g., churches, schools, restaurants, etc. How to distinguish these legitimate intensive SMS senders from SMS spammers is the main focus of our Greystar system.

4.2 Spammer Target Selection Strategy

We next study how spammers select spamming targets. We characterize their target selection strategies at two levels, i.e., how spammers choose area codes and how they select phone numbers within each area code.

We define the metric *area code relative uncertainty* (ru_a) to measure whether a spammer favors phone numbers within certain area codes. The ru_a is defined as:

$$ru_a(X) := \frac{H(X^a)}{H_{max}(X^a)} = \frac{-\sum_{q \in Q} P(q) \log P(q)}{log|Q|},$$

where $P(q)$ represents the proportion of target phone numbers with the same area code q and $|Q|$ is the total number of area codes in the US. Intuitively, a large ru_a (e.g., greater than 0.7) indicates that the spammer uniformly chooses targets across all the area codes. In contrast, a small ru_a means the targets of the spammer come from only a few area codes.

We next define a metric *random spamming ratio* to measure how spammers select targets within each area code. Let P^a be the proportion of active phone numbers[5] within area code a. For a particular spamming target sequence X^a of a spam number, if the spammer randomly chooses targets, the proportion of active phone numbers in X^a should be close to P^a. Otherwise, we believe the spammer has some prior knowledge (e.g., with an obtained target list) to select specific phone numbers to spam. Based on this idea, we carry out a one sided Binomial hypothesis test for each spammer and each area code to see if the corresponding target selection strategy is random within that area code. The random spamming ratio is then defined as the proportion of area codes within which a spammer selects targets randomly (i.e., the test fails to reject the randomness hypothesis with P-value=0.05). Note that, for each spam number, only area codes with more than 100 victims are tested to ensure the validity of the test.

[5] The active phone numbers are identified as all registered phone numbers inside the carrier’s billing database who have unexpired service plans. We find that the active numbers are uniform across all area codes, possibly due to frequent phone number recycling within carrier networks (e.g., phone numbers originally used by landlines are reassigned to mobile phones) and users switching between cellular carriers while retaining the same phone numbers.

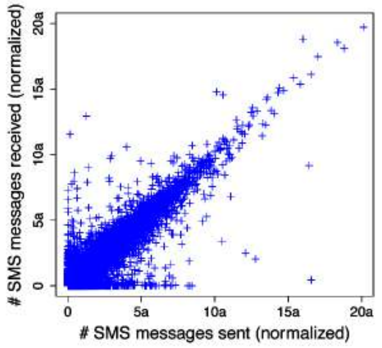

Figure 5: SMS sent vs. received.

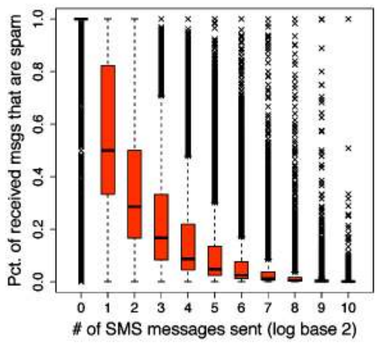

Figure 6: Activeness vs. spam prop.

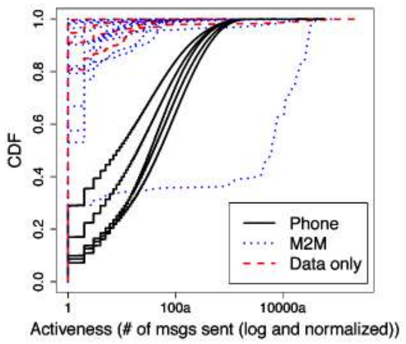

Figure 7: Device activeness (log).

Fig. 4 plots the ru_a (the x-axis) and the random spamming ratio (the y-axis) for individual spam numbers. For ease of visualization, we illustrate the marginal densities along both axes using boxplots. Based on the marginal density of ru_a, we find that many spam numbers (more than 60%, with $ru_a < 0.7$) concentrate on phone numbers within a few area codes. We find that spammers tend to focus on area codes with more users, i.e., those corresponding to large cities and metro areas, e.g., New York City, Chicago, Los Angeles, etc. In comparison, the remaining 40% of spam numbers select targets across many area codes or even the entire phone number space.

Meanwhile, based on the y-axis of Fig. 4, we find that, no matter how spam numbers choose area codes, a predominant portion of them select targets randomly within each area code. We refer to these spammers as *random spammers* hereafter. This also explains why spammers favor large metro areas, because they are likely to reach more active mobile users by randomly selecting phone numbers within these area codes.

In summary, due to the finite phone space, spammers can simply enumerate phone numbers to send spam messages. Compared to having a target phone number list before spamming, this random target selection strategy is effective and of low cost, and hence has been adopted by most SMS spammers. Due to their predominance, in this paper, we focus on detecting these random spammers. Meanwhile, the spammers who utilize non-random target selection strategies (e.g., the points at the bottom of Fig. 4) will be discussed in Section 7.3.

4.3 Mobile User SMS Activities

Since many SMS spammers adopt random target selection strategies, mobile users (within the same area code) have the same exposure to spam. In other words, given a fixed (long enough) observation period, these mobile users are expected to receive an equal amount of spam messages. In this section, we study the SMS activities of legitimate mobile users and demonstrate that certain users can be used for detecting spam activities.

We first obtain a general understanding on the volume of SMS activities from legitimate mobile users in the network. Fig. 5 shows the number of messages sent (x-axis) and received (y-axis) by each user over a month[6]. We observe that a majority of users send and receive a similar amount of SMS messages and thereby form an approximate diagonal line. However, there are mobile users who deviate from such a pattern noticeably. For example, the points close to the x-axis represent users who send far more SMS messages than the ones they receive. These users consist of senders who own a large subscriber base, e.g., cellular providers, university emergency contact lines, political campaign lines, etc. In contrast, we observe quite a few points that reside near the y-axis. Investigation shows that they are phone numbers which receive periodic updates (e.g., electricity readings) from machine-to-machine (M2M) devices through SMS messages (see Section 5.2 for discussion of M2M devices).

Fig. 5 implies the different magnitude that mobile users engaged in SMS communication. To quantify the intensity of SMS activities from mobile users, we define (SMS) *activeness* as the number of messages sent from a mobile user during the observation period. Intuitively, for users who are less active, the spam messages tend to account for a more dominant proportion of their overall SMS communication. We illustrate this point in Fig. 6, where we bin all users based on their activeness (x-axis, in log scale), and calculate the distribution of the proportion of spam messages out of all SMS messages received by each user within each bin. Note that spam messages are identified as the SMS messages originated from spam numbers contained in victim spam reports. From Fig. 6, we observe an upward shift of spam message proportions as the activeness decreases. Interestingly, we find quite a

[6] We note that the constants used for normalization (denoted as a and b) vary across individual figures.

few numbers which have sent no more than 1 SMS message during the one month period. For a majority of these numbers, all the messages they have received are spam (as indicated by the fact that most probability mass is squeezed to a small region close to 1). This implies that these SMS inactive numbers are good indicators of spamming activities, i.e., SMS senders who communicate with them are more likely to be spammers.

5 SMS Grey Phone Number Space

In order to utilize these SMS inactive numbers for spam detection, we want to first answer the following questions. Why do these numbers have a low volume of SMS activity? Is there an inexpensive way to identify a stable set of such numbers for building the detection system? To answer these questions, we carry out an in-depth analysis of SMS inactive users. We then define grey phone space and propose a method for identifying the grey phone space using CDR records. In the end, we study properties of grey phone space and show the potential of using it to detect spamming activities.

5.1 Investigating Service Plans

Cellular carriers often provide their customers with a rich set of features to build their personal service plans. Users are free to choose the best combination of features to balance their needs and the cost. For example, a frequent voice caller often opts in an unlimited voice plan and a user who watches online videos a lot can choose a data plan with a larger data cap. Therefore, service plans encode demographic properties of the associated users. We hence study the correlations between different service plan features and SMS activeness to understand these SMS inactive users.

More specifically, we extract all the service plans associated with the legitimate user samples, which include features related to voice, data and SMS services. We calculate the Pearson correlation coefficients of the SMS activeness and individual plan features (treated as binary variables). The features are then ranked according to the correlation values. We summarize the top 5 features that are positively and negatively correlated with SMS activeness in Table 1.

Top 5 negatively correlated	Top 5 positively correlated
Text restricted	Monthly unlimited voice/text
Voice restricted	Messaging unlimited
Text msg pay per use	Rollover family plan
Voice/data prepaid	Unlimited SMS/MMS
Large cap data plans	Small cap data plans

Table 1: Corr. of activeness and plan features.

The top 5 features with negative correlations are in the first column of Table 1. Many of these SMS inactive users are enrolled in the pay-per-use SMS plan, a common economical choice for users who rarely access SMS services. Interestingly, a large number of SMS inactive users have restrictions on their voice/text plans and have been simultaneously enrolled in large cap data plans. Such restrictions only apply for mobile users with data only devices, such as tablets and laptop data cards, etc. In contrast, the top 5 features with positive correlations are summarized in the second column. Most of SMS active users have unlimited SMS plans, a favorable choice of frequent SMS communicators. Many of them have also enrolled in small cap data plans and unlimited MMS plans, which are dedicated for smartphone users.

Though service plans demonstrate clear distinctions between SMS inactive and active users, relying on service plans to identify SMS inactive users is not effective in practice due to two reasons. First, service plans change frequently, especially when users upgrade their devices. Second, query service plan information persistently during run time can be very expensive. Fortunately, our analysis above also reveals that service plans are strongly correlated with the device types, e.g., data only device users are less active compared to smartphone users. Can we use device types as a proxy to identify SMS inactive users instead? We shall explore such possibilities in the following section.

SMS towards data only devices. Like phones, laptops and other data only devices are also equipped with SIM cards and hence, once connected to the network, are able to receive SMS messages. We therefore can capture CDR records to these devices at MSCs. However, manufacturers often restrict text usage on these devices by masking the APIs related to SMS functions. Meanwhile, at the billing stage, text messages to these data only devices (with a text restricted plan) are not charged by the carrier. There are exceptions such as laptops enrolled in regular text messaging plans, however, such cases are rare based on our observations.

5.2 Identifying Grey Phone Space

The device associated with each phone number can be found in the CDR data based on the first eight-digit TAC of the IMEI. We use the most updated TAC to device mapping from the UMTS carrier in January 2013 and have identified 27 mobile device types (defined by the carrier) which we summarize in Table 2. We note that finer grained analysis at individual device level is also feasible. However, we find that, except for the vehicle tracking devices which we shall see soon, devices within each category have strong similarity in their SMS activeness distributions. Hence we gain little by defining grey

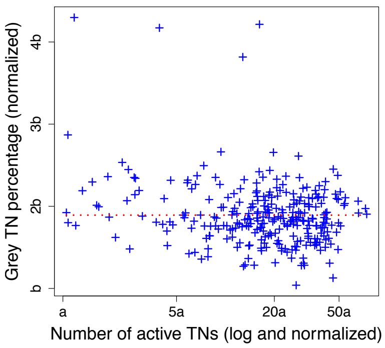

Figure 8: Grey number distribution.

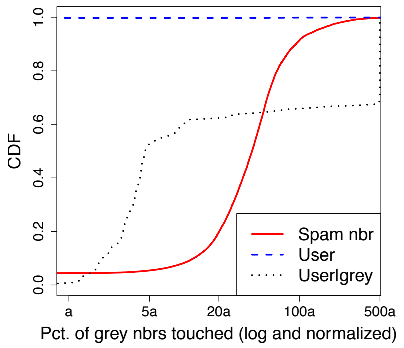

Figure 9: Grey ratio.

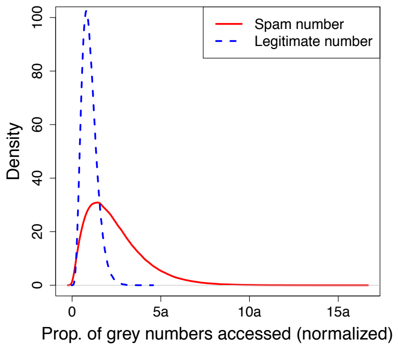

Figure 10: Distr. of θ and θ^*.

numbers at the device level.

Type	Examples
Data-only	Laptop data cards, tablets, netbooks, eReaders, 3G data modems, etc.
M2M	Security alarms, telematics, vehicle tracking devices, point-of-sale terminals, medical devices, etc.
Phone	Smartphones, feature phones, quick messaging phones, PDAs, etc.

Table 2: Device categories and examples.

Fig. 7 shows the CDF distributions of SMS activeness of phone numbers associated with different device types. We observe three clusters of CDF curves. The first one consists of curves concentrating at the top-left corner, representing devices with very low SMS activeness. This cluster covers all data only devices and a majority of machine-to-machine devices (see [11] for more discussions of M2M devices). The second cluster lies in the middle of the plot, which includes all phone devices. The third cluster contains only one M2M device type, which covers all vehicle tracking devices. Interestingly, the curve of such devices shows a bi-modal shape, where some devices communicate frequently using SMS while other devices mainly stay inactive. Based on Fig. 7, we define grey numbers as the ones that are associated with devices in the first cluster, i.e., data only devices and M2M devices excluding the vehicle tracking device category. The collection of all grey number are referred to as the grey phone space. The grey numbers are representatives of a subset of SMS inactive users[7]. Meanwhile, the grey phone space defined in this way is stable because it is tied to mobile devices instead of specific phone numbers, whose behaviors can change over time (e.g., when a user upgrades the device). Furthermore, grey numbers can be identified directly based on the IMEIs in the CDR data with little cost, as opposed to querying and maintaining service plan information for individual users.

5.3 Characterizing Grey Phone Space

We next study the distribution of grey numbers and show how grey phone space can help us detect spamming activities.

Fig. 8 shows the size of each area code in the phone space (the x-axis, in terms of the number of active phone numbers) and the proportion of grey phone numbers out of all active phone numbers in that area code (the y-axis). The correlation coefficient of two dimensions is close to 0, indicating that grey numbers exist in both densely and sparsely populated areas. The wide distribution of grey numbers ensures a better chance of detecting spam numbers equipped with random spamming strategies. To illustrate this point, we calculate the proportion of grey numbers out of all the numbers accessed by spam numbers (red solid curve) and legitimate users (blue dotted curve). We observe that a predominant portion of legitimate users never touch grey phone space. In fact, less than 1% of the users have ever accessed grey numbers in the 1 month observation period. In addition, we show the same distribution for legitimate users (who have sent to at least 50 recipients in a month) conditioned on having touched at least one grey number. Compared to the spam numbers which tend to access more grey numbers (red solid curve), these legitimate users communicate with much fewer grey numbers. In most cases, the access of grey numbers is triggered by users replying to spam numbers who usually use M2M devices to launch spam.

[7] We use devices in the first cluster as our definitions of grey space, however, as we have seen in Fig. 7, even within the grey number categories there are still (a very few) numbers that are highly active in SMS communication. The proposed beta-binomial classification model (discussed in detail in Section 6) will take into account this fact. Intuitively, the model detects a spam number only when it is observed to have significant interaction with the grey space. Given a majority of the grey numbers that are SMS inactive, the chance that a phone number is misclassified as a spam number due to its interaction with these outliers in the grey space is very small.

5.4 Discussion: Greyspace vs. Darkspace

In addition to the grey phone space, the "dark" phone space (i.e., formed by unassigned phone numbers) can also be a choice for detecting spam activities using the same technique proposed in this paper. Analogous concepts of grey IP addresses and dark IP addresses for detecting anomalous activities have been explored in [12,13]. However, unlike IP addresses which are often assigned to organizations in blocks (i.e., sharing the same IP prefix), the phone number space is shared by different cellular service providers, landline service providers and even (IP) TV providers. Even if some phone numbers are assigned in blocks initially to a certain provider, the frequent phone number assignment changes caused by new user subscription, old user termination, recycling of phone numbers and phone number porting in/out between different providers will ultimately result in the shared ownership of the phone number space as we have seen today. For example, different cellular and landline providers can have phone numbers under the same legitimate area code. It is difficult to tell which phone number belongs to which provider without inquiring the right provider.

This poses significant challenges when we want to identify dark (unassigned) phone numbers. As dark phone numbers can be anywhere in the phone number space (within legitimate area codes) and can belong to any provider, it is rather difficult to determine a dark number, at least from the perspective of a single provider. For instance, just because a phone number is not assigned to any user/device belonging to a particular provider, it does not necessarily mean that such a number is dark. In other words, accurate detection of dark numbers requires the collaboration of all the owners of the phone number space, which is an intractable task. Meanwhile, such dark number repository needs to be updated frequently to reflect the changes of phone number assignments.

In comparison, grey numbers can be defined easily with respect to a particular provider: these are phone numbers assigned to devices belonging to customers of that provider where there are usually less SMS activities originated from these numbers (devices). Meanwhile, whether a number is grey is readily available to us (based on the existing the IMEI numbers inside CDR records) without any extra work.

6 System Design

In this section, we first present an overview of Greystar. We then introduce the detection model and how we choose parameters for the model.

6.1 System Overview

The logic of Greystar is illustrated in Alg. 1, which runs periodically at a predefined frequency. In our experiment, we run Greystar hourly. Greystar employs a time window of W (e.g., W equals 24 hours in our studies). The footprint of each SMS originating number s, e.g., the sets of grey and non-grey numbers accessed by s (denoted as G_s and N_s, respectively), are identified from the CDR data within W. After that, a filtering process is conducted which asserts two requirements on originating numbers to be classified, i.e., in the past 24 hours: i) the sender is active enough (which has sent messages to no less than $M = 50$ recipients. Recall the high sending rates of known spam numbers in Fig. 3); and ii) the sender has touched at least one grey number. These two criteria, especially the second one, can help significantly reduce the candidates to be classified in the follow-up step. In fact, we find that, on average, less than 0.1% of users send SMS to grey numbers in each day. More importantly, these users cover a majority of active SMS spammers in the network as we shall see in Section 7. As a consequence, this filtering step can noticeably reduce the system load as well as potential false alarms.

Algorithm 1 Greystar algorithm.

```
Input: CDR records $D$ from the past $W = 24$ hours, $M$=50;
Output: Spam number candidates $C$;
From $D$, extract all SMS senders $Orig$;
for each $s \in Orig$ do
    Extract the CDR records associated with $s$: $D_s \subset D$;
    From $D_s$, identify the grey numbers $G_s$ and non-grey numbers $N_s$ accessed by $s$;
    if $|G_s| + |N_s| \geq M$ and $|G_s| > 0$ then
        if $detect\_spamnbr(G_s, N_s)$=1 then
            $C := C \cup \{s\}$;
        end if
    end if
end for
```

Once a sender passes the filtering process, the function *detect_spamnbr* is called to classify the sender into either a spam number or a legitimate number based on G_s and N_s associated with that sender. In this paper, we propose a novel Beta-Binomial model for building the classifier, which we explain in detail next.

6.2 Classifier Design

We assume a random SMS spammer selects spamming targets following a two-step process. First, the spammer chooses a specific target phone number block. Second, the spammer uniformly chooses target phone numbers from that block. Let θ denote the density of grey numbers in the target block and $X := \{x_i\}, 1 \leq i \leq n$ be the sequence of target phone numbers selected. Meanwhile, let k be the number of grey numbers in X. The target selection process can then be formulated as the following generative process.

1. Choose a target block with grey number density θ;
2. Choose $x_i \sim Bernoulli(\theta), 1 \leq i \leq n$;

We note that θ varies as a spammer chooses different phone number blocks. The choice of phone number blocks is arbitrary. For example, A spammer can choose a large phone block across multiple area codes or a small one consisting of only a fraction of phone numbers within one area code. Therefore, θ itself can be considered as a random variable. We assume θ follows a Beta distribution[8], i.e., $\theta \sim Beta(\alpha, \beta)$, with a probability density function as:

$$P(\theta|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha}(1-\theta)^{\beta},$$

where Γ is the gamma function. Therefore, the random variable k follows a Beta-Binomial distribution:

$$P(k|n,\alpha,\beta) = \binom{n}{k}\frac{\Gamma(k+\alpha)\Gamma(n-k+\beta)}{\Gamma(n+\alpha+\beta)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

The target selection process of legitimate users can be expressed using the same process. Because legitimate users tend to communicate less with grey numbers, their corresponding θ^*'s are usually much smaller. Let α^* and β^* be the parametrization of the Beta distribution associated with θ^*. For a phone number that has accessed n targets, out of which k are grey numbers, we classify it as a spam number (i.e., *detect_spamnbr* returns 1) if

$$\frac{P(spammer|k,n)}{P(legitimate|k,n)} = \frac{P(k|n,\alpha,\beta)P(spammer)}{P(k|n,\alpha^*,\beta^*)P(legitimate)} > 1,$$

where the first equation is derived using the Bayes theorem. It is equivalent to

$$\frac{P(k|n,\alpha,\beta)}{P(k|n,\alpha^*,\beta^*)} > \frac{P(legitimate)}{P(spammer)} = \eta$$

In practice, it is usually unclear how many spammers are in the network, therefore, to estimate η directly is challenging. We instead choose η through experiments.

6.3 Parameter Selection

There are five parameters to be estimated in the classifier, $\hat{\alpha}, \hat{\beta}, \hat{\alpha}^*, \hat{\beta}^*$ and η. We use the data from January 2012 to determine these parameters. To obtain ground truth, we submit to the fraud agents a list of all the SMS senders that i) have sent to more than 50 recipients in a 24 hour time window; and ii) at least one of the recipients is grey (recall the filtering criteria in Algorithm 1). Fraud agents carry out investigation on these numbers for us and label spam numbers in the list. We then divide the January data into two subsets, the first two weeks of data for fitting the Beta-binomial models (i.e., to determine the first four parameters) and the rest of data is reserved for testing the classifier to estimate η.

In particular, using the training data set, we estimate the parameters for two Beta-binomial models using maximum likelihood estimation. With the estimated parameters, we illustrate the probability density function $\theta \sim Beta(\alpha, \beta)$ and $\theta^* \sim Beta(\hat{\alpha^*}, \hat{\beta^*})$ in Fig. 10. The density functions agree with our previous observations in Fig. 9. The mass of the probability function corresponding to the legitimate users concentrates on a narrow region close to 0, implying that legitimate users communicate much less with grey numbers than non-grey numbers. In contrast, the density associated with spam numbers widely spreads out, indicating more grey numbers are touched by spam numbers due to their random target selection strategies.

We evaluate the accuracy of the classifier given different choices of η on the test data set and the Receiver Operating Characteristic (ROC) curve is displayed in Fig. 11. The x-axis represents the false alarm rate (or the false positive rate) and the y-axis stands for the true detection rate (or the true positive rate). From Fig. 11, with a certain η, Greystar can detect more than 85% spam numbers without producing any false alarm. We will choose this η value in the rest of our experiments[9].

7 Greystar Evaluation

In this section, we conduct an extensive evaluation of Greystar using five months of CDR data and compare it with the methods based on victim spam reports in terms of accuracy, detection delay and the effectiveness in reducing spam traffic in the network.

[8] In Bayesian inference, the Beta distribution is the conjugate prior probability distribution for the Bernoulli and binomial distributions. Instead of using the Bernoulli model, we can model the second stage of the target selection process as sampling from a multinomial distribution corresponding to different device types. In this case, the conjugate prior distribution of the multinomial parameters is the Dirichlet distribution. However, our preliminary experiments show little performance gain from applying the more sophisticated model in comparison to the increased computation cost.

[9] Note that the exact parameter values used in Greystar are proprietary and we are not able to release them in the paper. We have also tested the choice of η using different partitioning of the training/test data. The η remains stable across experiments.

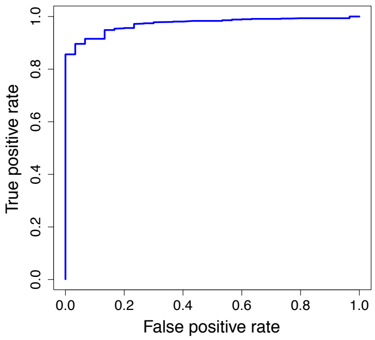

Figure 11: ROC curve (false positive rate vs. true positive rate.

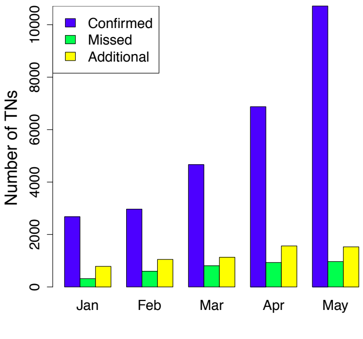

Figure 12: Accuracy evaluation (in comparison to victim spam reports).

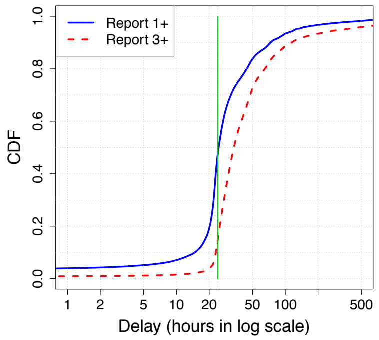

Figure 13: Detection speed compared to spam report based methods.

7.1 Accuracy Evaluation

To estimate the accuracy and the false alarm rate, we again consult with the fraud agents to check the numbers from Greystar detection results. False negatives (or missed detections), on the other hand, are more difficult to identify. Given the huge number of negative examples classified, we are unable to have all of them examined by the fraud agents to identify all missed detections because of the high manual investigation cost. As an alternative solution, we compare Greystar detection results with victim spam reports to obtain a lower bound estimate of the missed detections.

More formally, let S_g denote the detection results from Greystar and S_c be the spam numbers contained in the victim spam reports received during the same time period. We define *missed detections* of Greystar as $S_c - S_g$. In addition, we define *additional detections* of Greystar as $S_g - S_c$ to measure the value brought by Greystar to the existing spam defense solution. The monthly accuracy evaluation results are displayed in Fig. 12.

The blue bars in Fig. 12 illustrate the spam numbers validated by fraud agents in each month. Greystar is able to detect thousands of spam numbers per month. The ascending trend of detected spam numbers coincides with the increase of victim spam reports in the five-month observation window. This implies that Greystar is able to keep up with the increase of spam activities. In addition to the large number of true detections, Greystar is highly accurate given only two potential false alarms are identified by fraud agents in 5 months. Interestingly, these two numbers are associated with tenured smartphone users who suddenly behave abnormally and initiate SMS messages to many recipients whom they have never communicated with in the past. We suspect these users have been infected by SMS spamming malware that launch spam campaigns from the users' devices without their consent. To identify SMS spamming malware and hence removing such false alarms will be our future work.

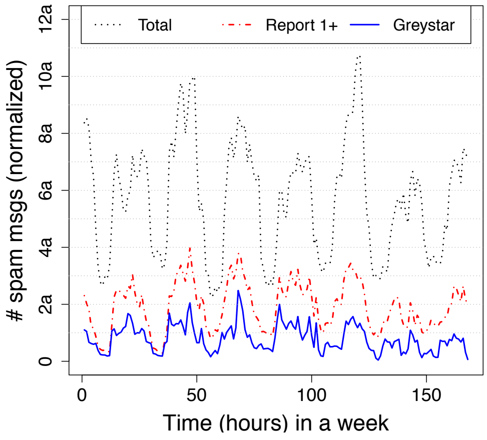

Figure 14: Number of spam messages after restriction.

In comparison to the victim spam reports, Greystar detects over 1000 addition spam numbers that were not reported by spam victims while missing less than 500 monthly. Meanwhile, although a majority of the spam numbers detected by Greystar are also reported by spam victims, Greystar can detect these numbers much faster than methods based on victim reports, and consequently can suppress more spam messages in the network. We illustrate this point in the next section.

7.2 Detection Speed and Benefits to Cellular Carriers

We note that, to reduce noise, cellular carriers often rely on multiple spam reports (e.g., K reports) from different victims to confirm a spam number. We refer to such a crowdsourcing method as the $K+$ algorithm. To evaluate the speed of Greystar, we compare it with two versions of the $K+$ algorithms, namely, $1+$ and $3+$. Comparing with

$1+$ supplies us with the lower bound of the time difference and comparison with $3+$ illustrates the real benefit brought by Greystar to practical spam defense solutions. More specifically, we measure how many hours Greystar detects a spam number ahead of $1+$ and $3+$, respectively. Fig. 13 shows the CDF curves of the comparison results, where we highlight the location on the x-axis corresponding to 24 hours with a green vertical line. We observe that Greystar is much faster than $K+$ algorithms. For example, Greystar is one day ahead of $1+$ in 50% of the cases and is one day before $3+$ in more than 90% of the times.

We find that, on average, it takes less than 1.2 hours for Greystar to detect a spam number after it starts spamming (i.e., starts sending messages to more than 50 victims in an hour). The fast response time of Greystar is accredited to the much larger population of grey numbers, from which Greystar can gather evidence to detect more spam numbers more quickly. In addition, collecting evidence passively from grey numbers eliminates the delay during the human reporting process (recall Fig. 2). Therefore, Greystar is characterized with a much faster detection speed than the $K+$ algorithm. Such a gain in the detection speed can lead to more successful reduction of spam traffic in the network. We illustrate this point next.

For simplicity, we assume a spam number can be instantly restricted after being detected. We run simulation on a one week dataset (the first week of January 2012) and calculate the number of spam messages appearing in each hour assuming a particular spam detection algorithm is deployed exclusively in the network. The results are illustrated in Fig. 14. The total spam messages are contributed by known spam numbers observed in that week. We observe that Greystar can successfully suppress the majority of spam messages. During peak hours when the total number of spam messages exceeds 600K, only around 150K remains after Greystar is deployed. In other words, Greystar leads to an overall reduction of 75% of spam messages during peak hours. In comparison, $1+$ only guarantees a spam reduction of 50% due to long detection delay. We note that, due to the noise in the spam reports, cellular providers often employ $K+$ ($K \geq 3$) instead of $1+$ to avoid false alarms. In this case, the benefit from Greystar is even more substantial.

7.3 Analysis of Missed Detections

In this section, we investigate the missed detections (false negatives) from Greystar, i.e., the spam number candidates that were not detected by Greystar but have been reported by spam victims. There are around 500 such numbers in each month and totally around 27K missed detections. We note that we focus only on a subset of the candidates who are customers of the cellular network under study, for whom we have access to a much richer set of information sources to carry out the investigation. We believe the conclusions from analyzing this subset of candidates also apply for other candidates outside the network.

We classify these candidates into three groups based on the volume of the associated CDR records.

No volume. We do not observe any CDR record for 19.5% of the numbers. We inquiry the SMS billing records for these numbers and find that many of them initiate a vast amount of SMS traffic to foreign countries, such as Canada and Jamaica, etc., and hence no CDR record has been collected to trigger Greystar detection.

Low volume. We find around 27% of the missed detections have accessed less than 50 recipients during the observation period. We study the text content inside the victim spam reports to understand the root cause of these missed detections. The most popular text content are party advertisements and promotions from local restaurants. Users are likely to have registered with these merchants in the past and hence received ads from them. For the rest of the numbers, we find many send out spam messages to advertise mobile apps and premium SMS services. From the users' comments posted on online forums and social media sites [14, 15], we find two of the advertised apps are messenger/dating apps which have issues with their default personal settings. Without manual correction, these apps, once initiated, will send out friend requests to a few random users of the apps. Spam messages from the remaining numbers are also likely to be sent out without users' consent, especially the ones that broadcast premium SMS services. We suspect they are caused by apps abusing permissions or even behaviors of malware apps. For example, one app advertised by spam is reported to contain malware that sent SMS text to the contact list on the infected device, where the text contains a URL for downloading that malware.

High volume. The rest of the phone numbers send SMS to a large number of recipients. From the reported spam text, we find 7.1% of them belong to legitimate advertiser who broadcast to registerred customers and are somehow reported by the recipients. For the rest of numbers, we find their spam topics are quite different from those of the detected ones. In particular, 11% of these numbers are associated with adult sites or hotlines, in comparison to only 0.06% among the detected numbers. Meanwhile, 17.6% of them advertise local shopping deals, as opposed to only 2.1% among the detected ones. Such difference suggests that these spam victims somehow gave out their phone numbers to spammers, e.g., while visiting malicious sites to register services or to purchase products. In addition, we extract the voice call history associated with these high volume candidates. Interestingly, we find that about 4% of these numbers have initiated phone calls to many terminating numbers in the past. We suspect that these spammers employ auto-dialers to harvest active phone numbers (i.e., the ones that have answered the calls) from the phone number space. With the list of active phone numbers, spammers can send spam more effectively and avoid detection in the mean time.

Admittedly, there are spam numbers in these three categories that are missed by Greystar because they are equipped with a target number list obtained through auto-dialing or social engineering techniques (for example, accurate target lists can potentially be obtained by applying techniques discussed in [16]). SMS traffic from these users is not differentiable from that of the legitimate users. However, we emphasize that these missed detections only account for less than 9% of all the spam numbers detected and they will not have a significant impact on the efficacy of Greystar for reducing the overall spam traffic. In fact, we find that, on average, the missed detections sent 37% less spam messages in comparison to the spam numbers detected by Greystar. On the other hand, we do see the needs of combining Greystar and other methods to build a more robust defense solution. For example, many malicious activities can be better detected by correlating different channels (e.g., voice, SMS and data). Meanwhile, cellular carriers can collaborate with mobile marketplace to detect and control suspicious apps that can potentially initiate spam.

8 Related Work

The demographic features and network behaviors of SMS spammers were analyzed in [6]. [16] investigated the security impact of SMS messages and discussed the potential of denying voice service by sending SMS to large and accurate phone hitlists at a high rate. Meanwhile, [16] also discussed several ways of harvesting active phone numbers, which can potentially be employed by SMS spammers to generate accurate target number lists to launch spam campaign more efficiently and to evade detection. Similar short message services carried by the data channel were also studied. For example, [17] characterized spam campaigns from "wall" messages between Facebook users. [18–21] analyzed Twitter spam. [22,23] studied talkback spam on weblogs. Meanwhile, akin to SMS spammers, the behaviors of email spammers were characterized in [24–27]. In comparison, we not only study the strategies of SMS spammers but also propose an effective spam detection solution based on our analysis.

In addition to the victim spam reports mentioned earlier, network behaviors of spammers, e.g., sending patterns, have been used in SMS spam detection, such as [28]. Similar network statistics based methods designed for email spam detection can also be applied for identifying SMS spam, such as [29–32]. However, these methods often suffer from large false positive rates, because many legitimate customers can exhibit SMS sending patterns similar to those of spammers. In contrast, Greystar utilizes a novel concept of grey phone space to detect spam numbers, which yields an extremely low false alarm rate.

Some systems have been developed in the form of smartphone apps to classify spam messages on user mobile devices [33–35]. However, not all mobile devices support executing such apps. Furthermore, from a user's perspective, this method is a late defense as the spam message has already arrived on his/her device and the user may already be charged for the spam message. Moreover, the high volume of spam messages that have already traversed the cellular network may have resulted in congestion and other adverse network performance impacts. Greystar is deployed inside the carrier network and hence do not have these drawbacks. As we have seen in Section 7, Greystar can quickly detect spam numbers once they start spamming and hence significantly reduce spam traffic volume in the network.

Similar to our work, many works have leveraged unwanted traffic for anomaly detection, such as Internet dark space [13,36], grey space [12], honeynet [37,38] and failed DNS traffic [39], etc. We are the first to advance the notion of grey phone space and propose a novel statistical method for identifying SMS spam using grey phone space.

9 Conclusion and Future Work

In this paper, we presented the design of Greystar, an innovative system for fast and accurate detection of SMS spam numbers. Greystar monitors a set of grey phone numbers, which signify impending spam activities targeting a large number of mobile users, and employs an advanced statistical model for detecting spam numbers according to their interactions with grey phone numbers. Using five months of SMS call detail records collected from a large cellular network in the US, we conducted extensive evaluation of Greystar in terms of the detection accuracy and speed, and demonstrated the great potential of Greystar for reducing SMS spam traffic in the network.

Our future work will focus on applying Greystar to detect other suspicious activities in cellular networks, such as telemarketing campaigns. Meanwhile, we will correlate Greystar detection results with cellular data traffic to detect malware engaged in such spamming activities.

Acknowledgments

The work was supported in part by the NSF grants CNS-1017647 and CNS-1117536, the DTRA grant HDTRA1-09-1-0050. We thank Peter Coulter, Cheri Kerstetter and Colin Goodall for their useful discussions and constructive comments. Finally, we thank our shepherd, Patrick Traynor, for his many suggestions on improving the paper.

References

[1] Federal communications commission. Spam: unwanted text messages and email, 2012. http://www.fcc.gov/guides/spam-unwanted-text-messages-and-email.

[2] Mobile spam texts hit 4.5 billion. http://www.businessweek.com/news/2012-04-30/mobile-spam-texts-hit-4-dot-5-billion-raising-consumer-ire.

[3] C. Baldwin. 350,000 different types of spam sms messages were targeted at mobile users in 2012, 2013. http://www.computerweekly.com/news/2240178681/350000-different-types-of-spam-SMS-messages-were-targeted-at-mobile-users-in-2012.

[4] 69% of mobile phone users get text spam, 2012. http://abcnews.go.com/blogs/technology/2012/08/69-of-mobile-phone-users-get-text-spam/.

[5] Y. Jin, N. Duffield, A. Gerber, P. Haffner, W.-L. Hsu, G. Jacobson, S. Sen, S. Venkataraman, and Z.-L. Zhang. Making sense of customer tickets in cellular networks. In *Proc. of the 30th IEEE International Conference on Computer Communications*, 2011.

[6] I. Murynets and R. Jover. Crime scene investigation: Sms spam data analysis. In *Proc. of the 12th ACM Internet Measurement Conference*, 2012.

[7] S. Sinha, M. Bailey, and F. Jahanian. Improving SPAM blacklisting through dynamic thresholding and speculative aggregation. In *Proc. of the 17th Annual Network and Distributed System Security Symposium*, 2010.

[8] J. Jung and E. Sit. An empirical study of spam traffic and the use of DNS black lists. In *Proc. of the 4th ACM Internet Measurement Conference*, 2004.

[9] A. Ramachandran, N. Feamster, and D. Dagon. Revealing botnet membership using dnsbl counter-intelligence. In *Proc. of the 2nd Workshop on Steps to Reducing Unwanted Traffic on the Internet*, 2006.

[10] Y. Xie, F. Yu, K. Achan, R. Panigrahy, G. Hulten, and I. Osipkov. Spamming botnets: signatures and characteristics. In *Proc. of the 2008 ACM SIGCOMM Annual Conference*, 2008.

[11] M. Shafiq, L. Ji, A. Liu, J. Pang, and J. Wang. A first look at cellular machine-to-machine traffic: large scale measurement and characterization. In *Proc. of the 2012 ACM International Conference on Measurement and Modeling of Computer Systems*, 2012.

[12] Y. Jin, G. Simon, K. Xu, Z.-L. Zhang, and V. Kumar. Gray's anatomy: dissecting scanning activities using ip gray space analysis. In *Proc. of the 2nd Workshop on Tackling Computer Systems Problems with Machine Learning Techniques*, 2007.

[13] R. Pang, V. Yegneswaran, P. Barford, V. Paxson, and L. Peterson. Characteristics of internet background radiation. In *Proc. of the 4th ACM Internet measurement conference*, 2004.

[14] Sms watchdog. http://www.smswatchdog.com.

[15] 800notes - Directory of unknown callers. http://www.800notes.com.

[16] W. Enck, P. Traynor, P. McDaniel, and T. La Porta. Exploiting open functionality in sms-capable cellular networks. In *Proc. of the 12th ACM Conference on Computer and Communications Security*, 2005.

[17] H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Zhao. Detecting and characterizing social spam campaigns. In *Proc. of the 10th ACM Internet Measurement Conference*, 2010.

[18] S. Ghosh, B. Viswanath, F. Kooti, N. Sharma, G. Korlam, F. Benevenuto, N. Ganguly, and K. Gummadi. Understanding and combating link farming in the twitter social network. In *Proc. of the 21st International World Wide Web Conference*, 2012.

[19] K. Thomas, C. Grier, V. Paxson, and D. Song. Suspended accounts in retrospect: an analysis of Twitter spam. In *Proc. of the 11th ACM Internet Measurement Conference*, 2011.

[20] C. Yang, R. Harkreader, J. Zhang, S. Shin, and G. Gu. Analyzing spammers' social networks for fun and profit: a case study of cyber criminal ecosystem on twitter. In *Proc. of the 21st International World Wide Web Conference*, 2012.

[21] C. Grier, K. Thomas, V. Paxson, and M. Zhang. @spam: the underground on 140 characters or less. In *Proc. of the 17th ACM Conference on Computer and Communications Security*, 2010.

[22] E. Bursztein, P. Lam, and J. Mitchell. Trackback spam abuse and prevention. In *Proc. of the 2009 ACM workshop on Cloud computing security*, 2009.

[23] E. Bursztein, B. Gourdin, and J. Mitchell. Reclaiming the blogosphere talkback a secure linkback protocol for weblogs. In *Proc. of the 16th European Symposium on Research in Computer Security*, 2011.

[24] C. Kreibich, C. Kanich, K. Levchenko, B. Enright, G. Voelker, V. Paxson, and S. Savage. On the spam campaign trail. In *Proc. of the 1st USENIX Workshop on Large-Scale Exploits and Emergent Threats*, 2008.

[25] C. Kreibich, C. Kanich, K. Levchenko, B. Enright, G. Voelker, V. Paxson, and S. Savage. Spamcraft: An inside look at spam campaign orchestration. In *Proc. of the 2nd USENIX Workshop on Large-Scale Exploits and Emergent Threats*, 2009.

[26] C. Kanich, C. Kreibich, K. Levchenko, B. Enright, G. Voelker, V. Paxson, and S. Savage. Spamalytics: An empirical analysis of spam marketing conversion. *Communications of the ACM*, 52(9):99–107, 2009.

[27] A. Pathak, F. Qian, C. Hu, M. Mao, and S. Ranjan. Botnet spam campaigns can be long lasting: evidence, implications, and analysis. In *Proc. of the 2009 ACM International Conference on Measurement and Modeling of Computer Systems*, 2009.

[28] Q. Xu, E. Xiang, Q. Yang, J. Du, and J. Zhong. Sms spam detection using noncontent features. *Intelligent Systems, IEEE*, 27(6):44 –51, 2012.

[29] T. Ouyang, S. Ray, M. Rabinovich, and M. Allman. Can network characteristics detect spam effectively in a stand-alone enterprise? In *Proc. of the 12th Passive and Active Measurement conference*, 2011.

[30] M. Sirivianos, K. Kim, and X. Yang. Introducing social trust to collaborative spam mitigation. In *Proc. of the 30th IEEE International Conference on Computer Communications*, 2011.

[31] S. Hao, N. Syed, N. Feamster, A. Gray, and S. Krasser. Detecting spammers with snare: spatio-temporal network-level automatic reputation engine. In *Proc. of the 18th USENIX Security Symposium*, 2009.

[32] A. Pitsillidis, K. Levchenko, C. Kreibich, C. Kanich, G.M. Voelker, V. Paxson, N. Weaver, and S. Savage. Botnet judo: Fighting spam with itself. In *Proc. of the 17th Annual Network and Distributed System Security Symposium*, 2010.

[33] K. Yadav, P. Kumaraguru, A. Goyal, A. Gupta, and V. Naik. Smsassassin: crowdsourcing driven mobile-based system for sms spam filtering. In *Proc. of the 12th Workshop on Mobile Computing Systems and Applications*, 2011.

[34] G. Cormack, J. Hidalgo, and E. Sánz. Feature engineering for mobile (sms) spam filtering. In *Proc. of the 30th international ACM SIGIR conference*, 2007.

[35] H. Tan, N. Goharian, and M. Sherr. $100,000 Prize Jackpot. Call now! Identifying the pertinent features of SMS spam. In *Proc. of the 35th Annual ACM SIGIR Conference*, 2012.

[36] E. Wustrow, M. Karir, M. Bailey, F. Jahanian, and G. Huston. Internet background radiation revisited. In *Proc. of the 10th ACM Internet measurement conference*, 2010.

[37] The honeynet project, 2012. http://project.honeynet.org/.

[38] G. Dunlap, S. King, S. Cinar, M. Basrai, and P. Chen. Revirt: enabling intrusion analysis through virtual-machine logging and replay. In *Proc. of the 2nd USENIX Symposium on Operating Systems Design and Implementation*, 2002.

[39] N. Jiang, J. Cao, Y. Jin, L. Li, and Z.-L. Zhang. Identifying suspicious activities through dns failure graph analysis. In *Proc. of the 8th IEEE International Conference on Network Protocols*, 2010.