Understanding SMS Spam in a Large Cellular Network: Characteristics, Strategies and Defenses

Nan Jiang[1], Yu Jin[2], Ann Skudlark[2], and Zhi-Li Zhang[1]

[1] University of Minnesota, Minneapolis, MN, {njiang,zhzhang}@cs.umn.edu
[2] AT&T Labs, Florham Park, NJ, {yjin,aes}@research.att.com

Abstract. In this paper, using a year (June 2011 to May 2012) of user reported SMS spam messages together with SMS network records collected from a large US based cellular carrier, we carry out a comprehensive study of SMS spamming. Our analysis shows various characteristics of SMS spamming activities, such as spamming rates, victim selection strategies and spatial clustering of spam numbers. Our analysis also reveals that spam numbers with similar content exhibit strong similarity in terms of their sending patterns, tenure, devices and geolocations. Using the insights we have learned from our analysis, we propose several novel spam defense solutions. For example, we devise a novel algorithm for detecting related spam numbers. The algorithm incorporates user spam reports and identifies additional (unreported) spam number candidates which exhibit similar sending patterns at the same network location of the reported spam number during the nearby time period. The algorithm yields a high accuracy of 99.4% on real network data. Moreover, 72% of these spam numbers are detected at least 10 hours before user reports.

1 Introduction

The past decade has witnessed an onslaught of unsolicited SMS (Short Message Service) spam [1] in cellular networks. The volume of SMS spam has risen 45% in the US in 2011 to 4.5 billion messages and, in 2012, more than 69% of the mobile users claimed to have received text spam [2]. In addition to bringing an annoying user experience, these SMS spam often entice users to visit certain (fraud) websites for other illicit activities, e.g., to steal personal information or to spread malware apps, which can inflict financial loss to the users. At the same time, the huge amount of spam messages also concerns the cellular carriers as the messages traverse through the network, causing congestion and hence degraded network performance.

Although akin to traditional email spam, SMS spam exhibit unique characteristics which render inapplicable classical email spam filtering methods. Unlike emails which are generally stored on servers and wait for users to retrieve them, SMS messages are delivered instantly to the recipients through the Signaling System 7 (SS7) network, leaving little time for cellular carriers to react to spam. Meanwhile, high operation cost also limits applying sophisticated spam filters which rely on inspecting SMS message content.

Filtering SMS spam at end user devices (e.g., using mobile apps) is also not a feasible solution given many SMS capable devices (e.g., feature phones) do not support

running such apps. In addition, a user (e.g., with a pay-per-use SMS plan) is already charged for the spam message once it arrives at her device. More importantly, the sheer volume of SMS spam, once entering the network, can significantly increase the traffic load and potentially deteriorate voice/data usage experience of other nearby mobile users. Due to these reasons, the focus of the SMS spam defense is to *detect and control phone numbers involved in initiating spam (i.e., spam numbers) quickly before they reach a large number of victims.*

Network behavioral statistics (e.g., sending patterns) have been applied for detecting spam numbers (e.g., [3–7]). However, many of these methods suffer from an unacceptable large false alarm rate, because many legitimate numbers who own a large subscriber base can exhibit similar SMS sending behaviors as those of spam numbers, e.g., cellular providers, university emergency contact lines, political campaign lines, etc. Due to this reason, many cellular network carriers have adopted and deployed a more accurate albeit conservative SMS spam reporting mechanism for mobile users, whereby after receiving a spam message, a victim can report it via a text message forward. Mobile carriers can then investigate and confirm these reported activities and restrict the SMS activities of the offending spam numbers. The user spam report based method produces much fewer false alarms, thanks to the human intelligence added while submitting these reports. However, as we shall see in Section 8, it suffers from significant delay due to the low report rate and slow user responses, rendering them less efficient in controlling spam.

Despite the drawbacks associated with user spam reports, they do provide us a unique information source for identifying spam numbers and studying their behaviors in order to build better spam defenses. Taking advantage of this SMS spam reporting mechanism, in this paper we collect spam messages reported to one of the largest cellular carriers in the US from May 2011 to June 2012 – which contains approximately 543K spam messages – and carry out an extensive analysis of spamming activities using these user reported spam messages together with their associated SMS network records. Our objectives are three-fold: 1) to characterize the spamming activities in today's large cellular networks; 2) to infer the intent and strategies of spammers; and 3) to develop effective spam detection methods based on lessons learned from our analysis.

To achieve these goals, we first identify more than 78K spam numbers from user-submitted SMS spam reports (referred to as user spam reports hereafter) and conduct an in-depth analysis of spamming activities associated with these numbers. We observe strong differences in behaviors between spammers and non-spammers in terms of their voice, data and SMS usage. We find that the tenure of the spam numbers to be less than one week old, and programmable devices are often used to deliver spam messages at various spam sending rates. More importantly, we find that most spammers select targets randomly, either from a few area codes or the entire phone number space. This is plausibly due to the *finite* phone number space which enables spammers to reach victims by simply enumerating their numbers. Meanwhile, we find spammers tend to concentrate at and select targets from densely populated geolocations (e.g., large metro areas), where they have access to more resources (e.g., high speed networks and spamming devices) and can reach live users more easily. As a consequence, at these locations, the huge volume of spam traffic can lead to more than a 20 times increase of SMS traffic

at some Node-Bs, and more than 10 times at some RNCs. The sheer volume of spam traffic can potentially have an adverse impact on the experience of normal users in these areas.

In addition to analyzing spamming behaviors of individual spam numbers, we carry out a multi-dimensional analysis of the correlations of spam numbers. More specifically, we apply a text mining tool, CLUTO [8, 9], to cluster spam numbers into various clusters based on similarity of spam content they generate. Our investigation shows strong similarity among the spam numbers contained in each cluster: for instance, the devices associated with these spam numbers are frequently of identical types, the spam numbers used are often purchased at nearly the same time; furthermore, the call records of these numbers also exhibit strong temporal and spatial correlations, namely, they occur at a particular location and close in time. All the evidence suggests that the spam numbers contained in the same cluster are likely employed by a single spammer to engage in the same SMS spam campaign, e.g., at a particular location using multiple devices such as laptops or 3G/4G cellular modems.

Based on the characteristics of spam numbers found in our analysis, we pinpoint the inefficacy of existing spam defenses based solely on user spam reports due to the associated low report rate and long delay. In addition to proposing solutions to enhance the existing user spam report mechanism, we innovative several spam defenses that rely less on user spam reports or do not require users' participation at all. For example, leveraging the strong temporal/spatial correlations among spam numbers employed by the same spammer, we propose a novel *related spam number* detection algorithm. The algorithm consists of two components. First, it maintains a watchlist of all potential spam numbers detected based on the SMS sending patterns of individual phone numbers. Second, upon receiving a user spam report, it identifies additional (unreported) spam number candidates which exhibit similar sending patterns at the same network location during the same or nearby time period. Evaluated on a month long dataset, the algorithm identifies 5.1K spam numbers with an extremely high accuracy of 99.4%, where more than 72% and 40% of the detection results are 10 hours and 1 day before the user reports, respectively. Moreover, 9% of the detected spam numbers have never been reported by users possibly due to the extremely low report rate. As another example, taking advantage of the random spamming strategies favored by most of the spammers, we propose to deploy honeypot phone numbers in the phone number space to trap spam messages and to detect spam numbers without the help of user spam reports.

The remainder of this paper is organized as follows. We briefly introduce the datasets in Section 2, and discuss related work in Section 3. In Section 4 we analyze user spam reports and extract spam numbers, which we use to study the characteristics of SMS spammers in Section 5 and their network behaviors in Section 6. In Section 7, we cluster spam numbers based on the spam content and further investigate correlations of spam numbers contained in each cluster. Analysis of existing solutions and proposal of new spam defenses are presented in Section 8. Section 9 concludes the paper.

2 Background and Datasets

In this section, we briefly introduce the SMS architecture of the cellular network under study. We then describe the datasets collected from this network for our analysis.

2.1 SMS Architecture in Large Cellular Networks

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. The (high-level) architecture for delivering (text-based) SMS messages[3] inside a UMTS network is depicted in Fig. 1. 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 reach UE_B. The return message will follow a reverse path from UE_B to UE_A.

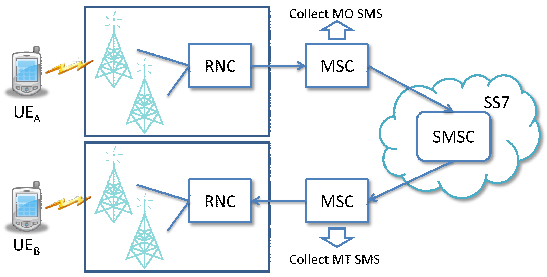


Fig. 1. SMS architecture in UMTS networks.

2.2 User Spam Report Dataset

The said cellular service provider 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*[4] hereafter). Once the above two-stage process is completed within a predefined time interval, a

[3] Note that we focus on studying text-based SMS messages, which are sent through the control (signaling) channel as opposed to messaging services which deliver content through data channels, like iMessage and Multimedia Message Service (MMS).

[4] We use the term "spam numbers" here to differentiate from spammers, where the latter term refers to the human beings who are in control of these phone numbers that initiate SMS spam.

spam record is created. The dataset used in our study contains spam messages reported by users over a one-year period (from June 2011 to May 2012). The dataset contains approximately 543K complete spam records and all the spam numbers reported are inside the said UMTS network (i.e., for whom we have access to complete service plan information and can hence observe all the SMS network records originated from these numbers). Each spam record consists of four features: the spam number, the reporter's phone number, the spam forwarding time and the spam text content.

2.3 SMS Spam Call Detail Records

To assist our analysis of spamming activities from multiple dimensions, we also utilize the SMS (network) records – SMS Call Detail Records (referred to as CDRs hereafter) – associated with the reported spam numbers over the same one year time period. These CDRs are collected at MSCs primarily for billing purposes: depending on the specific vantage point where call records are collected, there are two types of SMS CDRs (see Fig. 1): whenever an SMS message sent by a user reaches the SS7 network, a Mobile Originating (MO) CDR is generated at the MSC serving the sender (even when the terminating number is inactive); once the recipient is successfully paged and the message is delivered, a Mobile Terminating (MT) CDR is generated at the MSC serving the recipient. We note that unlike the user-generated SMS spam reports, these SMS CDRs do *not* contain the text content of the original SMS messages. Instead, they contain only limited network related information such as the SMS sending time, the sender's and receiver's phone numbers, the serving cell tower and the device International Mobile Equipment Identity (IMEI) number for the sender (in MO CDRs) or the receiver (in MT CDRs). Using SMS spam numbers identified from spam reports, we extract all CDRs associated with these spam numbers during the same one-year period, and use them to study the network characteristics of spam numbers and hence to infer the intents and strategies of the spammers. Recall that all the focused spam numbers are inside the cellular network under study, we only utilize MO CDRs for our studies, which cover the complete spamming history of each spam number.

We would like to emphasize that no customer personal information was collected or used in our study, and all customer identities were *anonymized* before any analysis was carried out. In particular, for phone numbers, only the area code (i.e., the first 3 digits of the 10 digit North American numbers) was kept; the remaining digits were hashed. Similarly, we only retained the first 8-digit Type Allocation Code (TAC) of the IMEIs in order to identify device types and hashed the remaining 8 digits. 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.

3 Related Work

In a related study [10], the authors characterized the demographic features and network behaviors of individual SMS spam numbers. Though we also conduct network-level

It will be shown later in this paper, spammers often employ multiple spam numbers for an SMS spam campaign. In contrast, a non-spammer (e.g., an airline notification service) typically uses only a single phone number when "broadcasting" an SMS notification to many recipients.

analysis of SMS spam, our purpose is to infer the intents and strategies of SMS spammers, and to identify and explain the correlation among different spam numbers.

In addition to the user spam reports mentioned earlier, network behaviors of spammers, e.g., sending patterns, have been used in SMS spam detection, such as [3]. Similar network statistics based methods designed for email spam detection were also applied for identifying SMS spam, such as [4–7]. Content-based SMS spam filters using machine learning techniques were also proposed in [11, 12]. However, the application of these methods is limited due to either the unacceptable false alarm rate associated or the large computation overhead on the end user devices. Based on the analysis of SMS spam in this paper, we propose several novel spam detection approaches for accurate and fast detection of SMS spam numbers.

As online social media sites become popular, many studies focus on understanding spam activities on these sites. For example, [13] quantified and characterized spam campaigns from "wall" messages between Facebook users. [14] studied link farming by spammers on Twitter. [15] analyzed the inner social relationships of spammers on Twitter. [16] characterized spam on Twitter. Though such IP-based short message spam are out of the scope of this paper, they often exhibit characteristics similar to SMS spam. Hence the proposed solutions are also applicable for detecting IP-based spam.

4 Analyzing User Spam Reports

In this section, we study the user reported spam messages. We first describe the data preprocessing step and explain how to extract spam numbers from these messages. We then illustrate statistics derived from the spam text content.

4.1 Data Preprocessing

Human users, unfortunately, may introduce noise and/or biases in the rather cumbersome SMS spam reporting process. For instance, a user may mistype a spam number in the second step, leave it blank, or simply enter an arbitrary alphanumeric string, say, xxxxxx, due to lack of patience. In addition, users may apply differing criteria in deciding what is considered as spam. To address these issues, we take a rather *conservative* approach and employ several preprocessing mechanisms to filter out the noise and potential biases introduced by human users during the reporting process.

To remove noise, we first filter out all spam reports that do not contain legitimate and valid 10-digit phone numbers[5]. In addition, we use the SMS CDRs to cross-validate the remaining spam numbers, i.e., we remove those that either have no corresponding SMS

[5] In fact, 12.2% of the user spam reports contain (valid) so-called *short code* numbers with fewer than 10 digits. The short codes are generally used as gateways between mobile networks and other (computer) networks and services. For instance, they are used for computer users (e.g., via Google voice or Yahoo messenger service) to send SMS messages to other mobile users, or for mobile users to send tweets to Twitter, or to vote for American Idol (in latter two cases, the messages are received by computers for further processing). Since this paper focuses on SMS spam sent/received by mobile users, we remove these short code related reports from further consideration, leaving analysis of them as our future work.

CDRs (within a week window of the user reporting). This filtering process removes roughly 15.6% of the spam reports from further consideration.

To address the potential biases introduced by users in reporting spam, we match the spam messages in the spam reports against a set of regular expressions defined by anti-fraud/anti-abuse human agents of the cellular carrier (e.g., "*.*you have won a XXX $1,000 giftcard.**"). These regular expressions are generated by these agents over time in a conservative manner based on manual inspection of spam reports and other user complaints, with the aim to restrict the offending spam numbers from further abuse. Hence these regular expressions have been tracked over years to ensure no false positives (the agents are notified of false alarms when legitimate customers call the customer care to complain about their SMS services being restricted). We obtain 384K spam reports after removing all reports that do not match any of the regular expressions.

4.2 Spam Number Extraction and Spam Report Volume

During a one year observation period, a phone number can be deactivated, e.g., abandoned by users or shut down by cellular providers, and can be recycled after a predefined time period. In other words, a phone number can be owned by some users for legitimate communication and by some others for launching SMS spam during the observation period. To address this issue, we consult the service plans of the phone numbers and identify their service starting times and ending times, which help uniquely identify each phone number. For instance, even with the same 10-digit sequence, a phone number which has a service plan that ends in January and is reopened in May will be counted as two different numbers in these two months. Hereafter we shall follow this definition to identify spam numbers.

After preprocessing, from the one-year user-generated spam reports, we extract a total of 78.8K spam numbers. Fewer than 1,000 spam messages were reported daily in 2011, and since 2012 this number has increased steadily and reached above 5K after April 2012. Furthermore, the number of new spam numbers reported has also increased over time (albeit not as significant). These increases are likely due to two factors: i) SMS spam activities have grown considerably over time; and ii) more users have become aware of – and started using – the spam reporting service. We also observe a clear day-of-week effect because spamming activities are more significant during week days.

4.3 Analyzing Spam Text Content

Our initial analysis on the text content of the reported spam messages reveals many interesting observations which we summarize as follows. We find among all the user reported spam messages, 23% of them contain reply phone numbers and 75.1% of them contain at least one valid URL, where 7.4% of these URLs used URL shortening service like TinyURL [17]. This is likely due to the limited SMS message length and spammers' intention of hiding the real phishing sites, which are much easier to be identified by mobile users. We find that 74.6% of the domain names associated with the embedded URLs are lookupable, i.e., they can be resolved to a total of 595 unique IP addresses. For these 595 IP addresses, 443 (74.4%) are associated with one domain name, while the rest of the 152 IP addresses are corresponding to multiple domain names. We find each

of these 152 IP addresses is usually associated with a relatively large number of domain names. For example, the largest one is associated with 50 domain names. Moreover, these IPs tend to come from similar subnets.

We further examine the domain names mapped to the same IP address. By looking at the keywords within these domain names, we find clusters of domain names belonging to different topics. For example, we find an IP address that hosts domain names related to free rewards and free electronic devices, where the corresponding domain names look very similar, such as *1k-reward.xxx* and *1krewards.xxx*, and *cell-tryouts.xxx* and *celltryout.xxx*. These observations imply that spammers are likely to rent hosting servers from certain IP ranges that are managed with loose policies. On each hosting server, they tend to apply for multiple domain names and create a separate website for each domain name. In this way, spammers can maximize the utilization of the phishing sites.

An interesting observation is that most spam messages are customized. Over 60% of the messages contain random numbers or strings. These random numbers or strings are often claimed as identification codes or are part of the URLs inside the spam messages. We suspect these random contents are used to differentiate spam victims for two purposes. First, when victims access the phishing sites through the URLs, such random content helps the spammer estimate the effectiveness of the spamming activities. We believe some spammers are paid based on how many unique victims are attracted to the phishing sites by the spam messages. Second, by recording the victims who reply to the spammers or access the phishing sites, spammers can obtain a list of active (or vulnerable in some sense) mobile phone numbers to increase the success rate of future spam activities.

5 Characterizing Spam Numbers

Using spam numbers extracted from the user spam reports, we gather various other sources of data associated with these numbers, such as account and device profiles, network and traffic level data and statistics (voice, SMS and data usage patterns, geolocations, and so forth). By analyzing and correlating these data sources, we study the various characteristics of individual spam numbers.

5.1 Device and Tenure

Device: In order to identify the devices employed by spammers, we extract the first 8-digit TAC from each IMEI associated with spam numbers and match it against a TAC lookup table. The table was created by the carrier in January 2013, which covers the most popular mobile devices in the cellular network under study.

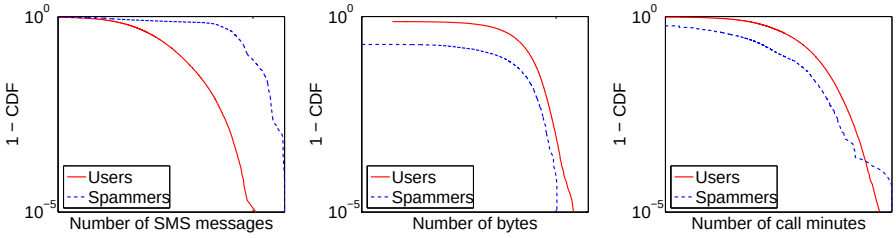
We find that nearly half of the devices are smartphones (44.5%). The rich functionality of these devices enables spammers to create apps to automate SMS spamming activities. There are 20.3% of the devices that have an *unknown* TAC type – this is likely due to either unpopular spam devices or random IMEI numbers generated by SIM boxes. Programmable devices such as 3G data modems, laptops/netbooks, data cards, etc. account for a total of 11.7% devices used in SMS spam. Interestingly, many

"M2M" (machine-to-machine) devices (e.g., used for vehicle tracking and vending machines) are also employed by spammers for sending SMS spam. Costs (both in terms of the devices and the account contracts/payment methods available to them) likely play a role in determining what types of devices are deployed for SMS spam campaigns.

Tenure. Here *tenure* is defined as the time from when the account of the spam number is first enrolled in the service until the first spam message from that spammer is reported. We find that a majority of the spammers hold new accounts. In particular, over half of spam numbers have a tenure of only one day and more than 60% of them have a tenure less than a week (similar observation was made in [10]).

5.2 SMS, Voice and Data Usage Patterns

We now study the overall SMS, voice and data usage patterns of spam numbers, and compare them with the rest of legitimate numbers [6]. For data usage patterns, only those spam numbers with data activities are used. Figs. 2[a-c] display the comparison in terms of the number of SMS messages [a], the number of bytes of data [b] , and the total call duration [c] over the same one month observation period. Not surprisingly, spam numbers initiated far more SMS messages than legitimate ones (Fig. 2[a]). In fact, we observe that 80% of the spam numbers send more than 10K SMS's, and half of the spam numbers send more than 100K SMS's. In comparison to SMS usage, spam numbers consume very little data as represented by the much fewer number of bytes (Fig. 2[b]). However, among the spam numbers which do initiate data communications, the data activities more often than not involve financial sites such as banks. Further investigation of whether such data traffic is associated with security attacks or other illicit financial transactions is left to future work.



(a) Number of SMS's (b) Number of data bytes (c) Total call minutes

Fig. 2. Compare monthly SMS/data/voice usage of reported spam numbers to legitimate numbers.

The total call minutes of spam numbers are generally shorter than those of legitimate ones (Fig. 2[c]). However, we find some spam numbers may initiate even far more (though generally short) voice calls than legitimate ones do. We count the out-going

[6] Though we have checked the tenure and device information of the legitimate numbers to remove likely spam numbers, there is still a chance that a few spam numbers are included in these legitimate numbers. However, we believe this does not affect our analysis of the usage behaviors of legitimate numbers given their large population size.

voice calls from spam numbers and find 10 spam numbers which have initiated more than 10K voice calls. All of them were reported by users on popular online forums [18] as being involved in telemarketing and other voice related fraud activities [19]. It is possible that these spam numbers harvest live mobile numbers through voice calls in order to increase the efficiency of spamming.

6 Network Characteristics of Spam Numbers

Using the SMS CDRs, we next study the network characteristics of spam numbers and infer the spamming strategies adopted by spammers.

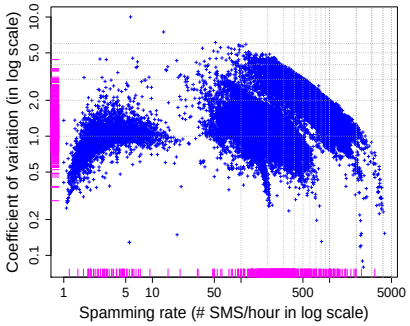


Fig. 3. Spamming rate and variability.

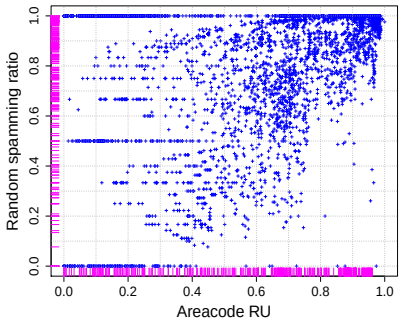


Fig. 4. Target selection strategies.

6.1 Spam Sending Rate

We measure the SMS spamming rate using the average number of SMS messages sent from each identified spam number per hour. We assess the variability of spamming rates using the *coefficient of variation*, which is defined as $c_v = \sigma/\mu$, where σ and μ represent the standard deviation and mean spamming rate of each spam number, respectively. The coefficient of variation shows the extent of variability relative to the mean sending rate. Fig. 3 displays the mean spamming rate and the corresponding coefficient of variation for individual spam numbers. For ease of visualization, we illustrate the marginal densities along both axes using rug plots. We observe that the spamming rate varies from a few to over 5,000 spam messages per hour. In addition, while the majority of spamming activities are at a constant rate (i.e., with a low c_v close to the x-axis), some numbers exhibit more bursty spamming behaviors, i.e., with a c_v greater than 3. From these two metrics, we observe three distinct regions, which we refer to as "slow," "moderate," and "fast" spammers (i.e., three clusters from left to right in Fig. 3). "Moderate" spammers cover 63% of all spam numbers, while "fast" spammers and "slow" spammers account for 20% and 17%, respectively. Further investigation shows that the spamming rates often depend on the devices used and the network locations of the spammers.

6.2 Target Selection Strategies

We next study how spammers select spamming targets. Let $X = \{x_t\}, 1 \leq t \leq T$, denote the sequence of phone numbers that a spam number sends messages to over time. Given the fact that each phone number is a concatenation of two components: the 3-digit area code x_t^a, which is location specific, and the 7-digit subscriber number x_t^s, we also characterize the target selection strategies at two levels, i.e., how spammers choose area codes and phone numbers within each area code.

We use 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:

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

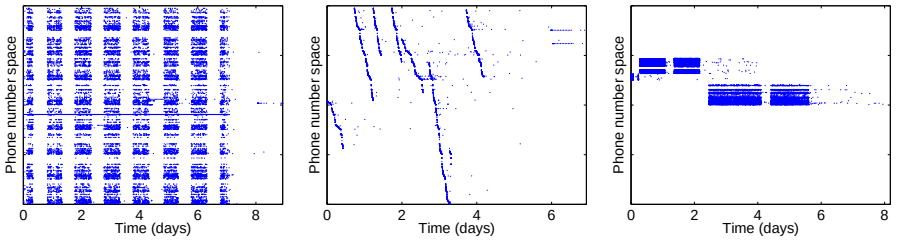
where $P(q)$ represents the proportion of target phone numbers with the same area code q and $|Q|$ is the total number of area codes in the phone number space. Intuitively, a large ru_a (e.g., greater than 0.8) indicates that the spammer uniformly chooses targets across all the area codes. In contrast, a small ru_a means the targets of the spammer are concentrated by sharing only a few area codes.

We next define a metric *random spamming ratio* to study how spammers select targets within each area code. Let P^a be the proportion of active phone numbers with area code a. For a particular spamming target sequence X^a of a spam number, if the spammer randomly choose 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 with random spamming strategies (i.e., when 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.

Fig. 4 plots the ru_a (the x-axis) and the random spamming ratio (the y-axis) for individual spam numbers. Based on the marginal density of ru_a, we find that a majority of spam numbers (78%, using $ru_a = 0.8$ as a cut-off threshold) concentrate on phone numbers within certain area codes. We refer to such a spamming strategy as *block spamming*. In comparison, the remaining 22% spam numbers adopt a *global spamming* strategy, i.e., selecting targets from the entire phone number space. We rank area codes by their popularity among spam numbers, i.e., how many spam numbers select the most target numbers from a particular area code. In fact, we investigate the top 20 popular area code among spammers and find that most of them correspond to large cities and metro areas, e.g., New York City (with 3 area codes) , Chicago (2), Los Angeles (2), Atlanta, and so on.

Based on the y-axis, we find that, no matter how a spam number chooses area codes, a predominant portion of them select targets randomly within each area code. This is likely accredited to the finite phone number space, which enables spammers to enumerate phone numbers to send spam messages to. Such random spamming strategies are

of almost zero cost and hence are the most economic strategies for spammers. Furthermore, this explains why spammers favor large metro areas, because they are likely to reach more active mobile users by randomly selecting numbers from these area codes.



(a) Global random (b) Global sequential (c) Block random

Fig. 5. Foot prints of most representative target selection strategies

We illustrate in Fig. 5 the "footprints" of three most popular target selection strategies, where the x-axis represents time and the y-axis stands for numbers in the phone number space. The *global random spamming* is shown Fig. 5[a], where a spammer randomly chooses phone numbers from the entire phone number space [7]. In comparison, in the *global sequential spamming* strategy (Fig. 5[b]), a spammer enumerates numbers in the phone number space in an ascending order and sends spam messages to each phone number sequentially. Different from the above two strategies, *block random spamming* only focuses on victims within certain area codes, and selects victims from each area code randomly; see Fig. 5[c] for an example (the *block sequential spamming* strategy, observed less frequently, is omitted due to space limit).

6.3 Spamming Locations and Impact on the Cellular Network

We end this section by an assessment of the sending locations of spam messages and the potential impact of spamming traffic on the cellular network. We define the location of a spam number as the serving node-B from which a spam message is sent by that spam number. We find there are a few spam numbers (4.9%) which are highly mobile, i.e., they utilize more than 10 node-B's and distribute their workload among these node-B's (i.e., with the proportion of spam messages from the most dominant node-B less than 40%). However, most spam numbers initiate spam at less than 5 node-B's (78.2% spam numbers) and the most dominant node-B carry more than 60% of the traffic (74.5%). We hence refer to these dominant node-B's as the *primary spamming locations* for spam numbers. In fact, many of these node-Bs reside in densely populated metro areas (e.g.,

[7] Note that most spam numbers are programmed to avoid well known area codes that are unlikely to contain active mobile users or inflict extra cost when sending SMS to, e.g., 900 area codes and area codes of foreign countries which adopt the North American Numbering Plan (NANP). This results in ranges of phone numbers never assessed by the spam number (i.e., shown as the blank horizontal regions in Fig. 5[a]).

New York City and Los Angeles). We suspect that concentrating on densely populated urban areas enables spammers to easily obtain resources, like used phone numbers. In addition, spammers can take the advantage of the high-speed 3G/4G network at these locations to spam in much higher rates.

At these node-B's, we find that the sheer volume of spamming traffic is astonishing. The spamming traffic can exceed normal SMS traffic by more than 10 times. Even at the RNC's, which serve multiple node-B's, the traffic from spamming may 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. These observations also emphasize the necessity of restricting spam numbers earlier before they reach many victims and inflict adverse impact on the cellular network.

7 Investigating the Correlations between Spam Numbers

So far we have focused on the characteristics of *individual* spam numbers. In this section we will cluster spam numbers based on the content similarity of the spam messages they generate, and characterize and explain the correlations between spam numbers.

7.1 Clustering Spam Messages with CLUTO

Recall that, through our initial manual content inspection, we have observed that many spam numbers are reported to have generated the same or similar spam messages. We hence apply a text mining tool–CLUTO [8, 20]–to cluster spam messages with similar content into spam clusters. CLUTO contains many different algorithms for a variety of text-based clustering problems, which have been widely applied in research domains like analyzing botnet activities [21]. After testing different clustering algorithms implemented in CLUTO, we choose the most scalable k-way bisecting algorithm, which yields comparable clustering results to other more sophisticated algorithms.

Table 1. Example spam messages from the same clusters.

Raymond you won ... Go To apple.com.congratsuwon.xxx/*codelrkfxxxxxx*
Laurence you won ... Go To apple.com.congratsuwon.xxx/*codercryxxxxxx*
You have been chosen ... Goto ipad3tests.xxx. Enter: *68xx* on 3rd page
You have been chosen ... Goto ipad3tests.xxx. Enter: *16xx* on 3rd page

Before applying CLUTO, we first compute a similarity matrix for all the spam messages, using the *tf-idf* term weighting and the cosine similarity function. Operating on the similarity matrix, the k-way bisecting algorithm repeatedly selects one of the existing clusters and bi-partitions it in order to maximize a predefined criterion function. The algorithm stops when K clusters are formed. We explore different choices of K's

and select the largest K such that trivial clusters (i.e., which contain only one message) start to appear after further increasing K. Details regarding how to apply CLUTO for clustering spam messages can be found in [22].

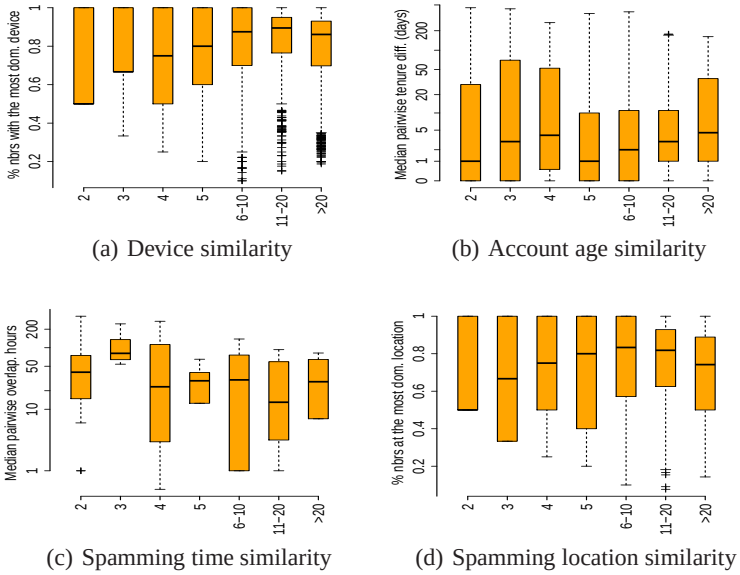
We manually investigate and validate the clusters identified by CLUTO. Not surprisingly, we find that spam messages within the same cluster are generally similar except for one or two words. Table 1 demonstrates examples of spam messages that belong to two different clusters, where the variant text content is highlighted in blue italics. We suspect that such variant content is specific to each spam victim. Spammers rely on such content to distinguish and track responses from different victims and possibly get paid according to the number of unique responses. In the end, we obtain 2,540 spam clusters that cover all the spam messages. We observe that most of the clusters (92%) contain multiple spam numbers and 48% can cover more than 10 spam numbers. In the follow-up analysis, we focus on the top 1,500 clusters which exhibit an intra-cluster similarity greater than 0.8, and investigate the correlations of the spam numbers inside these clusters. These clusters cover totally over 85% of the reported spam messages.

7.2 Correlation of Spam Numbers

Device similarity. We start by comparing the device types associated with individual spam numbers. We define the *device similarity* as the proportion of spam numbers within each cluster that use the most dominant device of that cluster. Fig. 6[a] shows the distribution of device similarities. For ease of comparison, we bin spam clusters based on their sizes with the purpose of ensuring enough samples in each bin. We note that in the rest of our analysis, we shall follow the same binning scheme for consistency. We observe that all the bins exhibit strong device similarities, i.e., all with a median similarity greater than 0.5. Meanwhile, device similarity strengthens as the spam clusters become larger. For example, the median device similarity is above 0.8 for clusters with more than 5 spam numbers. This suggests that spam numbers within each cluster tend to be associated with the same cellular device for launching spam.

Account age difference. We next consult the account information of the spam numbers and identify their most recent account initiation dates prior to the occurrence of spam traffic. We note that after purchasing a spam number, a spammer may spend some time preparing for spamming by sending out a few test messages. Taking this into consideration, we refer to the *account age* of a spam number as the time span from the account initiation date to the first date with observed active spamming behaviors (i.e., the first date with a spamming rate above 50 messages per hour based on Fig. 3).

We measure the *account age difference* of spam numbers in each cluster using the their median pairwise absolute account age difference (in days). From Fig. 6[b], we see the median values of such difference in all the bins are below 5 days. Such a small difference indicates that most spam clusters employ spam numbers acquired within a short time period, e.g., purchased from the same retailer at the same time. In fact, for 30% of the clusters, spammers start spamming actively at the same date when all the spam numbers are initiated, 73% within 3 days and 82% within one week. This implies



(a) Device similarity (b) Account age similarity

(c) Spamming time similarity (d) Spamming location similarity

Fig. 6. Correlation of spam numbers belonging to the same spam clusters.

that monitoring and tracking purchases of bulks of phone numbers by the same user can be an effective way of alerting potential spam clusters.

Spamming time similarity. After investigating the similarity of demographic features, we next compare the spamming patterns of spam numbers. We first explore whether spam numbers within each cluster tend to send spam actively during the same time period. We define the time similarity as the median pairwise overlapping time (in hours) with active spamming behaviors (i.e., more than 50 messages per hour), which is displayed in Fig. 6[c]. In most of the bins, the median values are above 20 hours, which implies a strong temporal correlation among these spam numbers.

Spamming location similarity. Another spamming pattern we investigate is the spamming locations of spam numbers. We define the *location similarity* as the proportion of spam numbers within a cluster with primary spamming locations being the most dominant one in that cluster. Fig. 6[d] displays the distribution of the location similarity, which again appears to be very significant. The similarity reaches 0.8 when the cluster size equals 5 and drops slightly as cluster size further increases. We investigate the clusters with more than 20 spam numbers and find that many of these phone numbers have primarily locations in closeby node-B's. We suspect that this is because spammers want to increase the spamming speed by deploying multiple numbers at nearby locations.

To summarize, various independent evidences from our analysis above of the spam clusters demonstrate that spam numbers within the same cluster are strongly correlated. We believe that the spam numbers contained in the same clusters are very likely employed by the same spammers. These spammers purchase a bulk of spamming devices and phone numbers and program them to initiate spam. These spam numbers thus ex-

hibit strong spatial and temporal correlations. Meanwhile, we observe that for more than 80% of the clusters, the spam numbers in the cluster employ similar spamming rates and target selection strategies (i.e., in the same category defined in Fig. 4[a][b]). It implies that spammers often program their spamming devices in a similar way (often at the maximum speed allowable for the devices at the locations of the network). In comparison, spam numbers exhibit little correlation across clusters, indicating that different clusters are likely caused by different spammers (likely) from different locations.

8 Implications on Building Effective SMS Spam Defenses

Based on our previous analysis on various aspects of SMS spam numbers, in this section, we pinpoint the inefficacy of existing solutions solely replying on user spam reports. We then propose several novel and effective spam defense methods.

8.1 Are User Spam Reports Alone Sufficient?

As we have mentioned, many cellular carriers today rely primarily on user spam reports for detecting and restricting spam numbers. Unfortunately, such a user-driven approach inevitably suffers from significant delay. For example, the black solid curve in Fig. 7 measures how long it takes for a spam number to be reported after spam starts (i.e., *report delay*). We consider a spam number starts spamming when it first reaches at least 50 victims in an hour. From Fig. 7, we observe that only less than 3% of the spam numbers are reported within 1 hour after spam starts. More than 50% of the spam numbers are reported 1 day after. This is likely due to the extreme low spam report rate. Compared with the huge volume of spam messages, less than 1 in 10,000 of spam messages were reported by users in the 1-year observation period.

While most of the report delay is due to the extremely low spam report rate, even users who do report spam may also introduce delay on their side, partly due to the inconvenient two-stage reporting method. The red dotted curve in Fig. 7 shows how fast a user reports a spam message after receiving it. Since each user can receive multiple spam messages from the same spammer and can report the same report number multiple times, we define *user delay* as the time difference between when the user reports a spam message and the *last* time that the user receives spam from that particular spammer before the report. We observe in Fig. 7, among the users who report spam, half of their reports arrive more than 1 hour after they receive the spam messages. Around 20% of the spam messages occur after one day. In fact, even for those users who report spam, we find around 16.8% of them stop at the first stage and fail to supply the corresponding spam numbers, not to mention the inaccurate spam records caused by users mistyping spam numbers.

Such report delay is amplified when used for detecting multiple spam numbers employed by the same spammers. For example, we measure the earliest report times of all spam numbers in each of the clusters which we identified in Section 7 that contain at least 5 spam numbers. Fig. 8 demonstrates the total time (in hours) required for users to report 50%, 80% and all spam numbers in each cluster, respectively. We again observe a significant delay in user reports. In particular, for 80% of the clusters, it takes 20 hours

for users to report half of the spam numbers in them. It takes even more than 38 hours for users to report 80% of the spam numbers in them.

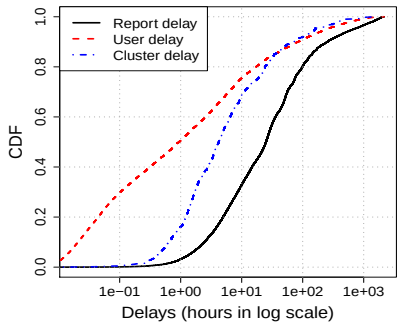


Fig. 7. Different kinds of delays associated with user reported spam messages.

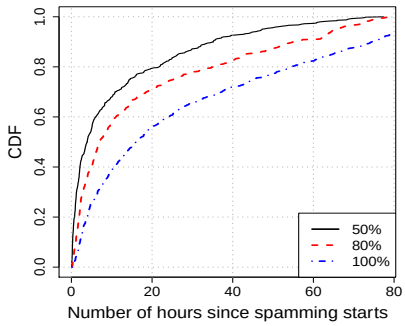


Fig. 8. Time for users to report multiple spam numbers in each cluster.

Therefore, spam defenses relying solely on the current user spam reports can be late and can miss many spam numbers due to both the low report rate and report delay. Advertising can be useful to increase the users' awareness of the spam reporting service and hence can help increase the report rate. Meanwhile, incentives (e.g., credits) provided by cellular carriers can encourage more users to report spam they have received. In addition, an enhancement of the existing cumbersome two-stage reporting method is also important to prevent mistakes during spam reporting and ultimately increase spam report rate. As an example, on smartphones, we are currently developing a mobile-app based solution which enables users to report spam via one single click.

8.2 Detecting Spam Numbers through Spatial/Temporal Correlations

In addition to improving the existing spam reporting, we can also design more efficient spam defenses that are less dependent on user spam reports. For instance, although it takes a long time for a majority of the spam numbers in each cluster to be reported by users, the first report regarding a particular spam number often comes much faster. In Fig. 7, we show for the top 1500 clusters in Section 7, how long it takes for the first number in each cluster to be reported after any number in the cluster starts spamming (i.e., *cluster delay*). For 15% of the top 1500 clusters, we find the earliest report comes within an hour and for 70% of them the first report comes within 10 hours. Given our observation that spammers often employ multiple spam numbers, once a number has been reported, we can detect other related numbers earlier by exploring their temporal and spatial correlations with the reported number, instead of waiting for users to report them.

We illustrate our idea in Algorithm 1, which consists of two components. First, we continuously monitor all SMS senders in the network and maintain a watchlist of

phone numbers at different geolocations (node-B's) that have sent SMS messages to more than β recipients in each time interval of length T[8]. Second, the detection part is triggered by a confirmed spam number (e.g., from user spam reports). In particular, when a spam number in the watchlist is confirmed, we look for all the other numbers from the watchlist whose primary spamming locations (i.e., node-B's) is the same as the confirmed number and report them as spam number candidates.

Algorithm 1 Detecting correlated spam numbers.

```
Input: T, β
//Maintaining a watchlist
for all Locations l do
    Within the observation window T, identify W_l={nbr: nbr at location l has sent SMS's to
    more than β recipients}, and W := ∪W_l;
end for
//Detecting spam numbers by geo/temporal correlations;
loop
    if A spam number x is confirmed and x ∈ W then
        Obtain the location l associated with x;
        Output spam number candidates W_l − {x};
    end if
end loop
```

We simulate the detection process on a month long dataset consisting of CDRs and spam reports received during that month. The proposed algorithm detects 5,121 spam number candidates, 4,653 (90.9%) of which were reported later by mobile users via spam reports. We have the remaining unreported candidates investigated by fraud agents. The investigation combines information sources such as spam reports from online forums (e.g., [24]), service plans, devices as well as the expert knowledge. In the end, 465 of them have been validated to be spam numbers. In other words, the proposed algorithm is highly accurate, with only 3 (less than 0.06%) candidates not yet verified. In addition, we observe that in more than 93% of the cases, the proposed algorithm detects spam numbers an hour ahead of user reports. More than 72% and 40% of the detection results are 10 hours and 1 day before user reports arrive. In fact, more than half of the spam messages can be reduced by detecting and restricting spam numbers using our method. From the perspective of spammers, the proposed method can only be evaded by either reducing the spamming speed, employing a single number for spamming or distribute numbers at different network locations. Nevertheless, any of them will either limit the impact of spamming or significantly increase the management cost.

[8] We note that, the process of maintaining watchlists is similar as running a real-time spam detection purely based on behavioral statistics associated with individual phone numbers. Here we only utilize SMS volume (fan-out) as the feature and apply a hard threshold for detecting suspicious phone numbers. However, more sophisticated features, e.g., SMS message inter-arrival time, entropy based features, etc., and more intelligent thresholds [6,23], can be applied to further improve the accuracy of the watchlists. For proprietary reasons, the specific choices of parameters β and T will not be released in this paper.

8.3 Trapping Spammers using Honeypots in the Phone Number Space

Because random spamming is the most dominant target selection strategy adopted by spammers, we can explore such randomness to detect spam numbers without relying on user spam reports at all. One idea is to employ *unassigned* phone numbers owned by the carrier as *honeypot numbers* to trap spam messages. These honeypot numbers apparently do not participate in SMS communications and hence any SMS messages towards these numbers are likely to be spam. Spammers, on the other hand, are hard to avoid touching these numbers due to the random spamming strategies they employ. Therefore, by correlating SMS messages collected at different honeypot numbers (with an adequate density), we can potentially detect spam numbers much faster and more accurately, without acquiring the assistance from user spam reports.

Deploying honeypot numbers can sometimes be costly and collecting spam messages targeting these numbers often require additional resources. One alternative is to monitor messages to existing *SMS inactive* phone numbers, referred to as *grey phone numbers*. These grey phone numbers are associated with data only devices like laptops, data modems, ereaders, etc., and machine-to-machine communication devices, such as vending machines, security alarms and vehicle tracking devices, etc. Because these devices rarely communicate through SMS, they behave like honeypot numbers and hence any messages towards them are also likely to be spam. For details regarding the grey phone number based spam detection method, please see [25].

9 Conclusion and Future Work

In this paper, we carried out extensive analysis of SMS spam activities in a large cellular network by combining user reported spam messages and spam network records. Using thousands of spam numbers extracted from these spam reports, we studied in-depth various aspects of SMS spamming activities, including spammer's device type, tenure, voice and data usage, spamming patterns and so on. We found that most spammers selected victims randomly and spam numbers sending similar text messages exhibit strong similarities and correlations from various perspectives. Based on these facts, we proposed several novel spam detection methods which demonstrated promising results in terms of detection accuracy and response time. Our future work involves designing user friendly spam reporting framework to encourage more reports and developing a system for real-time spam detection based on our analysis results.

Acknowledgement

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.

References

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

2. 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/.
3. 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.
4. T. Ouyang, S. Ray, M. Rabinovich, and M. Allman. Can network characteristics detect spam effectively in a stand-alone enterprise? PAM'11, 2011.
5. M. Sirivianos, K. Kim, and X. Yang. Introducing Social Trust to Collaborative Spam Mitigation. In *INFOCOM'11*, 2011.
6. S. Hao, N. Syed, N. Feamster, A. Gray, and S. Krasser. Detecting spammers with snare: spatio-temporal network-level automatic reputation engine. USENIX Security Symposium'09, 2009.
7. 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 *NDSS'09*, 2010.
8. Cluto - software for clustering high-dimensional datasets. http://glaros.dtc.umn.edu/gkhome/views/cluto.
9. Y. Zhao and G. Karypis. Criterion functions for document clustering: Experiments and analysis. Technical report, University of Minnesota, 2002.
10. I. Murynets and R. Jover. Crime scene investigation: Sms spam data analysis. IMC'12, 2012.
11. K. Yadav, P. Kumaraguru, A. Goyal, A. Gupta, and V. Naik. Smsassassin: crowdsourcing driven mobile-based system for sms spam filtering. HotMobile '11, 2011.
12. G. Cormack, J. Hidalgo, and E. Sánz. Feature engineering for mobile (sms) spam filtering. SIGIR '07, 2007.
13. H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Zhao. Detecting and characterizing social spam campaigns. IMC '10, 2010.
14. 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. WWW '12, 2012.
15. 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. WWW '12, 2012.
16. C. Grier, K. Thomas, V. Paxson, and M. Zhang. @spam: the underground on 140 characters or less. CCS '10, 2010.
17. Tinyurl. http://tinyurl.com/.
18. 800notes - Directory of unknown callers. http://www.800notes.com.
19. N. Jiang, Y. Jin, A. Skudlark, W. Hsu, G. Jacobson, S. Prakasam, and Z.-L. Zhang. Isolating and analyzing fraud activities in a large cellular network via voice call graph analysis. MobiSys'12, 2012.
20. Y. Zhao, G. Karypis, and U. Fayyad. Hierarchical clustering algorithms for document datasets. *Data Min. Knowl. Discov.*, 2005.
21. G. Jacob, R. Hund, C. Kruegel, and T. Holz. Jackstraws: picking command and control connections from bot traffic. SEC'11, 2011.
22. A. Skudlark N. Jiang, Y. Jin and Z.-L. Zhang. Understanding and detecting sms spam through mining customer reports. Technical report, AT&T Labs, 2012.
23. A. Ramachandran, N. Feamster, and S. Vempala. Filtering spam with behavioral blacklisting. CCS '07, 2007.
24. Sms watchdog. http://www.smswatchdog.com.
25. N. Jiang, Y. Jin, A. Skudlark, and Z.-L. Zhang. Greystar: Fast and accurate detection of sms spam numbers in large cellular networks using gray phone space. USENIX SEC'13, 2013.