# MobiPot: Understanding Mobile Telephony Threats with Honeycards

Marco Balduzzi
Trend Micro Research

Payas Gupta
New York University Abu Dhabi

Lion Gu
Trend Micro Research

Debin Gao
Singapore Management University

Mustaque Ahamad
Georgia Institute of Technology and New York University Abu Dhabi

## ABSTRACT

Over the past decade, the number of mobile phones has increased dramatically, overtaking the world population in October 2014. In developing countries like India and China, mobile subscribers outnumber traditional landline users and account for over 90% of the active population. At the same time, convergence of telephony with the Internet with technologies like VoIP makes it possible to reach a large number of telephone users at a low or no cost via voice calls or SMS (short message service) messages. As a consequence, cybercriminals are abusing the telephony channel to launch attacks, e.g., scams that offer fraudulent services and voice-based phishing or vishing, that have previously relied on the Internet. In this paper, we introduce and deploy the first *mobile* phone honeypot called MobiPot that allow us to collect fraudulent calls and SMS messages. We implement multiple ways of advertising mobile numbers (honeycards) on MobiPot to investigate how fraudsters collect phone numbers that are targeted by them. During a period of over seven months, MobiPot collected over two thousand voice calls and SMS messages, and we confirmed that over half of them were unsolicited. We found that seeding honeycards enables us to discover attacks on the mobile phone numbers which were not known before.

## 1. INTRODUCTION

According to reports from the University of Manchester [17] and the International Telecommunication Union [34], mobile phone subscriptions have grown over 7% yearly in the last ten years. Since October 2014, there have been more mobile phones than people [7]. As of November 2015, the GSMA's real-time tracker sets the number of mobile devices to 7.58 billion [16], overtaking the 7.24 billion estimated world population [8]. Countries like China and India have experienced a huge growth in mobile technologies [27, 6]. For example, China has over 1.2 billion active mobile phones with 93% penetration rate [32].

Cybercriminals, who traditionally relied on the Internet to commit fraud, consider telephony an attractive target not only due to its wider reach, but the fact that people have traditionally trusted it more, making it prone to more effective social engineering attacks *a-lá-Mitnick* for stealing private information or accessing protected systems. As we fortify defenses on the Internet side, telephony provides an alternative path to potential victims for the cybercriminals. They can easily reach such victims with unsolicited calls and spam SMS messages, which has become a serious problem in many countries. Social engineering attacks over the telephony channel to reset online banking credential and steal money have already been reported [20]. Voice phishing attacks can exploit the telephony channel to lure their victims into revealing confidential information like birthday, residence, and credit card numbers [22].

Lately, advances in Internet telephony technologies like VoIP have provided miscreants a fast, cheap, and easy way to conduct large-scale attacks. For example, fraudsters can dial and reach victims via voice calls worldwide at very low cost. Telephony denial-of-service attacks [14] or massive number of robocalls (one-ring calls) [2] have become another form of telephony threats. Recently researchers introduced a telephony honeypot (Phoneypot) aimed at investigating telephony threats, and found evidence of telephony denial-of-service, unsolicited telemarketing, and debt collector abuse [13]. Authors used unassigned telephone numbers – numbers that do not belong to real users – to collect evidence of unwanted calls targeting people in North America. Their work confirmed the existence of a wide variety of telephony abuse, but did not differentiate landline and mobile numbers or actively invite or engage attackers for more in-depth study. In addition to that, the aim of the work was to study accuracy, completeness, and timeliness of data collected to understand telephony abuse.

In this paper, we introduce and deploy a novel *mobile* telephone honeypot that we name *MobiPot* (Mobile HoneyPot) to gain better understanding of mobile telephony threats. First, we configure MobiPot with honeycards (honeypot simcard numbers) to monitor, engage, and record activities of potential attackers who target our honeycards via calls and SMS messages. Unlike email spam, voice calls require active engagement with the callers to understand their goals. To the best of our knowledge, MobiPot is the first system to



*ASIA CCS '16, May 30-June 03, 2016, Xi'an, China*



DOI: http://dx.doi.org/10.1145/2897845.2897890

provide the ability for automated engagement with potential attackers which enables it to record longer conversations and therefore gain better insights into the attacks. Second, we propose and implement several ways to actively advertise or seed honeycards. This allows us to evaluate the effectiveness of various seeding techniques, including those that were not investigated in previous studies. For example, malicious mobile apps are known to steal contact details (phone numbers). By using this seeding method for honeycards, we can find out if such stolen phone numbers actually get calls or messages by fraudsters. Finally, we analyze the call and SMS records to investigate how mobile telephony attackers behave. This led to multiple insights into attacks, including the use of both SMS messages and calls in certain coordinated scams.

In summary, our paper makes the following contributions:

- We propose and deploy the first reported mobile telephony honeypot system called MobiPot with honeycards that come from multiple regions and providers in China.
- We seed honeycards in three distinct ways, including mobile malware, social networks, and abuse lists and analyze effectiveness of these seeding mechanisms.
- Over a period of seven months, we collected 1,021 SMS messages from 215 senders and 634 calls from 413 callers. By using a semi-automated approach, we verify that 82.95% of the SMS messages and 57.73% of the calls are unsolicited and indeed represent mobile telephony abuse/threats.
- We validate our results with public complaint databases and show that a large fraction of source numbers that we classify as malicious were previously unknown.
- We also identify a number of interesting cases that help us better understand the mobile telephony threat.

The remainder of the paper is structured as follow. We first present related work and background in Section 2 with discussions on the difference between an earlier telephony honeypot called Phoneypot [13] and MobiPot. We then introduce MobiPot and its deployment in a realistic setting in Section 3. Section 4 presents the data that we collect with the deployment of MobiPot and our analysis of this data. We present some interesting case studies in Section 5 and discuss additional seeding options in Section 6. We conclude the paper in Section 7.

## 2. RELATED WORK AND BACKGROUND

Honeypots have been extensively used for collecting threat intelligence in computer networks to fight email spam [31], malware [10], and attacks in general [30]. They have also been used to investigate VoIP threats, including spam over IP and Telephony (SPIT) [35] and other VoIP abuse [9].

In contrast to the use of honeypots for Internet threats, there has been limited research on telephony honeypots. There is a MAAWG best practices paper to demonstrate the benefits of having a telephony honeypot [12]. Close to our work, Gupta et al. [13] introduced and deployed a telephony-based honeypot called Phoneypot aimed at investigating telephony-specific threats like telephony denial-of-services (TDoS), telemarketing, and telesurveys. Phoneypot used unallocated telephone numbers to collect evidence of voice abuse. It did not explore SMS abuse and callers were not actively engaged. Call audio was not recorded either.

MobiPot differs from Phoneypot in a number of ways and explores several new areas. Firstly, it specifically focuses on mobile telephony threats by implementing a dedicated simcard-based honeypot. Secondly, in addition to calls, MobiPot extends the analysis to include SMS sent to honeycards and resulting in a collection of twice as many messages as calls. Thirdly, in addition to the callers' source number, MobiPot records the content of both messages and calls. We go beyond Phoneypot by seeding the honeypot numbers in multiple ways and investigating effectiveness of the seeding approaches. In particular, because we focus on mobile phones, we are able to study if phone numbers leaked by malicious applications actually do receive calls.

Besides Phoneypot, there is other related work that targets telephony threats. Jiang et al. [18] performed analysis on voice call graphs to detect telephony frauds from call records. In a following work, the same authors designed and applied a statistical model to detect spam numbers based on their footprints on the grey telephone space [19]. Maggi [22] and Griffin [11] analyzed the voice pishing (vishing) phenomenon on a collection of detailed reports submitted by the victims through a website they deployed. They showed that vishing is very popular in the United States and often conducted by humans, as we confirmed in our work.

## 3. SYSTEM OVERVIEW AND DEPLOYMENT

As discussed in Section 2, there has been a recent deployment of a telephony honeypot that investigated and confirmed the existence of a wide variety of telephony abuse [13]. In this paper, we propose a mobile specific honeypot system called MobiPot that differs from this prior work in terms of its design in a numbers of ways.

First, we focus on mobile phone numbers as the victim and try to identify attacks specifically targeting mobile users. For example, we include SMS messages into our study which were not considered in the previous system [13]. Second, we want to take a more active approach in engaging the sources of abuse calls so that we could extract more information from them. MobiPot does this by engaging the callers and recording the call audio. The passive approach taken by prior work did not do this and simply recorded the caller and called numbers with a timestamp. Third, we consider seeding the phone numbers (making our honeycards known to attackers) part of the deployment process, whereas existing work only passively monitored unused phone numbers. Figure 1 shows an overview of our systems design.

### 3.1 System Architecture

To interface with real mobile phone numbers (on a GSM network with simcards) and at the same time enable automatic recording of calls and SMS messages, we follow a hybrid approach in the system design. We use a GSM-VoIP gateway to virtualize the mobile telephony infrastructure – including its stack – and real mobile phone numbers in the form of simcards (i.e., the *honeycards*) to implement the physical layer. With the GSM-VoIP gateway, we manage multiple honeycards and concurrently receive/transmit over each of them in a single installation. We rely on 8-simcard version of GoIP (`GoIP-8` [21]) as the GSM-VoIP gateway to register our GSM honeycards with the VoIP soft-switch

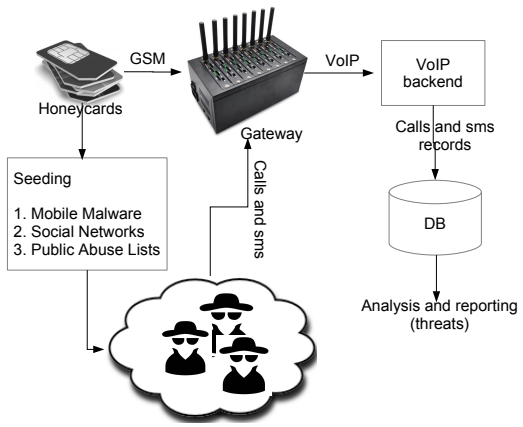



**Figure 1: MobiPot Architecture**

system running Asterisk [1]. The cost of which is approximately 1,000 USD. Asterisk is a well-known open source telephony switching and private branch exchange service for Linux. We use SIP as the communication protocol between Asterisk and `GoIP-8`. Our implementation runs on a standard Linux Ubuntu-32bit installation with 4GB of RAM and 500GB of hard-drive.

### 3.1.1 MobiPot Deployment in China

We configure `GoIP-8` with eight honeycards registered in some of the largest cities of China across two telecom providers (see Table 1). We use the VoIP soft-switch system to emulate a person interacting with the caller. A major challenge with telephony honeypots, as much with honeypots in general, is to keep the attacker (busy) in the system as long as possible. For that purpose, when a call is received, we play an automated engaging message with the goal of incentivizing the caller to keep the call running. We emulate a receiver with hearing difficulties by playing pre-recorded messages (in Chinese) that read *Hello. [...] Hello? [...] Do you hear me? [...] Better now? [...]*. Caller and callee numbers as well as the content of SMS messages and calls were recorded and stored in a database. We focus our evaluation on China for the following reasons.

- China has the largest adoption of mobile phones. Accordingly to Statista, China tops the world with over 1.2 billion active mobile telephones and 93% penetration rate [32]. As of now, there are more mobile phones than traditional landline installations in China.
- Mobile telephony fraud is a serious problem in China. The Ministry of Public Security in China confirmed that over 400,000 users reported to be victims of telephony abuse, and accounted for a loss of 10.7 billion RMB ($\approx$ 1.67 billion USD) in 2014 [28].
- The recording of calls is regulated in many countries, making it hard to perform a similar study. In US, for example, the Federal Communications Commission (FCC) dictates a one-party consent for most of its states, with the exclusion of California and few others [3]. However, in China, there is no specific law for telephony privacy protection which allows us to record calls legally [23, 5]. We recognize that legitimate users may misdial honeycards and could send an SMS message by mistake. This is no different than phone users reaching strangers by mistake. We made sure that any information received by MobiPot was securely maintained (in an encrypted database) and was not shared or used for any purpose other than this study.

In view of these considerations, we chose to have our first deployment of MobiPot in China. We collected the data using our system in China for over seven months from August $22^{nd}$ 2014 to March $27^{th}$ 2015 (both dates included).

### 3.1.2 Seeding MobiPot's Phone Numbers

Another challenge consists of "advertising" phone numbers used by MobiPot to make them appealing for the attackers for abuse. Telephony users are known to receive unwanted calls without the need of advertising their numbers, e.g., from telemarketers that automatically call a multitude of numbers, or simply spam. This could be because the phone number space is limited and a high fraction of possible numbers are allocated. However, to understand how attackers choose phone numbers that are targeted by them (especially in a country like China that has a huge population and a massive number of mobile phones), we design and deploy MobiPot by carefully exposing its phone numbers. The goal is to attract as many unsolicited calls as possible for recording and analyzing; therefore, we investigate various ways of promoting the phone numbers so that more attackers consider these numbers as their targets. Using a PPP model (*P*assive, *P*ublic and *P*rivate), we classify honeycards based on how they are seeded to be attractive for the scrapers and not for legitimate users.

We organized our eight mobile numbers in groups of two and seeded six of them with three different techniques – i.e., a pair of mobile numbers for each seeding technique. The remaining two numbers were not seeded. Table 1 shows the details of our numbers and how and when they were seeded.

#### *Passive honeycards .*

Passive honeycards (`nsd1` and `nsd2`) are never seeded. Calls and SMS messages to these numbers are typically misdialed or randomly targeting phone numbers without any pre-qualification, or attempts to qualify a phone number as "interesting/active". Another reason of unwanted calls and messages could be prior history, where these numbers might have been issued previously to some other entity.

#### *Seeding Public Honeycards– Social Network (`soc1` and `soc2`).*

These honeycards are seeded by actively publishing them at websites in a public domain with the assumption that this will make them attractive to fraudsters but not to legitimate users.

Social networking sites like Facebook, Google+, Twitter, and personal web blogs/web sites could potentially be the targets for fraudsters to scrape and obtain phone numbers. Moreover, some of the online dating websites allow users to provide phone numbers to be a part of their public profile which can be misused by fraudsters. This idea stems from research which suggests that fraudsters may be using social networking sites to entice users to call the numbers they

[1] http://www.asterisk.org

| Label | Honeycard number | Provider | Province | Seeding | | |
|---|---|---|---|---|---|---|
| | | | | Technique | Date(s) / Period(s) | Type |
| nsd1 | 15621192273 | China Unicom | Shandong | Unseeded | N/A | Passive |
| nsd2 | 13477033614 | China Mobile | Hubei | | | |
| soc1 | 18757194227 | China Mobile | Zhejiang | Social Networks | Dec 4 2014 | Public |
| soc2 | 13860141274 | China Unicom | Fujian | | | |
| mal1 | 18701408339 | China Mobile | Beijing | Mobile Malware | Nov 19 2014 – Dec 4 2014 | Private |
| mal2 | 15602228631 | China Unicom | Guangdong | | Dec 11 2014, Feb 19 2015 | |
| abs1 | 13160067468 | China Unicom | Jiangsu | Abuse Lists (Call) | Feb 2 2015, Feb 10 2015 | Private |
| abs2 | 15921962935 | China Mobile | Shanghai | Abuse Lists (Sms) | Dec 30 2014, Jan 15 2015 | |

**Table 1: Seeding of our honeycards**

publish on these sites on false pretexts, like free services or highly discounted articles [1]. However, the challenge here is that the fake profile should be popular enough to be chosen by the fraudsters.

The process of faking profiles and popularizing them is slower as compared to commenting on popular posts/videos as the fraudsters are highly likely to be already scraping the popular sites and blogs. Honeycards can be posted as comments on existing popular sites and blogs.

We identified social networking websites in China that are expected to be crawled by cybercriminals. In particular, we advertised our numbers via three of the most popular social networking platforms, namely a micro-blogging site `Weibo` (Chinese version of Facebook)[2], a video streaming site `Youku` (Chinese version of Youtube)[3], and a blogging site `Baidu Space` (discussion forum)[4]. We publicly advertised honeycards with a message simulating a change of number on these websites.

For example, we promoted on Weibo our "change of number" by embedding popular hashtags in our tweets. *Mandela, the first anniversary of the death of the ### my cell phone is lost, replaced with a new phone number: 18757194227.* For Youku, we used a script to automatically comment popular videos – one of this has been rendered over 5 million times[5].

### *Seeding Private Honeycards– Mobile Malware (*`mal1` *and* `mal2`*).*

Private honeycards are defined as tokens which are not seeded in the public domain but directly to fraudsters.

The popularity and adoption of smartphones has greatly increased the spread of mobile malware, especially popular platforms like Android. According to a recent threats report [24], almost 800,000 new mobile malware are observed per quarter. At the end of 2014, over 6 million samples are known to be in the wild and 10% of them come from Asia, in particular China. In a study of 1,200 Android malware apps [36], the authors show that more than 50% of these malicious apps steal personal information including phone numbers and contacts.

Our second seeding technique consisted of running mobile malware on a testing device which we configured with the honeycards in the contact list. We tracked the leakage of the contact list in two ways: a) by configuring the handset with the `TaintDroid` analysis framework [15]; b) by collecting the network traffic generated by the malware samples that, e.g. connected to C&C servers controlled by the attackers.

We obtained from 369 unique samples of malicious Android applications (from 60 families) known to be leaking private information from Trend Micro. Out of the 60 families, one half consisted of trojanized versions of legit software (i.e., repackaged with malware) and the other half were "standalone" malicious applications offering for example fake messaging, free wallpapers, ring tones, games, and sexual content. We ran each malware on a Nexus 4 running Android 4.3 for 5 minutes with manual interactions in order to trigger possible leakages by, e.g., registering with the applications or engaging in gaming.

The samples were given to us in batches of three with 220, 140, and 9 samples respectively. We ran the first batch on Nov $19^{th}$ 2014 and Dec $2^{th}$ 2014 (repeating), the second batch on Dec $11^{th}$ 2014, and the third batch on Feb $19^{th}$ 2015. The third batch consisted of malware used in the *sextortion* campaign that was ongoing at the time when we conducted the experiments [25]. Out of the 369 samples, 248 samples (i.e., 67%) successfully executed on our testing device. By analyzing the 438MB of network traffic collected at the gateway and the alerts generated by TaintDroid, we identified 264 leakages towards 140 unique C&C servers. All leakages occurred over the HTTP protocol.

### *Seeding Private Honeycards– Abuse Lists (*`abs1` *and* `abs2`*).*

There are large number of web sites publishing suspicious caller numbers, e.g., http://800notes.com. Making calls to those numbers from the numbers associated with honeycards is another approach to seed the honeycards.

We extracted 2,236 unique numbers (1,683 of which are mobiles) from `Lajidianhua`[6] – the largest provider of abuse call lists in China – and contacted them with two honeycards. We used one of the two to send them an engaging SMS message and the other to make a one-ring call to them. Our engaging SMS message reads as *I am fine with our discussion. How do we proceed?*

In Section 6, we discuss some other seeding methods that we leave as future work.

## 4. EVALUATION

In this section, we present the results of our deployment of MobiPot in China over seven months from August $22^{th}$, 2014 to March $27^{th}$, 2015. We collected 1,021 SMS mes-

[2]http://www.weibo.com
[3]http://www.youku.com/
[4]Now re-branded as `Baidu Cloud`: http://yun.baidu.com/
[5]http://v.youku.com/v_show/id_XODM4NzE2NDE2
[6]http://www.lajidianhua.com

sages from 215 senders and 634 voice calls from 413 callers. We also received 66 MMS messages that we ignored because they were not supported by the GSM-VoIP gateway. We first describe our pre-processing of the collected data to filter out noise. We also provide volume and temporal characteristics of unsolicited calls/SMS messages. Thereafter, we discuss the effectiveness of our seeding techniques. Finally, we present a few interesting case studies as results of our analysis.

## 4.1 Unsolicited Calls and SMS Messages

We set up MobiPot to understand the ecosystem behind unsolicited SMS messages and calls that potentially come from fraudsters who abuse the telephony channel. However, not all the calls and SMS messages received on MobiPot are unsolicited. There are multiple reasons why some of them are not. a) The calls and SMS messages received could be the result of misdialing by legitimate users. b) honeycards could have been previously assigned to another legitimate entity, which may lead to SMS messages and calls received during our experiments which are meant for those entities. To decide if a call or SMS message is unsolicited, we adopted a semi-automated approach. Note that in our definition of unsolicited, we included *unwanted* content like spam or robocalls that are not necessarily malicious per-se, but are generally annoying to the user and not wanted.

We first transcribe all calls with an external transcription service called `Wanbo Steno` [33]. With all calls transcribed into Chinese text, we translated SMS and call content into English using `Google Translate`. Before the classification into unsolicited and benign which is a manual process that is tedious and error prone, we automatically cluster the SMS messages and calls into groups to aid the manual process. We used a hierarchical clustering algorithm with Levenshtein as the distance metric and Dunn index to cut the dendrogram. In cases where a URL is embedded in an SMS message, we adopted the web-reputation service offered by Trend Micro to classify it. We automatically labeled all SMS messages as unsolicited that include malicious URL. Calls and the remaining SMS messages were manually classified by two researchers with the aid of the automatic clustering results.

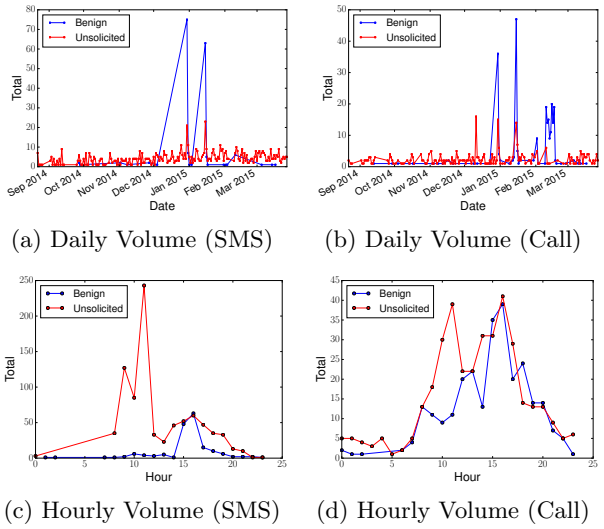
Using this approach, we classified 847 (82.95%) SMS messages and 366 (57.73%) calls as unsolicited. In total, there were 215 sources who sent messages to at-least one of the honeycards.

## 4.2 Volume and Temporal Characteristics

In this subsection, we provide insights into the temporal calls and SMS messages patterns received on MobiPot (see Figure 2). We show the diurnal volume for both benign and unsolicited SMS messages and calls. As it can be noticed, on almost all days, MobiPot received more unsolicited calls and SMS messages as compared to the benign ones.

We collected on average 3.88 unsolicited SMS messages and 1.68 unsolicited calls per day. There was an increase in daily SMS message volume from December 2014 onward during which we performed various seeding exercises on the honeycards (see Figure 2(a)). We explain the effects of seeding in the following subsections in more detail.

An interesting observation of the benign SMS message and call volume is that it was very high on two occasions – much higher than that for unsolicited one. These sharp




(a) Daily Volume (SMS) (b) Daily Volume (Call)

(c) Hourly Volume (SMS) (d) Hourly Volume (Call)

**Figure 2: Overall volume**

increases also coincide with the dates of our seeding, especially with seeding of the two honeycards with abuse lists `abs1` and `abs2`. By checking the sources of these benign SMS messages and calls, we confirm that almost all of them are from the abuse lists that we called during the seeding period. They were not classified as unsolicited because the content appears to be benign. This suggests that a large portion of the numbers on the abuse lists are actually benign. We further investigate this suspicion and discuss it in Section 4.8.

Figure 2(c) and 2(d) report the hourly distribution of calls and SMS messages. We observe that most of the traffic was made during business hours. This confirms the findings reported by [13] that most attack sources blend in with the normal telephony traffic to appear legitimate.

## 4.3 Honeycards

In this section, we provide details of the unsolicited SMS messages and calls received on each of the honeycards. Table 2 shows a breakdown of the SMS messages and calls received over the period when the MobiPot was running.

The first interesting result is that there is a higher likelihood of getting an unsolicited message as compared to unsolicited call. As we can see from Table 2(a), all the messages received on `nsd1`, `mal1` and `mal2` were unsolicited. Moreover, there were 664 out of 679 messages (97.79%) received on six out of eight honeycards (excluding `abs1` and `abs2`, due to reasons explained in Section 4.2) were unsolicited. On the other hand, 241 out of 286 (84%) calls were unsolicited after excluding `abs1` and `abs2`.

During the entire timeframe when the MobiPot was running, `soc1` received the largest number (and percentage) of abuse calls and messages – 303 out of 317 or 95.58% of the hits on `soc1` were unsolicited. On the other hand, `abs1` received the smallest number of unsolicited calls and messages (49 out 190, or 25%). `abs1` was also the worst performer in terms of percentage of unsolicited calls and messages received.

As we can notice, the two numbers seeded in the same way differ considerably in the messages and calls received. We

**Table 2: Breakdown of SMS messages and calls before and after seeding. Mann-Whitney statistical test was used to compute statistical significance. (*) - Tendency, (**) - Significant, (↤) - Before seeding, (↦) - After seeding.**

(a) Based on call and message volume

| Label | # of SMS messages | | | | | # of calls | | | | | # of SMS and calls combined | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Total | Unsolicited | Seeding ↤ | Seeding ↦ | p-value | Total | Unsolicited | Seeding ↤ | Seeding ↦ | p-value | Total | Unsolicited | Seeding ↤ | Seeding ↦ | p-value |
| nsd1 | 209 | 209 | | | | 12 | 10 | | | | 221 | 219 | | | |
| nsd2 | 20 | 16 | | | | 60 | 54 | | | | 80 | 70 | | | |
| soc1 | 281 | 278 | 23 | 255 | 0.0** | 36 | 25 | 6 | 19 | 0.025** | 317 | 303 | 29 | 274 | 0.0** |
| soc2 | 30 | 22 | 4 | 18 | 0.005** | 53 | 37 | 12 | 25 | 0.066* | 83 | 59 | 16 | 43 | 0.004** |
| mal1 | 81 | 81 | 31 | 50 | 0.383 | 97 | 95 | 31 | 64 | 0.27 | 178 | 176 | 62 | 114 | 0.231 |
| mal2 | 58 | 58 | 14 | 44 | 0.028** | 28 | 20 | 5 | 15 | 0.155 | 86 | 78 | 19 | 59 | 0.035** |
| abs1 | 16 | 6 | 4 | 2 | 0.217 | 174 | 43 | 12 | 31 | 0.0** | 190 | 49 | 16 | 33 | 0.0** |
| abs2 | 326 | 177 | 77 | 100 | 0.061* | 174 | 82 | 25 | 57 | 0.068* | 500 | 259 | 102 | 157 | 0.019** |
| Total | 1,021 | 847 | | | | 634 | 366 | | | | 1,655 | 1,220 | | | |

(b) Based on callers and senders

| Label | # of senders | | | | | # of callers | | | | | # of senders and callers combined | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Total | Unsolicited | Seeding ↤ | Seeding ↦ | p-value | Total | Unsolicited | Seeding ↤ | Seeding ↦ | p-value | Total | Unsolicited | Seeding ↤ | Seeding ↦ | p-value |
| nsd1 | 5 | 5 | | | | 10 | 9 | | | | 15 | 14 | | | |
| nsd2 | 9 | 7 | | | | 50 | 45 | | | | 59 | 52 | | | |
| soc1 | 22 | 19 | 5 | 14 | 0.056* | 32 | 22 | 5 | 17 | 0.053* | 54 | 41 | 10 | 31 | 0.01** |
| soc2 | 14 | 9 | 2 | 7 | 0.06* | 50 | 37 | 12 | 25 | 0.066* | 64 | 46 | 14 | 32 | 0.015** |
| mal1 | 2 | 2 | 1 | 1 | 0.398 | 89 | 87 | 30 | 57 | 0.251 | 91 | 89 | 31 | 58 | 0.267 |
| mal2 | 11 | 11 | 2 | 9 | 0.059* | 24 | 19 | 5 | 13 | 0.203 | 35 | 30 | 7 | 22 | 0.078* |
| abs1 | 14 | 5 | 3 | 2 | 0.214 | 46 | 22 | 10 | 11 | 0.049** | 60 | 27 | 13 | 13 | 0.061* |
| abs2 | 147 | 32 | 3 | 29 | 0.048** | 121 | 65 | 22 | 36 | 0.145 | 268 | 97 | 25 | 65 | 0.111 |
| Total | 215 | 84 | | | | 413 | 300 | | | | 583 | 373 | | | |

believe this is because of their history and attackers abuse them differently because they are not equally "dirty". For example, there is a big difference between the number of calls and messages received on `soc1` and `soc2`. `soc1` and `soc2` received a total of 317 and 83 unsolicited calls and SMS messages, respectively. `soc1` received 281 SMS messages and 36 calls, while `soc2` received 30 and 53 only, respectively. Out of the 281 and 30 SMS messages received by `soc1` and `soc2`, 278 (98.9%) and 22 (73.33%) of them were unsolicited respectively. Similar patterns are observed with `nsd1` - `nsd2`, `mal1` - `mal2`, and `abs1` - `abs2`. At this point, we do not know the reason behind this bias, however, as pointed out by previous literature [13], history of the phone number (or honeycard in our case) could be one possible reason. Also note that, these numbers were from different telecom providers in China, and we suspect that it might play a role in the observed differences.

Another interesting finding is that 289 out of 301 SMS messages and calls on `nsd1` and `nsd2` combined were found to be unsolicited. This shows a probability of 0.96 of receiving an unwanted call or SMS message even if the number is not seeded (with slightly higher chances of receiving an unwanted message than that for calls at 98.25% v/s 88.88%). Again, this could have been heavily influenced by the history of the honeycard and might not be interpreted as a finding with general applicability.

We further explain data presented in Table 2 in the next subsection to show effectiveness of seeding.

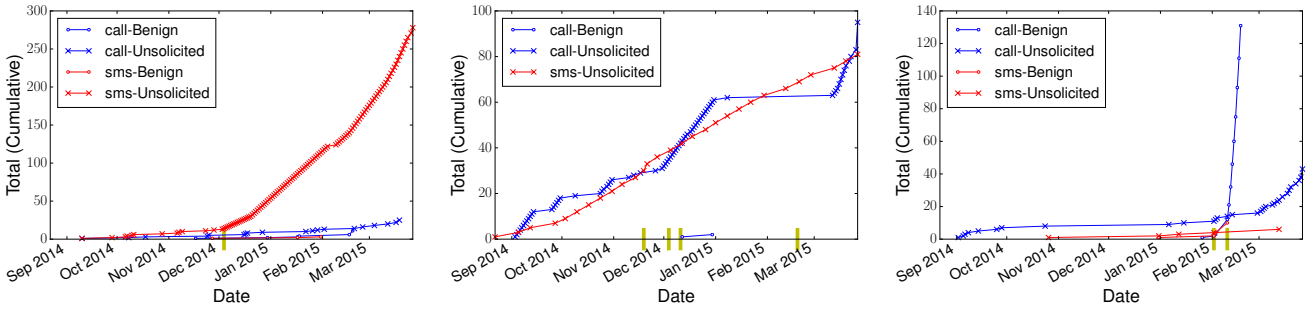
## 4.4 Effect of Seeding on Honeycards

In this subsection, we focus on analyzing the effectiveness of our seeding mechanisms. We first show the per-token volume of SMS messages and calls, see Figure 3 (a–f). To provide easy reference, we indicate in the figure (bars on the x-axis) the time when each honeycard was seeded using different seeding methodology as explained in Section 3.1.2.

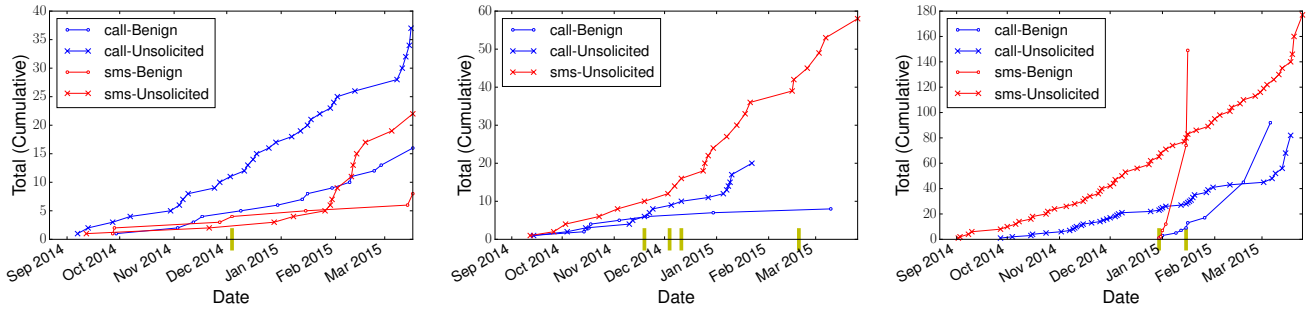
We notice sharper increases in the volume and sources right after seeding in many cases, most noticeable in `abs1` and `abs2`. We separately explain the reasons for each of honeycard later in this section. Although (cumulative) volume gives us a general idea of the total number of unsolicited SMS messages and calls received, in this section, we will focus on finding out whether the contribution comes from more unique senders/callers or more messages/calls per sender/caller. Figure 3 (g–l) plots the cumulative number of sources. It shows that there is an significant increase in the number of unique sources during seeding of honeycards that had contributed to the increase in volume. This serves as a clear indication that telephony fraudsters are actively looking for new targets by, e.g., contact leakage from mobile malware, instead of simply targeting numbers that are "alive".

### 4.4.1 *Abuse on* `soc1` *and* `soc2`

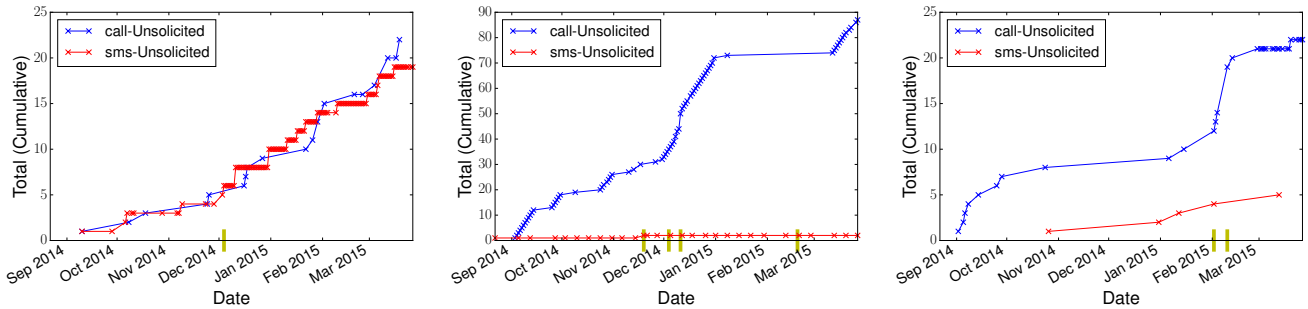
The use of social networks in seeding was very effective, especially in the case of `soc1` where the total number of messages and calls received after seeding was statistically



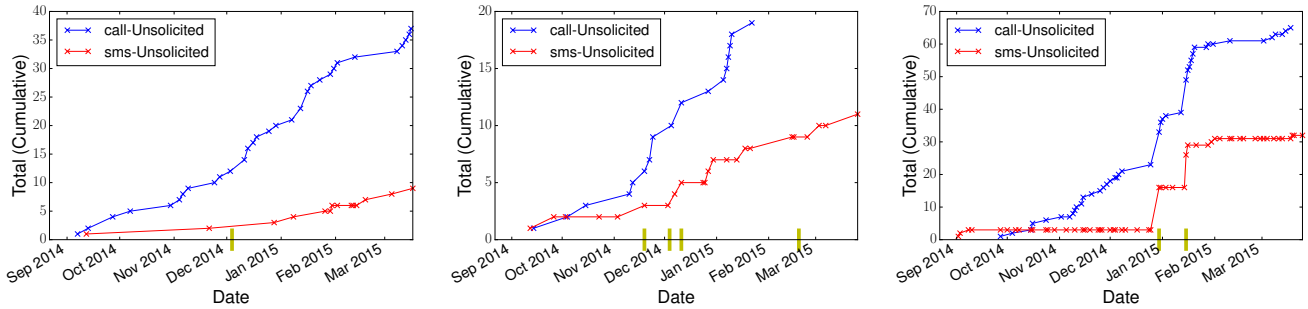

(a) Volume - `soc1` (b) Volume - `mal1` (c) Volume - `abs1`




(d) Volume - `soc2` (e) Volume - `mal2` (f) Volume - `abs2`




(g) Sources - `soc1` (h) Sources - `mal1` (i) Sources - `abs1`




(j) Sources - `soc2` (k) Sources - `mal2` (l) Sources - `abs2`

**Figure 3: (Cumulative) number of unsolicited calls (a–f) and sources (g–l) for SMS messages and calls per honeycard. On the x-axis, the green vertical bar (|) denotes the time(s) when the honeycard was seeded. Note that we do not show the benign sources from figure (g–l). Calls are represented by blue color and SMS messages with red. Benign is represented by `o-o` and unsolicited by `x-x`. For example, unsolicited calls are represented by blue `x-x` and benign SMS messages by red `o-o`.**

significantly higher ($p = 0.0$ and $p = 0.025$, respectively) than the total number of messages and calls received before seeding. This can easily be observed by the sudden increase in the rate in which SMS messages were received right after seeding (indicated by green bars on the x-axis). The increase in the case of `soc2` was not as dramatic, but the Mann-Whitney statistical test also shows significant difference ($p = 0.005$) for SMS messages received before and after seeding, and a distinct trend toward significance ($p = 0.066$) in the case of calls. This is the first sign of our success of seeding the public honeycards `soc1` and `soc2` on online social media which resulted in more unsolicited SMS messages and calls than benign ones.

The profile of account used for `soc1` was picked up by a media website called `Xinhua Quanmei`[7] which broadcasts daily news in the form of spam. We received a total of 221 messages related to websites involved in adware campaigns. This is also the most prevalent cluster of messages from the contributor `106582622` (see Table 4). All the messages from this source were received right from the day on which `soc1` was seeded.

In addition to the volume of calls and messages received by `soc1` and `soc2`, we found that the number of sources contacted either `soc1` or `soc2` is significantly higher after seeding as compared to that before seeding ($p = 0.01$ and $p = 0.015$, respectively). However, we found that individual numbers of sources for calls and messages for both `soc1` and `soc2` has approached acceptable levels of statistical significance but definitely not statistically significant. `soc2` received fewer hits as compared to `soc1`, but at the same time were targeted by more sources.

#### 4.4.2 *Abuse on* `mal1` *and* `mal2`

We seeded `mal1` and `mal2` four times on different dates (see Table 1 and green bars on x-axis on Figure 3(b), 3(e), 3(h) and 3(k). In Table 2(a), we use the first seeding date as a reference point to calculate the statistical significance.

From Table 2(a) we notice that there is no significant effect of seeding on the dirtiness of `mal1`. However, we do see a statistical difference on the number of messages received when `mal2` was seeded. We also tested all four time periods for computing the statistical significance level; however do not see any difference and found the same results. There is a sudden rise in the number of unsolicited calls for `mal1` (see Figure 3(b)) and unsolicited messages for `mal2` (see Figure3(e)) after the second round of seeding. As discussed in Section 3.1.2, the second and third rounds of seeding were more effective as we found more recent malicious set of applications which leaked the phone numbers to the Internet.

`mal1` received a total of 81 messages from two sources, out of which 79 were from `106588302`. All of these messages were advertisement messages with no URL links in them. The other two messages were from `10690123590110` and have a URL in it http://wap.guanxi.me. Apparently, these two messages appeared on the same date when `mal1` was seeded the first time (see Figure 3(h)). We believe that this was the result of our seeding exercise and some application did leak our honeycards. Interestingly, we found that `mal2` also received the same messages from a different sender `1065502004955590110` on the same date when `mal2` was seeded the first time. We believe the two source numbers are spoofed and owned by the same attacker. Our efforts show some early insights on seeding honeycards privately through malicious applications. This serves as the second sign of our successful seeding exercises. We also found that there is a clear tendency to significance ($p = 0.059$) between the total number of messages before and after `mal2` was seeded the first time.

[7]http://www.xhqm.cn/

#### 4.4.3 *Abuse on* `abs1` *and* `abs2`

The third and final seeding exercise was performed using `abs1` and `abs2`, where these two honeycards were seeded directly by contacting the known abuse phone numbers. Figure 3(c) and 3(f) show the total (cumulative) call and SMS message volume before and after seeding dates. Note that unlike other seeded honeycards, `abs1` and `abs2` were seeded at different times.

As it can be noticed from both figures that there has been a noticeable increase in the number of calls and messages on both `abs1` and `abs2`. We believe that the more noticeable increase in `abs1` and `abs2` are characteristics of the seeding methodology of calling and sending SMS messages to numbers on the abuse list, in that the call and SMS messages most likely attracted human attention immediately since it requires human interaction with the attacker. Note that the increase in volume for `abs1` and `abs2` applies to benign SMS messages and calls as well.

We used only the first seeding date as a reference to generate population before and after seeding, and use Mann-Whitney to compute the statistical significance. We found that there is a statistically significant difference between the number of calls received by `abs1` before and after seeding ($p = 0.0$). On the other hand, we only found a margin at the edge of significance for messages and calls received by `abs2` before and after seeding ($p = 0.061$ and 0.068 respectively). We also found that the total number of senders in `abs2` and total number of callers in `abs1` be statistically higher after seeding ($p = 0.048$ and $p = 0.049$, respectively).

Immediately after seeding on both dates, there were many fraudulent transaction messages received by `abs1` and `abs2`. Some examples of these messages (translated in English and masked for privacy) are

- *ICBC: 621226400000XXXXXXXX; Account: accounting Liping; received, please return!*
- *Confirmed Kazakhstan, also grew a position of waiting before QQ news, I just called and asked, and transfers it to me. The fifth branch of the Sichuan branch of China Construction Bank 52409438XXXXXXXX.*
- *Agricultural Bank; number: 622848012080XXXXXXXX; Beneficiary: Lu Yudan; Longgang, Pingshan Branch, Shenzhen Branch*

Moreover, there were a lot of calls and SMS messages from legitimate users to verify our identity, e.g., in asking whether we know each other (see Figure 3(c) and 3(f)). We believe this happened because these (legitimate) peers had their phone numbers listed on the dedicated sites of telephony abuse lists. Following are possible reasons of having these numbers listed on the abuse lists: (a) an adversary spoofed legitimate users' phone numbers and reached out to other people; (b) their mobile phones were infected by malware that used the phones as a bridge (e.g., to send malicious messages without notifying the user); (c) their telephone numbers were previously employed by a malicious actor; and

(d) an adversary voluntarily published the number in form of a complaint.

## 4.5 Classification of Threats

We perform some simple classification on the unsolicited SMS messages and calls, and show our results in Table 3. We find that most unsolicited SMS messages are spam with fewer than 10% being scam messages. On the other hand, scams and spam contribute about equal share in unsolicited calls. We believe this is a sign that attackers find it more effective to scam victims by calling. Another interesting finding is the use of URL in unsolicited SMS messages. It has been very effective to scam victims.

**Table 3: Breakdown of honeycards by threats. (HC) - Human Call, (RC) - Robocall**

| Label | SMS | | | | Calls | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Spam | | | Scam | Spam | | Scam | | Others |
| | URL | Ad | Survey | | HC | RC | HC | RC | HC |
| nsd1 | 193 | 15 | 0 | 1 | 0 | 0 | 1 | 2 | 7 |
| nsd2 | 0 | 16 | 0 | 0 | 5 | 15 | 3 | 3 | 28 |
| soc1 | 223 | 37 | 0 | 18 | 2 | 6 | 1 | 9 | 7 |
| soc2 | 0 | 20 | 0 | 2 | 4 | 5 | 0 | 10 | 18 |
| mal1 | 2 | 79 | 0 | 0 | 1 | 1 | 13 | 4 | 76 |
| mal2 | 3 | 52 | 0 | 3 | 1 | 2 | 1 | 3 | 13 |
| abs1 | 0 | 5 | 0 | 1 | 1 | 20 | 3 | 1 | 18 |
| abs2 | 0 | 85 | 61 | 31 | 5 | 14 | 15 | 12 | 36 |

The results of categorization into human calls and robocalls are also interesting. We found a significantly bigger share of spamming attacks be robocalls (63 against 19), while scam calls have a more equal share between human calls and robocalls (44 and 37, respectively). This seems to be intuitive as scams probably requires higher sophistication in terms of social engineering and phishing, and will have higher success rate when it is conducted by human beings.

There are more than 50% of the unsolicited calls that are classified as others, which include a large number of robocalls with limited or no conversations.

## 4.6 Hot Sources

Table 4 shows the biggest contributors of the unsolicited SMS messages and calls. As expected, these SMS messages are mostly spam with URL and advertisements. Interestingly, the senders of the SMS messages are either the service provider itself (`10086` is the number of China Mobile) or SMS gateways (services offered by the mobile service provider with a prefix of `106`). It appears that attackers use such services to reach out to a large number of victims in a convenient and low-cost way, and the mobile service provider does not stop such spamming with its services. One example is `106558000623`, a marketing content provider called Brisk Hayat[8], which broadcasted 193 messages advertising subscription services.

The top contributor to unsolicited calls is from a landline number in Guangzhou, the third largest city in China. This number has been reported to various abuse lists, and the content of the call is mostly investment invitation into stock market and companies.

All mobile calling numbers were from China, with the exception of four foreign-looking numbers. We validated our results by performing HLR lookups and querying the SS7 signaling network with `Number Portability Lookup` [4]. We confirmed that only one of these (`+6698240898`) was an international number registered with a sim-card in Thailand. Even though the other three numbers pretended to be US based (i.e., starting with `+1`), they were actually invalid numbers likely spoofed by the attackers to avoid easy detection. All these international-looking mobile numbers were fraudulent callers/senders, e.g., a fake provider extorting money with a pretext of contract expiration or a fake postal service requesting a fee to release a parcel retained in customs.

[8]http://dysh.qingk.cn/

**Table 4: Hot Sources**

(a) SMS messages

| Sender | Total | Description | Classification |
|---|---|---|---|
| 106582622 | 223 | Spam | URL |
| 106558000623 | 193 | Spam | URL |
| 10086 | 89 | Spam | Ad |
| 106588302 | 79 | Spam | Ad |

(b) Calls

| Caller | Total | Description | Classification |
|---|---|---|---|
| 02066335588 | 20 | Spam | Robocall |
| 15813449813 | 6 | Others | Robocall |
| 15342606832 | 5 | Others | Robocall |
| 15719210386 | 4 | Scam | Human call |

## 4.7 Campaigns Detection

In this section, we dig deeper into the SMS message content in hope of revealing connectivity among various SMS senders. We believe that attackers are typically using multiple mobile numbers to send out SMS messages due to, e.g., maximum number of free SMS messages each simcard could send and to remain low profile to avoid detection. Here we exploit similarities among the SMS messages to detect senders that are part of the same campaign and, consequently, which honeycards they target.

Our analysis starts with splitting a message $\alpha$ to form a set of words $\mathbb{R}_\alpha$ (see Figure 4 for the pseudo code). We then create a graph where the nodes represent the messages and an edge between two nodes denotes that there is similarity (in terms of Jaccard similarity index between corresponding two sets of words) between the two messages. Finally, we use the Leading Eigenvector community detection algorithm [26] to find clusters, which are the campaigns detected.

With the set of sources that form a campaign and target one or more honeycards detected, we present the results in another graph as shown in Figure 5. In this graph, the bigger nodes represent honeycards and the smaller nodes represent various campaigns detected using the algorithm shown above, with the number inside a small node representing the total number of sources in that campaign. An edge between a campaign node and a honeycard node indicates that a subset of sources of the campaign targeted the honeycard. An edge between two campaign nodes denotes that there are common sources between the two campaigns.

From Figure 5, we notice:

- There are multiple campaigns targeting a particular honeycard (e.g., five campaigns targeting explicitly `soc1`).
- There are multiple sources which are part of the same

```
1: procedure CampaignDetection(𝕊, λ)
2:     𝕊 ← TranslateChineseToEnglish(𝕊)          ▷ 𝕊 is a collection of messages
3:     ℝ ← Tokenize(sms ∈ 𝕊)                     ▷ Tokenize and remove stopwords
4:     G ← Graph()
5:     G.add_vertices(α)  ∀  α ∈ ℝ               ▷ α is sms id
6:     for all α, β ∈ (ℝ 2) do
7:         sim(α, β) ← |ℝα ∩ ℝβ| / |ℝα ∪ ℝβ|     ▷ Jaccard similarity between each pair of message
8:         if sim(α, β) > λ then
9:             G.add_egde(α, β, sim(α, β))
10:        end if
11:    end for
12:    ℂ ← EigenvectorCommunityDetection(G)      ▷ Community detection algorithm to cluster graph
13: end procedure
```

**Figure 4: Campaigns Detection Algorithm**

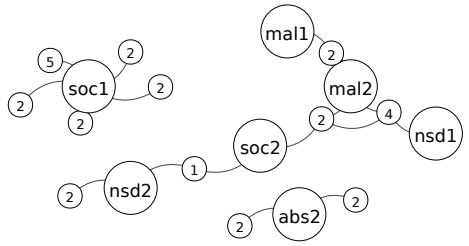



**Figure 5: Social graph of campaigns and honeycards based on content of the messages**

campaign. This confirms our intuition that attackers use multiple numbers to perform their attacks.

- There are common sources between campaigns (an edge between the two campaigns targeting `mal2`). This suggest that there are multiple campaigns run by the same attacker.
- There are campaigns that target multiple honeycards (e.g., the same campaign targeting both `mal2` and `soc2`. This suggests that attackers may use multiple ways of identifying potential victim numbers.

Our findings suggest that the attacks exploiting the telephony channel are complicated and organized activities, similar with cyber-attacks conducted on the Internet. We believe that MobiPot is useful in collecting evidence of such organized activities and the data collected as well as analysis like this provide a first step in better understanding them. Note that, we only show campaigns in Figure 5 which have multiple sources.

## 4.8 Complementing Current Technology

Our deployment of MobiPot has been effective in recording many unsolicited SMS messages and calls with the effective seeding of honeycards. In this subsection, we correlate the findings with existing technologies, namely, public complaint lists, to see if there are new numbers found.

We cross-checked the unsolicited source numbers obtained on MobiPot with public databases of complaint numbers, in particular, `Lajidianhua`. We also installed `Sogou Haoma`[9], a popular mobile application that informs the user whether the incoming call or SMS is untrusted, and exported its abuse list consisting of 57,441 numbers. Lastly, we wrote scripts to search occurrences of these source numbers on Google to find any evidence of the abuse of them.

Overall, 77.47% of the unsolicited numbers recorded by MobiPot were unreported in three public sources we accessed. The percentage goes up to 89% when excluding `abs1` and `abs2` which were seeded via Lajidianhua (also used for this measurement). This suggests that MobiPot opened up a new and effective avenue in finding sources of unsolicited SMS messages and calls, and could potentially be used to complement existing technologies in better understanding and defending against the attacks.

# 5. CASE STUDIES

Besides the main results presented in the previous section, we also find some interesting cases that we believe are worth sharing. In particular, we report cases that are specific to mobile phone users which are not found in existing work of telephone-based honeypots [13].

## 5.1 Re-using Mobile Numbers

In a big country like China, it is not uncommon that previously allocated and terminated mobile numbers are reassigned to new subscribers. This is annoying to the new subscriber since she might receive calls and messages intended for the ex-owner; when it happens to our honeycards, the effect is two-fold.

On one hand, legit calls and messages intended for the ex-owner add to the pool of those that need to be identified and filtered for the purpose of analyzing unsolicited ones. For example, we identified two different numbers calling and sending messages to `nsd2` asking for the same person, which we believe to be highly likely legit calls and messages to the ex-owner. On the other hand, attackers may pretend to be the ex-owner and, e.g., request one-time passwords to be forwarded to the pretended ex-owner as an attempt to compromise the two-factor authentication system. We found 22 requests of this type via SMS, out of which 18 were for Alipay and 4 were for QQ. For example, `soc2` received the following verification message:
*[Tencent] Verification code 658339. Use it to change the password of the QQ number 64******5. Leaking the verification code has a risk. The QQ Security Center.*

Clearly, this represents a security issue for the new subscriber.

## 5.2 Sophisticated Scamming

We found a potentially scamming call that appears to be the first step of a sophisticated scam earlier reported in China[10]. In this first step of the attack, the caller pretends to be the big boss of the victim and demands that the victim visit his office the next morning. On the way to meet the "big boss", the victim will receive a second call directing her to the secretary of the big boss to settle banking accounts for commission or reimbursement issues first.

[9] http://haoma.sogou.com/

[10] http://www.chinanews.com/sh/2015/01-14/6969548.shtml and http://blog.sina.com.cn/s/blog_5d881ee30102v7xh.html

## 5.3 Attacks Specifically Targeting Mobile Users

A primary objective of the *mobile* telephony honeypot is to collect evidence of attacks targeting mobile users. For example, in a scenario reported in the previous subsection, the scam took place when the victim was on the move to meet the "big boss". In general, all attacks employing short message services are mobile victim specific. SMS enables the attacker to launch more sophisticated attacks that might not be possible against traditional landline installations. In this section, we report cases where an attacker uses both call and SMS to conduct the attack.

We found more than 10 scamming cases where the same attacker makes phone calls and at the same time sends SMS messages, with contents of the call and SMS message being about the same. As an example, `18069953481` gave on Dec. $30^{th}$ a first call with the pretext of knowing her peer. A couple of minutes after, the same source sent a message asking for QQ messaging's account details with the excuse of refunding some credit. Lately during the day, we received an additional call from the same number in which the author asked if the message was received and demanding for the messaging account, including personal name and surname.

There were also six cases where the attacker started with a casual conversation regarding some transactions, and then followed with an SMS message that gave the account numbers for bank transfers or instant messaging. Intuitively, this is to take advantage of SMS in its clarity and convenience in sharing large numbers that are hard to memorize. For example, `13691049676` performed three social-engineered calls before sending an SMS message with the information for the money deposit on Alipay.

# 6. OTHER SEEDING TECHNOLOGIES

In this section, we briefly present other seeding technologies that potentially provide incentives for fraudsters to scrape or steal phone numbers. Due to various constraints, especially the limited number of honeycards we have, we did not implement these techniques and leave them for further investigation in future work.

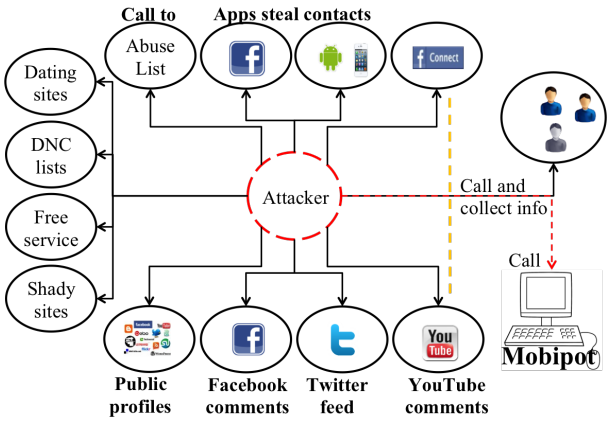



**Figure 6: Prominent seeding targets for MobiPot**

- **Public Discussion Forums:** Posting phone numbers on public forums like USENET is another way of seeding the honeycards on the web. The discussion groups in USENET are ranked by the largest number of articles read, largest number of articles posted, and largest of bytes posted, etc.
- **DNC List:** The Do Not Call lists or registry is a list of personal phone numbers which the telemarketers are prohibited from calling. However, this list is potentially another source of phone numbers for fraudsters. Therefore, some numbers can be posted on DNC lists of the country where the MobiPot is setup.
- **Questionable websites:** Websites in a few categories are more likely to be scraped by fraudsters. Such categories include those of drugs, gambling, and adult dating, porn, or sex related sites. Most of these sites are known to be spam/malware intensive and thus come under the classification of "questionable".
- **Advertisements:** Advertisement websites which may have widespread use like Craigslist is also a viable target. Researchers had showed how Nigerian scammers target Craigslist to harvest phone numbers [29]. Craigslist has numerous sections devoted to jobs, housing, community, gigs, resumes, and discussion forums. Thus, another option is posting advertisements with honeycards in such websites. However, since such sites have legitimate usage, it is important that the advertisement does not attract the legitimate users. Alternatively, phone numbers could be publicized by filling in fake details in the advertisement forms that pop up while visiting particular websites or during URL redirection.
- **Free Services:** Websites offering free services are other places from which the phone numbers are likely to be harvested. These sites may give free music, free ringtones, free or discounted coupons, etc. Some sites may offer rewards to fill out surveys, in which case it is important to distinguish between the legitimate and illegitimate survey sites.

# 7. CONCLUSIONS

In this work, we introduced the first *mobile* telephony honeypot. We implemented this mobile telephony honeypot system with honeycards in a real system called *MobiPot*. We seeded these honeycards in three distinct ways and studied the effectiveness of the seeding mechanisms in attracting previously unknown fraudsters.

Overall, we collected 1,021 SMS messages from 215 senders and 634 calls from 413 callers. We confirmed that over half of them were unsolicited. We investigated the biggest contributors, classified threats, and studied the connectivity among SMS senders. Finally, we also described a number of interesting cases that we hope will help us gain a better understanding regarding how mobile telephony threats are conducted.

**Acknowledgments:** Mustaque Ahamad's participation in this research was supported in part by US National Science Foundation award 1318167. Any opinions, findings and conclusion or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF. Additional acknowledgments go to Trend Micro's Forward-Looking Threat Research (FTR) and Mobile Research & Development teams who supported the research in different forms.

# 8. REFERENCES


[1] Data-slurping Facebook Graph Search flaw revealed. http://www.net-security.org/secworld.php?id=15147.

[2] Wangiri Fraud. http://www.xintec.com/wangiri-fraud/.

[3] P. C. S. 630-638. PENAL CODE SECTION 630-638. http://www.leginfo.ca.gov/cgi-bin/displaycode?section=pen&group=00001-01000&file=630-638.

[4] H. L. API. Number portability lookup. http://www.numberportabilitylookup.com.

[5] Baidu. In mainland China call recording crime? http://zhidao.baidu.com/question/919394718333901099.html.

[6] A. Bhatia. Decline of landline in India. http://telecomtalk.info/decline-of-landline-in-india/66093/.

[7] Z. D. Boren. There are officially more mobile devices than people in the world.

[8] U. S. C. Bureau. World Population Clock. http://www.census.gov/popclock/.

[9] A. Costin, J. Isacenkova, M. Balduzzi, A. Francillon, and D. Balzarotti. The role of phone numbers in understanding cyber-crime schemes. In *Privacy, Security and Trust (PST), 2013 Eleventh Annual International Conference on*, pages 213–220. IEEE, 2013.

[10] D. Dagon, X. Qin, G. Gu, W. Lee, J. Grizzard, J. Levine, and H. Owen. Honeystat: Local worm detection using honeypots. In *Recent Advances in Intrusion Detection*, pages 39–58. Springer, 2004.

[11] S. E. Griffin and C. C. Rackley. Vishing. In *Proceedings of the 5th Annual Conference on Information Security Curriculum Development*, InfoSecCD '08, pages 33–35, New York, NY, USA, 2008. ACM.

[12] P. Gupta, M. Ahamad, J. Curtis, V. Balasubramaniyan, and A. Bobotek. M3AAWG Telephony Honeypots: Benefits and Deployment Options. Technical report, 2014.

[13] P. Gupta, B. Srinivasan, V. Balasubramaniyan, and M. Ahamad. Phoneypot: Data-driven understanding of telephony threats. In *22nd Annual Network and Distributed System Security Symposium, NDSS 2015, San Diego, California, USA, February 8-11, 2014*. The Internet Society, 2015.

[14] K. J. Higgins. The TDos Attack. http://www.darkreading.com/attacks-breaches/hacking-the-tdos-attack/d/d-id/1139863?

[15] L. Intel, S. Penn, and U. Duke. TaintDroid. Realtime Privacy Monitoring on Smartphones. http://appanalysis.org/.

[16] G. Intelligence. Definitive data and analysis for the mobile industry. https://gsmaintelligence.com/.

[17] ITU. International Telecommunication Union (ITU). https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx.

[18] N. Jiang, Y. Jin, A. Skudlark, W.-L. Hsu, G. Jacobson, S. Prakasam, and Z.-L. Zhang. Isolating and analyzing fraud activities in a large cellular network via voice call graph analysis. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*, pages 253–266. ACM, 2012.

[19] 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. In *USENIX Security*, pages 1–16. Citeseer, 2013.

[20] A. Litan. Gartner Survey: U.S. Banks Are Improving Much Needed Online Security, but Their Phone Channels Need More Attention. https://www.gartner.com/doc/1861016/gartner-survey-banks-improving-needed.

[21] D. T. Ltd. GoIP-8 VoIP-GSM Gateway. http://www.dbltek.com/products/goip-8.html.

[22] F. Maggi. Are the con artists back? a preliminary analysis of modern phone frauds. In *Computer and Information Technology (CIT), 2010 IEEE 10th International Conference on*, pages 824–831. IEEE, 2010.

[23] MBALib. Privacy laws in China. http://wiki.mbalib.com/zh-tw/\%E9\%9A\%90\%E7\%A7\%81\%E6\%9D\%83.

[24] McAfee. Threats Report, February 2015.

[25] T. Micro. Sextortion in the far east. https://www.trendmicro.com/cloud-content/us/pdfs/security-intelligence/white-papers/wp-sextortion-in-the-far-east.pdf.

[26] M. E. Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical review E*, 74(3):036104, 2006.

[27] C. M. of Industry and I. Technology. 2014 Communications Operation Industry Statistical Bulletin. http://www.miit.gov.cn/n11293472/n11293832/n11294132/n12858447/16414615.html.

[28] C. M. of Public Security. Annual Report. http://www.npc.gov.cn/npc/xinwen/2015-03/14/content_1927536_9.htm.

[29] Y. Park, J. Jones, D. McCoy, E. Shi, and M. Jakobsson. Scambaiter: understanding targeted nigerian scams on craigslist. *system*, 1:2, 2014.

[30] N. Provos et al. A virtual honeypot framework. In *USENIX Security Symposium*, volume 173, 2004.

[31] A. Ramachandran and N. Feamster. Understanding the network-level behavior of spammers. *ACM SIGCOMM Computer Communication Review*, 36(4):291–302, 2006.

[32] Statista. Number of mobile cell phone subscribers in China from March 2014 to March 2015 (in millions) . http://www.statista.com/statistics/278204/china-mobile-users-by-month/.

[33] W. Steno. Wanbo Steno Transcription Service. http://item.taobao.com/item.htm?spm=2013.1.1998246701.4.hae4nK&scm=1007.10152.6216.1i36829088458&id=19324170391&pvid=b899b50a-1d15-42d8-9ad7-f9a0e1898b81.

[34] I. T. Union. Mobile subscriptions near the 7 billion mark. Does almost everyone have a phone?

[35] G. Zhang and S. Fischer-Hübner. Detecting near-duplicate spits in voice mailboxes using hashes. In *Information Security*, pages 152–167. Springer, 2011.

[36] Y. Zhou and X. Jiang. Dissecting android malware: Characterization and evolution. In *Security and Privacy (SP), 2012 IEEE Symposium on*, pages 95–109. IEEE, 2012.