Mobile Telephony Threats in Asia

Black Hat Asia 2017

Marco Balduzzi[*], Payas Gupta[+], Lion Gu[*], and joint work with Debin Gao and Mustaque Ahamad[-]

[*]Trend Micro
[+]Pindrop
[-]SMU and GaTech

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[1] and the International Telecommunication Union[2], mobile phone subscriptions have grown over 7% yearly in the last ten years. Since October 2014, there have been more mobile phones than people[3]. As of November 2015, the GSMA's real-time tracker sets the number of mobile devices to 7.58 billion[4], overtaking the 7.24 billion estimated world population[5]. Countries like China and India have experienced a huge growth in mobile technologies[6,7]. For example, China has over 1.2 billion active mobile phones with 93% penetration rate[8].

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[9]. Voice phishing attacks can exploit the telephony channel to lure their victims into revealing confidential information like birthday, residence, and credit card numbers[10].

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[11] or massive number of robocalls (one-ring calls)[12] 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 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.

2 System Overview and Deployment

In this paper, we propose a mobile specific honeypot system called MobiPot that differs from prior work[13] 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.

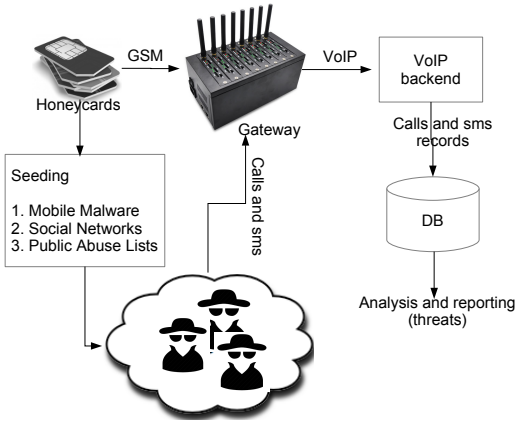


Figure 1. MobiPot Architecture

2.1 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`[14]) as

the GSM-VoIP gateway to register our GSM honeycards with the VoIP soft-switch 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.

2.1.1 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.

2.1.2 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.

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

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 publish on these sites on false pretexts, like free services or highly discounted articles[15]. However, the challenge here is that the fake profile should be popular enough to be chosen by the fraudsters.

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

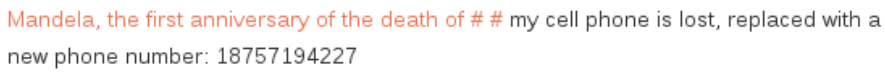


Figure 2. Social networking seeding via `Weibo`

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 2.

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[16], 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[17], 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[18]; 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[19]. 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`[5] – 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?*

3 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 messages 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

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

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.

3.1 Unsolicited Calls and SMS Messages

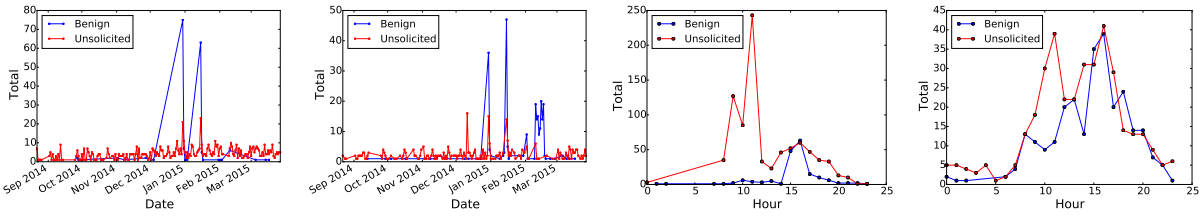
We set up MobiPot to understand the ecosystem behind unsolicited SMS messages and calls that potentially come from fraudsters who abuse the telephony channel. However, not all the calls and SMS messages received on MobiPot are unsolicited. There are multiple reasons why some of them are not. a) The calls 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`[20]. 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.

3.2 Volume and Temporal Characteristics

In this subsection, we provide insights into the temporal calls and SMS messages patterns received on MobiPot (see Figure 3). 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.



(a) Daily Volume (SMS) (b) Daily Volume (Call) (c) Hourly Volume (SMS) (d) Hourly Volume (Call)

Figure 3. Overall volume

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 3(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 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 3.6.

Figure 3(c) and 3(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.

3.3 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.

Table 2. Breakdown of SMS messages and calls before and after seeding.

(a) Based on call and message volume

Label	# of SMS messages		# of calls		# of SMS and calls combined	
	Total	Unsolicited	Total	Unsolicited	Total	Unsolicited
nsd1	209	209	12	10	221	219
nsd2	20	16	60	54	80	70
soc1	281	278	36	25	317	303
soc2	30	22	53	37	83	59
mal1	81	81	97	95	178	176
mal2	58	58	28	20	86	78
abs1	16	6	174	43	190	49
abs2	326	177	174	82	500	259
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	Total	Unsolicited	Total	Unsolicited
nsd1	5	5	10	9	15	14
nsd2	9	7	50	45	59	52
soc1	22	19	32	22	54	41
soc2	14	9	50	37	64	46
mal1	2	2	89	87	91	89
mal2	11	11	24	19	35	30
abs1	14	5	46	22	60	27
abs2	147	32	121	65	268	97
Total	215	84	413	300	583	373

The first interesting result is that there is a higher likelihood of getting an unsolicited message as compared to unsolicited call. 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 3.2) 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 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.

3.4 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 4 (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 2.1.2.

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 4 (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".

3.4.1 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 significantly higher. 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).

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[6] 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. 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 higher after seeding as compared to that before seeding.

3.4.2 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 4(b), 4(e), 4(h) and 4(k). In Table 2(a), we use the first seeding date as a reference point to calculate the statistical significance.

There is a sudden rise in the number of unsolicited calls for mal1 (see Figure 4(b)) and unsolicited messages for mal2 (see Figure4(e)) after the second round of seeding. As discussed in Section 2.1.2, 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 4(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.

3.4.3 Abuse on abs1 and abs2

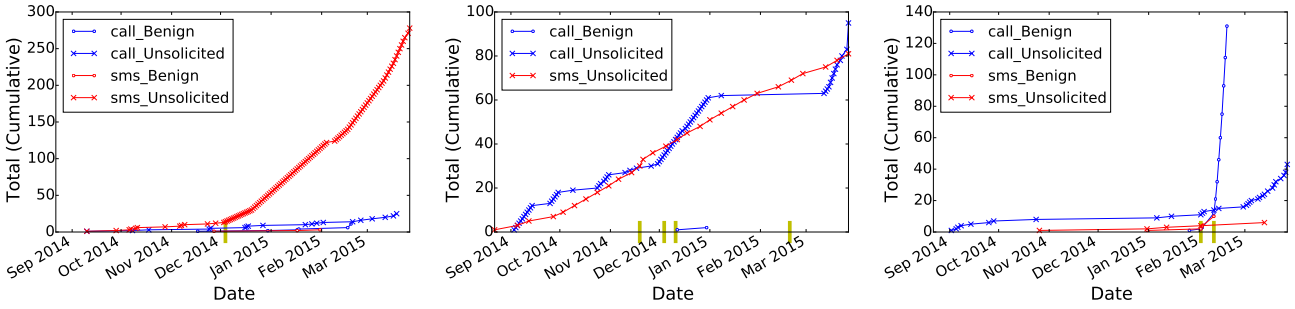
The third and final seeding exercise was performed using abs1 and abs2, where these two honeycards were seeded directly by contacting the known abuse phone numbers. Figure 4(c) and 4(f) show the total (cumulative) call and SMS message volume before and after seeding dates. Note that unlike other seeded 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.

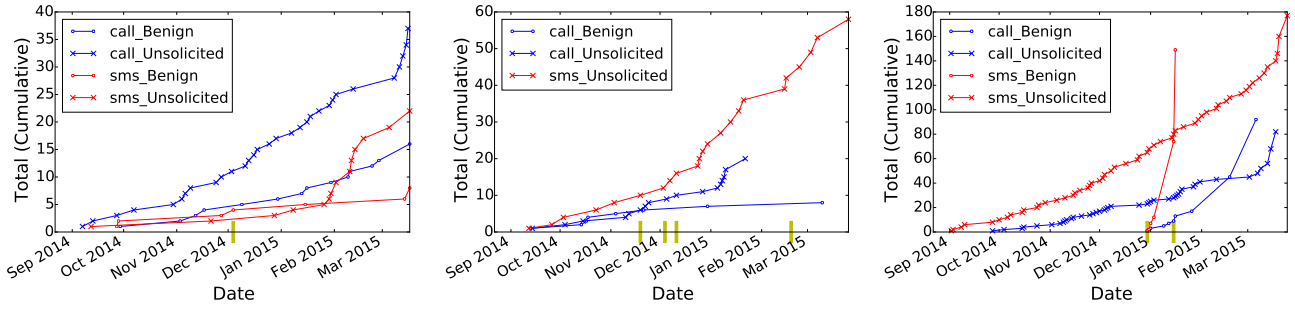
Immediately after seeding on both dates, there were many fraudulent transaction messages received by abs1 and abs2. Some examples of these messages (translated in English and masked for privacy) are

- *ICBC: 62122640000XXXXXXXX; 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: 62284801208XXXXXXXX; Beneficiary: Lu Yudan; Longgang, Pingshan Branch, Shenzhen Branch*

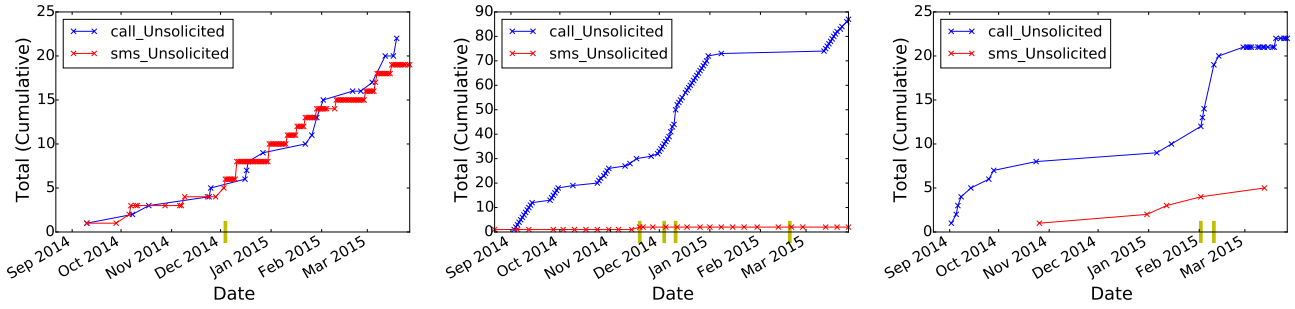
[6] http://www.xhqm.cn/



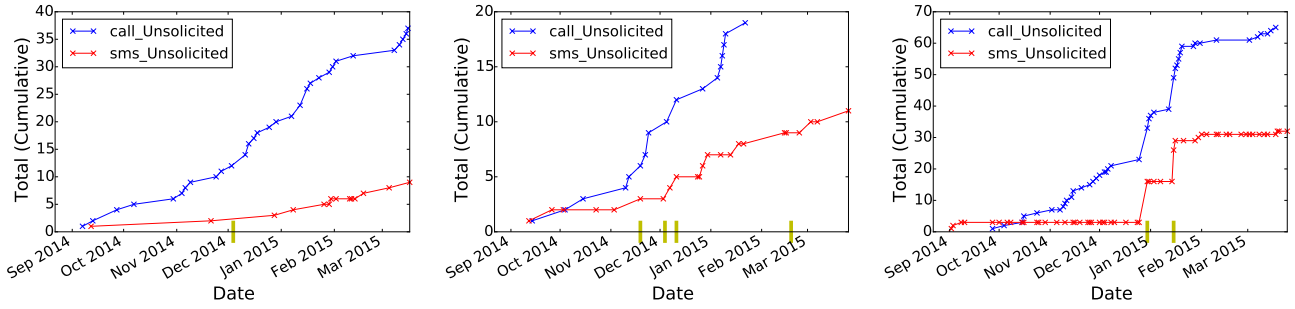
(a) Volume - `soc1` (b) Volume - `mal1` (c) Volume - `abs1`



(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 4. (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`.

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 4(c) and 4(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.

3.5 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.

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.

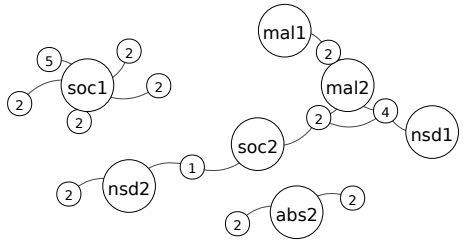


Figure 5. Social graph of campaigns and honeycards based on content of the messages

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 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.

3.6 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`[7], a popular mobile application that informs the user whether

[7] `http://haoma.sogou.com/`

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.

4 Mobile Threats

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].

4.1 Fraud

We identified 48 frauds targeting all MobiPot's numbers, with a prevalence on the numbers that we leaked through the social networks and the abuse lists experiments.

All telephony frauds were *manually* initiated by frauders who used social engineering as primary form of tricking us into performing actions like money transfers. All calls originated from counterfeit identities like fake banks, no-profit associations or old friends. We often received multiple calls from the same frauder, asking for the status of the payment process.

Some frauders pretended to be our telephony provider (e.g. China Mobile) and informed us that our contract was due to suspend because the bill was not paid. In a second version, they leveraged the introduction of the new 4G technology to offer us a fake upgrade for 399 RMB.

In another example, the frauders impersonalized postal services like ChinaPost EMS[8] and express mail services like SF-Express[9] claiming that there was a package for us; we were requested to pay a fee to have it delivered, because in customs hold.

With the excuse of being lucky winners of (fake) lotteries, others offered us free memberships or prizes upon registering with their "special" membership programs. Further investigations revealed that these programs had regular monthly fees and lasted for a year, at last.

As said, all these examples were manually conducted by frauders over mobile telephone, with the goal of money extortion.

4.2 Malicious Sites

We processed the URLs embedded in the received content, and confirmed two cases of malicious websites: `478808.com` and `240044.com`.

The first one was advertised through an automated call as a profitable casino located in Macao. The access to the casino's website was protected with authentication and only granted to its members. The call invited us into subscribing for 39 RMB/day by pressing the keypad's number 1. Because of the limitation of our current deployment, apart from playing the engaging message, we did not actively interacted with the caller – e.g., by typing the requested number. We manually called back the sourcing number, without success. We leave this as future work.

In the second case, after we seeded *malw*2 via a mobile application dedicated to gambling, and probably hijacked with malware, we received via sms the malicious URL. With further investigation, we confirmed this website being a shady online casino[10].

4.3 Scam

We collected 28 scam related content (i.e., 11 sms and 17 calls). Unlike frauds – earlier discussed –, the scam messages (and calls) we collected were delivered in automated fashion, similarly to spam, and did not actively engage their victims. On the other hand, similarly to frauds, the actors behind the scams were intended in collecting money, likely from ingenuous users.

Some examples of messages were:

Do a good job and send your money to this bank's account <IBAN>. Ye Jianling and *We finally finalized it! We agree with the first paragraph of the contract. 20 thousand is fair. Our secretary will take care of it. <IBAN>. Good to call me*.

With respect to the calls, a counterfeit bank called Beijing Rongxin Guarantee promoted a low-interest investment by opening a `10,000 USD` account with them. In another case, 14 calls were related to investing in gold and silver; a shady broker advertised guarantee earnings and gave a IBAN for the deposit.

[8] `http://www.ems.com.cn/english.html`
[9] `http://www.sf-express.com/cn/en/`
[10] Offline, at the time of writing

Interestingly, these scammers made used of Alipay as additional form of cashing out to bank transfers. We collected 5 Alipay accounts associated with scammers.

4.4 Social Engineering

Social engineering is known to be used to lure a victim into performing actions of interest for the attackers, such as revealing personal information or transferring money – commonly associated to attacks performed via telephone.

In our analysis, we classified as social engineering those attacks manually initiated by a person, but for which it was not possible to identify the intention behind the attack. For example, because the attacker hanged up the call beforehand or the call lacked of context.

In particular, we collected 2 sms and 2 calls. The sms originated from the same person asking whether we received the money she sent us. For the calls, both actors pretended of knowing us (old friends) and asked whether we can meet to discuss an unfinished business.

4.5 Spam

Spam content accounted for the majority of junk messages spam calls, e.g. from telemarketers, mobile plans promotions, shop advertisements, university programs, job offers, escort services, online games and quizzes, and any sort of commercial and ads.

We collected a multitude of self-assessment tests offered over mobile telephony. In the following, for example, our peer pretended to guess our ability as entrepreneur by analyzing our answers to questions like: which animal we prefer to reborn to, what is our favorite landscape or popsicle's color, e.g.: *What place? A view of a beautiful mountain, B summer warm place, C a place to relax, D games can be fun places. Reply with letter or cancel receiving this information with QX3*.

Other content was related to propagandas of political groups like in the following call: *I wish you a New Year of health and peace. I called to tell you that the Chinese disasters continued. How we will be able to not spend money? [...] Love to the Chinese Communist Party. In our program, we want to reform the land [...]*

Sms were also used for trading illicit goods. `15692600442` traded credit cards and hijacked payments accounts (i.e., paypal) with verified balances up to 5000 RMB (about 805 USD). `13846477853` and `15501623869` offered authentic (filled and signed) invoices in different amounts and formats like construction, technology, advertising and services. Interestingly, in these cases, the criminals provided different contact numbers to receive the orders – a sign of organizations made of different actors and structured networks (e.g., multiple simcards).

We also discovered official (i.e., verified) applications leaking personal data like the contact list. One of them, a messaging platform similar to `WhatsApp` and `Viber`[11], exploited the collected numbers to publicize itself with the following message: *You have a network of friends already using <name>. Download <name> and connect with them. <url>*.

Finally, we recorded private investigators offering shadowing and surveillance services, and others providing hacking services like accessing personal emails and spying on mobile telephones' users.

4.6 Re-using Mobile Numbers

In a big country like China, it is not uncommon that previously allocated and terminated mobile numbers are re-assigned 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.

4.7 Sophisticated Scamming

We found a potentially scamming call that appears to be the first step of a sophisticated scam earlier reported in China[12]. 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.

[11] We voluntarily omit its name. We are open to share the name with the reviewers, if needed

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

4.8 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.

5 Conclusions

In this work, we introduced the first *mobile* telephony honeypot that we then implemented 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.

References

1. ITU. International Telecommunication Union (ITU). https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx.
2. Union, I. T. Mobile subscriptions near the 7 billion mark. Does almost everyone have a phone?
3. Boren, Z. D. There are officially more mobile devices than people in the world.
4. Intelligence, G. Definitive data and analysis for the mobile industry. https://gsmaintelligence.com/.
5. Bureau, U. S. C. World Population Clock. http://www.census.gov/popclock/.
6. of Industry, C. M. & Technology, I. 2014 Communications Operation Industry Statistical Bulletin. http://www.miit.gov.cn/n11293472/n11293832/n11294132/n12858447/16414615.html.
7. Bhatia, A. Decline of landline in India. http://telecomtalk.info/decline-of-landline-in-india/66093/.
8. 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/.
9. Litan, A. 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.
10. Maggi, F. Are the con artists back? a preliminary analysis of modern phone frauds. In *Computer and Information Technology (CIT), 2010 IEEE 10th International Conference on*, 824–831 (IEEE, 2010).
11. Higgins, K. J. The TDos Attack. http://www.darkreading.com/attacks-breaches/hacking-the-tdos-attack/d/d-id/1139863?
12. Wangiri Fraud. http://www.xintec.com/wangiri-fraud/.
13. Gupta, P., Srinivasan, B., Balasubramaniyan, V. & Ahamad, M. 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). URL http://www.internetsociety.org/doc/phoneypot-data-driven-understanding-telephony-threats.

14. Ltd., D. T. GoIP-8 VoIP-GSM Gateway. http://www.dbltek.com/products/goip-8.html.

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

16. McAfee. Threats Report (2015).

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

18. Intel, L., Penn, S. & Duke, U. TaintDroid. Realtime Privacy Monitoring on Smartphones. http://appanalysis.org/.

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

20. Steno, W. 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.