Are the Con Artists Back?
A Preliminary Analysis of Modern Phone Frauds

Federico Maggi
Dipartimento di Elettronica e Informazione
Politecnico di Milano
Milano, Italy
fmaggi@elet.polimi.it

Abstract—**Phishing is the practice of eliciting a person's confidential information such as name, date of birth or credit card details. Typically, the phishers use simple technologies (e.g., e-mailing) to spread social engineering attacks with the goal of persuading a large amount of victims into voluntarily disclose sensitive data. Phishing based on e-mail and web technologies is certainly the most popular form. It has indeed received ample attention and some mitigation measures have been implemented.**

In this paper we describe our study on *vishing* (voice phishing), a form of phishing where the scammers exploit the phone channel to ask for sensitive information, rather than sending e-mails and cloning trustworthy websites. In some sense, the traditional *a-lá-Mitnick* phone scams are streamlined by attackers using techniques that are typical of modern, e-mail-based phishing. We detail our analysis of an embryonic, real-world database of vishing attacks reported by victims through a publicly-available web application that we build for this purpose. The vishing activity that we registered in our preliminary analysis is targeted against the U.S. customers. According to our samples, we analyzed to what extent the criminals rely on automated responders to streamline the vishing campaigns. In addition, we analyzed the content of the conversations and found that words such as "credit", "press" (a key) or "account" are fairly popular. In addition, we describe the data collection infrastructure and motivate why gathering data about vishing is more difficult than for regular e-mail phishing.

Keywords-**phishing, phone phishing, social engineering, measurements.**

I. INTRODUCTION

The term *phishing* was coined in 1996, when attackers used to refer a compromised account to as *phish* [1]. At the time, malicious people —nowadays known as *phishers*— used to trade phishes as a form of electronic currency. This practice has now evolved into a widespread, serious Internet threat. For instance, a recent advisory published by the Anti Phishing Working Group estimates that, in 2009, 30,131 domain names and 3,563 unique IPs were used for phishing purposes [2]. Normally, phishing targets very valuable data. Unsurprisingly, Symantec estimated that financial information is the phishers' favorite target in 79% of the campaigns [3] (between 2008 and 2009).

Typically, phishing involves some kind of social engineering. More precisely, a phisher sends large amounts of luring e-mails that appear as if they were sent from a trusted party (e.g., a bank or a large, popular website such as PayPal or eBay). These e-mails usually contain a link to a clone of the trusted website. Fooled by the familiar look and feel of the page, the victim feels comfortable at submitting sensitive data (e.g., username and password or PIN). Unfortunately, the cloned website is controlled by the attacker who has previously deployed a computer program that automatically collects the phishes for later use. The most valuable phishes may be passport numbers, credentials, e-mail accounts, or financial information.

The consequences and the financial losses are further exacerbated by the enormous scale reached by malicious activity on the Internet. In fact, the cyber criminal organizations control vast amounts of computational and networking resources. Most notably, by spreading malicious software that automatically infects unprotected computers, the criminals have learned to build networks of remotely-managed, compromised hosts, also known as botnets [4]. These resources are nowadays very accessible thanks to the recent pay-per-use business model adopted by criminals [5], which allows people with no or little technical skills to run phishing campaigns and other forms of scam at very convenient prices. For example, a scam hosting solution ranges between $3 and $40 per week according to the latest Symantec report [6]). Interestingly, as of March 2009, "phishing tutorial" is the 8th most popular Google suggestion for "phishing". The criminals immediately recognized that phishing works very well [7] and is indeed a very profitable "business": according to a recent Gartner survey, more than 5 millions U.S. customers lost money due to phishing in the 2008 [8], a 40 percent increase in one year [9].

During the past five years the blackhats have begun to leveraging alternative spreading channels such as instant messaging [10], [11], forums, blogs and even text messages [12], [13]. Interestingly, the criminals are also resorting to the telephone channel to achieve the same objectives of traditional e-mail phishing. This practice is referred to

as *vishing* (voice phishing) [14]. More precisely, vishing is the activity of systematically defrauding account holders using social engineering over the telephone system. Real-world facts from the past have extensively proven how the timing of the message can be exploited with success in a live telephone call [15]. Indeed, a live (telephone) conversation enhances the effectiveness of social engineering techniques significantly. This does not happen normally with e-mailing because the e-mails have to be read and thus leave less chances to the attackers to lure the victims.

Similarly to phishing, the goal of vishing is to eliciting a person's confidential information. However, vishing is inherently more difficult to analyze as opposed to traditional, e-mail phishing. In fact, collecting e-mails suspected of phishing is relatively easy. For instance, from a purely technical perspective, e-mails can be intercepted, dispatched, filtered, stored and so forth. On the other hand, gathering evidence of phone-based phishing is difficult. First of all, while the content of e-mails is stored on servers and clients, the nodes of the telephone system do not store the *content* of the calls on a regular basis. Second, while e-mails are made of strings and, as such, can be scanned and parsed, phone calls cannot be analyzed so easily. Note that, although a phone conversation may be translated into text using automatic programs, this process may introduce inaccuracies that would make it difficult to extract significant features (especially in the case of phone calls between live humans, which voices are less predictable).

Unfortunately, the countermeasures against this emerging threat are limited to one approach proposed to mitigate the malicious SMS received by customers [16]. This work is interesting because it adopts an automated technique to profile the users' normal behavior. However, to our knowledge, no thorough analyses of *voice* phishing have been done. The first motivation is that social engineering attacks are, by their nature, difficult if not impossible to analyze. Second, the awareness about this phenomenon is limited. Therefore, we believe that a solid understanding of vishing is important. Given the above observations, the objective of this paper is to analyze real-world vishing activity.

The main contribution of this paper is, in summary, the analysis of social engineering attacks carried over the phone. More precisely:

- we collected *structured* data about vishing from the real-world.
- To this end, we selected a relatively small set of interesting attributes (e.g., the calling identifier, transcribed conversation, subject, nationality, language) useful to infer the typical characteristics of a vishing attack and the vishers' modus operandi.
- The secondary contribution of this paper is the first publicly-available service for collecting user-submitted vishing reports: this service is briefly described in Section III-A.

We think that the preliminary work described in this paper is the first effort to understanding the techniques adopted by the vishers, toward promoting better user-awareness and, most importantly, to devising simple countermeasures.

II. PHONE FRAUDS: BACK FROM THE PAST

In one of his books, *"The Art of Deception: Controlling the Human Element of Security"* [15], Mitnick transcribed and described in details many live telephone calls he made to steal confidential information. He resorted to his strong social engineering skills to convince the interlocutor to reveal sensitive data such as a secret phone number or a numeric code, etc. At the time, this information was very valuable, since the telephone lines were used to access computer systems such as mainframes. Thus, people relied on this subtle form of attacks to gain initial access to the target organization's network.

Ten years later, the spread of the *World-Wide Web* (WWW), e-mail, and, later, e-commerce, promoted the use of computers to perform several task, even financial transactions. As a consequence, usernames, passwords, and credit card numbers became immediately a very valuable piece of information. At the time, people were familiar with e-mails for communicating. Thus, by leveraging simple social engineering strategies, the attackers began to write templates of e-mails —that masquerade themselves as trust-worthy entities— to elicit the recipients' confidential data for fraudulent purposes (e.g., to make illegal transactions). As opposed to live phone calls, deceptive e-mails need not very sophisticated social engineering techniques, since users have learned very quickly to follow hyperlinks paying little or no attention to the consequences and, in addition, are not capable of distinguishing real web-pages from cloned ones [7].

Nowadays, online chats and online communities (notably, the first was AOL, then MySpace, and now FaceBook) are very widespread. Very soon, attackers started to exploit such live communication channels to maneuver the users [10], [11] with persuasive, written conversations. Apparently, their goal is to reach more than just regular e-mail users (e.g., mobile device consumers such as young individuals). Hence, phishing techniques become more sophisticated and share similarities with both old-fashioned and e-mail phishing. More precisely, the exploiting of instant messaging (e.g., Windows Live Messenger, Skype, the FaceBook chat) is similar to voice phishing as it involves *live* a conversation with a human (or a computer program that leverages natural language processing to mimic a real person). On the other hand, the use of modern Internet-based tools allow the criminals to reach a large share of victims using mass messaging. This makes new phishing techniques also similar to e-mail phishing, which relies on mass mailing to spread deceptive links.

Surprisingly, the last step involves an old technology: the telephone system. Recently, the cyber criminals have indeed rediscovered the phone channel to reach their victims [14] but, interestingly, there are differences and similarities with the past. More precisely, the mechanisms utilized by the cyber criminals have a wider spread, yet the technique is still basically the same. In particular, our experience (described in details in Section IV) suggests that the phishers rely on automated responders, and not only on human operators, with the objective of streamlining the dissemination. On the other hand, traditional *a-lá-Mitnick*, phone scams mostly leveraged social engineering and, thus, were relatively slow. To increase the speed, modern phone scammers have partially abandoned social engineering and, instead, take advantage of the customers' familiarity with "new technologies" such as Internet-based telephony, text-messages [12], and automated telephone services. It is interesting to note that, in the case of e-mail phishing, such "new technologies" were e-mails, which became so popular and convenient that attackers decided to use them instead of the phone to reach as many victims as possible. In a similar vein, the widespread of automated telephone services and call centers has made the customers more used to provide information to strangers (or machines) who just ask for it.

In addition to the aforementioned reasons, the fast spread of *Voice over IP* (VoIP) telephony has been having an important role in the raising of vishing and other threats that target voice communication [5]. First, because of the significant reduction of call rates. As a matter of fact, VoIP calls are just slightly more expensive than e-mails yet, certainly, much more effective. Second, VoIP makes it more difficult to track the attackers, since clients do not terminate in a well-known location in the physical world (e.g., a house, a building) as opposed to traditional telephony. Third, VoIP is not a secure protocol [17], and criminals can certainly take advantage of its vulnerabilities for obtaining the resources to run vishing campaigns and masquerade their activity by spoofing the calling identifiers. In addition, as noticed in [14], the telephone channel allows to reach a share of victims that is not typically reachable using e-mail (e.g., the elderly).

III. OVERVIEW OF THE COLLECTION SYSTEM

The objectives of this work can be summarized as follows. First, we want to determine to what extent the telephone system is leveraged for perpetrating phishing. Second, we want to understand whether the blackhats rely on automated mechanisms to streamline this process. Third, we focus on extracting interesting words from the transcribed conversations.

To achieve these objectives, the first phase focuses on collecting high-quality data, suitable for analysis. Our proposal to this phase is described in Section III-A. The second phase, is to define a set of simple attributes that characterize

a typical voice phishing attack. Note that, such attributes may be insufficient for defining detection techniques or countermeasures against vishing. However, at this stage of the work, we focus only on the analysis phase, thus the development of countermeasures is out of the scope of this paper.

A. Collecting vishing data

Research on e-mail phishing can rely on large data-collection infrastructures [18]. Moreover, e-mail service providers and modern client software incorporate anti-spam and anti-phishing mechanisms that also collect information on e-mails deemed as suspicious. This information can be easily shared into worldwide databases used to understand the scammers' modus operandi or to improve the protection mechanisms.

On the other hand, collecting data on vishing is more difficult. For this reason, we designed and implemented a very simple yet useful infrastructure to collect reports of vishing attempts. Our system is inspired by PhishTank [19], a web application to collect full dumps of suspicious e-mails submitted by users. In addition to the submission phase, PhishTank also analyzes the data and calculates aggregated indicators and shows trends. For example, statistics on daily verified phishes are available along with the total number of active, online phishing sites. Thanks to an API, this data feed is used to provide blacklist-based protection tools such as PhishTank SiteChecker [20], a Mozilla Firefox extension that automatically block the display of pages known to be malicious.

Similarly to PhishTank, a beta version of our system is publicly available as web application [21] where anyone can send anonymous reports of phishing received via telephone. The goals of PhonePhishing.info are very similar to those of PhishTank: we are aimed at collecting data about phone phishing attacks, including (but not limited to) the calling identifier, the subject of the call, the date and time, the country, to analyze the conversations and calculate useful statistics.

There are other resources for reporting suspicious calling identifiers. Notably, CallerComplaints.com is the largest one and stored about 500,000 complaints so far. The calling identifiers are grouped by area code and ranked by popularity, i.e., number of reports. Another example is 800notes.com, which keep records of calling identifiers that are unknown to official listings. These services are fairly popular and thus have been fed with a quite large amount of data. However, they are not without drawbacks. For instance, 800notes.com only keeps numbers along with an optional message (typically utilized by the users to report their complaints). Unfortunately, by itself, a free-text field is not sufficient for analyzing vishing thoroughly. Although CallerComplaints.com collects more details about the suspicious calls reported, some relevant aspects are ignored, namely the date and time,

the country which the call is received in, the language spoken by the caller, and whether or not the call is automated. We believe that such details are important, since more accurate conclusions can be drawn about the vishers' strategies. Last, both the aforementioned collection services focus on generic calls, not on vishing calls. In fact, CallerComplaints.com accepts four different values to characterize the type of call: telemarketers, debt collectors, pranks, political calls. Another common drawback is that the user is considered trusted and this issue affects all the system that collects user-contributed content and no effective mitigation exist. However, to minimize the aforementioned problems all the data submitted to PhonePhishing.info is manually reviewed before approval. This prevents non-relevant reports to bias the analysis (e.g., reports that, according to the submitted information, qualify as "annoying calls" are discarded). At this stage, the manual revision process follows a conservative approach, that is, every report is discarded by default unless striking signs of being a vishing are present. Accepted reports are along the line of the following:

"your ATM card from [...] has been suspended, push 1 to connect to our security"

or, for instance:

"[...] asking to press one in order to receive a prize [...] speaking in bad English [...] asks for credit card info"

Differently from existing services, PhonePhishing.info focuses only on scams rather than on calls that may certainly be annoying yet pose no real threats. In particular, we collect the following details.

- **Phone number:** we validate the numbers' format so to avoid the submission of reports regarding calls with blocked or unknown calling identifiers.
- **Conversation:** the user must transcribe the conversation held during the call.
- **Subject:** this can take only one of the following values: "bank account", "car insurance", "car warranty", "health insurance", "credit card", "other". These options help at identifying only those callers that typically attempt to elicit a user's confidential information. It is important to underline that, during the manual revision process, we pay particular attention to those calls which subject is "other". If possible, the contributor is contacted via e-mail for clarifications and, in any case, we attempt to find the correct classification for the call. If none is found, the report is rejected.
- **Country:** this is the country the phone call is received in. This is useful to identify the most targeted countries.
- **Human?** This Boolean attribute helps to distinguish fully-automated calls from those that also involve live humans.
- **Spoken language:** this is simply the language spoken by the caller (or the automated voice).

The PhonePhishing.info project is at an early stage of development and no API is provided, yet. However, we have recently begun to publish a feed of suspicious numbers through a Twitter stream[1].

B. The human factor

The PhonePhishing.info website collected about 360 phishing reports over one year. At a first glance, this may appear a limited amount of data. However, it must be taken into account that people typically do not tend to voluntarily provide information unless such task requires little or no effort (e.g., push a button when a suspicious call is received and have the system automatically retrieve and send out the relevant information for analysis). To this end, as detailed in Section VII, an extra effort is being undertaken to automatize the submission process as much as possible by leveraging the call history details already available on smart-phones rather than having the users to type them in.

IV. ANALYSIS

In this section, we describe the preliminary analysis we run on the dataset collected through PhonePhishing.info between late 2008 and late 2009. Our analysis focuses on three aspects. First, we analyze the popularity of certain calling identifiers and, in some cases, also correlate this information to the reports available on other websites that collected related data. Second, we compare the calls that have been reported as coming from automated responders with those that involved a conversation with a human operator. Third, we analyze the popularity of the terms extracted from the transcribed conversations.

A. Popular calling identifiers

As shown in Figure 1, most of the reports refer to unique numbers. However, two numbers appear to be very popular. More precisely, `8007540961` has been reported several times also on other websites[2]. The caller claimed to be from a telephone provider located in the United States. Our reports confirm that the operator asked for the victim's account number, password and social security number. The second most popular number is `2024597122` (from District of Columbia), which seems to be related to a dangerous credit card scam. In particular, according to our reports, the caller attempted to elicit the victim's credit card number along with its expiration date. Several debt collection companies advertise themselves through the phone in the United States and consumers are also protected by laws [22], since this phenomena became very popular. However, according to the reports that we and other websites received, these callers are deliberately attempting to steal the victim's credit card number and expiration date rather than asking questions regarding their debts or less valuable, personal information.

---

[1]http://twitter.com/vishing
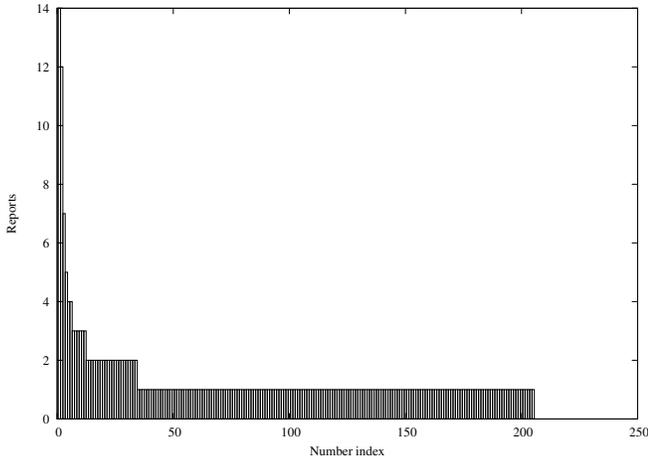[2]e.g., http://whocallsme.com/Phone-Number.aspx/8007540961, http://800notes.com/Phone.aspx/1-800-754-0961

Figure 1. Distribution of calling identifiers. Instead of reporting the real identifiers, we report indexes that correspond to unique numbers.

In addition, we analyzed the popularity of the prefixes (i.e., area codes) with focus on the United States. Such prefixes are useful to characterize what type of numbers the vishers have registered. As shown in Figure 2, the most popular numbers are toll-free numbers (i.e., 800, 866, 877 [23]). Many online companies provide VoIP numbers (also toll-free) and anyone can register them very easily (e.g., a dummy name and an anonymous, pre-paid credit card is the only information required for immediate activation) and at very low rates. For instance, Google Voice is free of charge. The second most used numbers have 202 (from District of Columbia), 289, 876 prefixes.

Last, we measured the vishing activity of each state in the United States. As shown in Figure 3 and Table show, the states of California and Florida are the most active ones, with 12.56% and 9.04% of the reports, respectively.

B. Use of automated responders

To infer the use of automated responders we use the following simple check. The report must have the "human?" flag set to "False" or the transcribed message must contain signs that the caller used a recorded message or an automated responder. To this end, we resort to a simple heuristic. In particular, while reviewing the reports, we noticed that some words were used to describe calls with automated responders. Representative examples of such words are "robo", "automated", "machine", "recorded". According to this criterion, we identified that about one half of the reports referred to automated calls. It is interesting to note that while prefixes such as 202 (from District of Columbia), which are fairly popular according to Figure 2, were reported only three times as using automated responders. On the other hand, prefixes associated to toll-free numbers are the most popular among the automated calls.

Last, we measured the number of calls that relied on both automated responders and human operators. To this end, we took into account those reports having (1) the "human?"
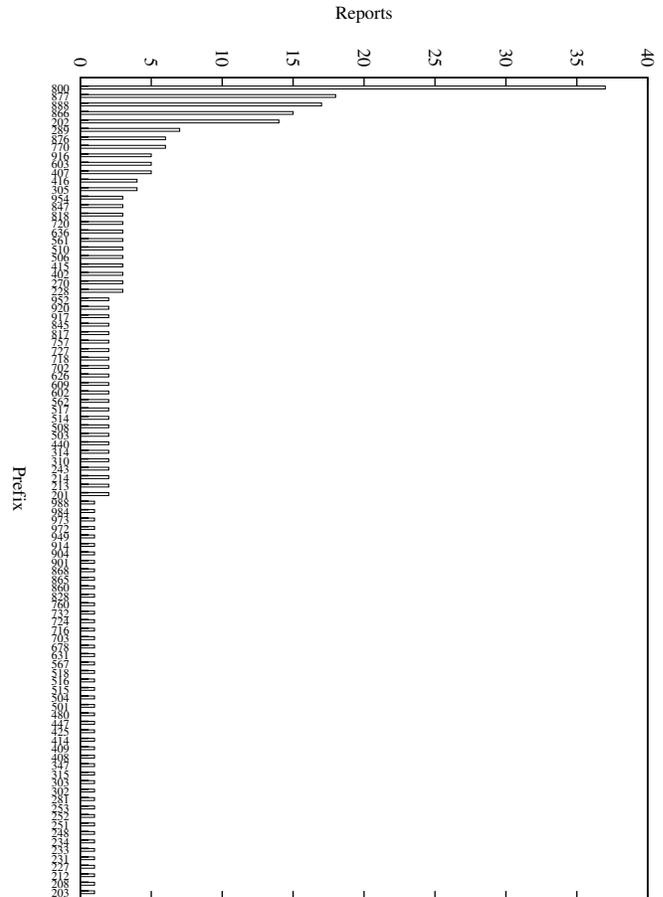

Figure 2. Distribution of prefixes.

flag set to "True" and (2) at least one occurrence of the aforementioned terms (e.g., "robo", "machine"). Among the 7 reports found, only 1 out of 7 unique numbers has a toll-free prefix. The remainder of the numbers are located in Florida, Texas, California, District of Columbia, and New Brunswick.

It is important to underline that this analysis may be biased because, typically, customers immediately hung up the phone, thus, the call is never transferred to a human operator.

C. Analysis of terms

In this analysis we focus on the terms contained in the transcribed conversations. First, we partition the reports into two groups: those with live human callers and those that rely on automated responders. To this end, we apply the simple criterion described previously. Second, we extract all the words contained in the transcribed conversations from the two groups of reports. Stemming and simple filters are applied to remove negligible tokens. In particular, we first applied the stemming algorithm described in [24] and then removed common, non-relevant words. Notable examples of such (stemmed) words are "phone", "about", "call", "hello", "minute", etc. The resulting ranking of the 15 most popular

Figure 3. Distribution of popular prefixes on the United States territory (non-toll-free numbers only). Darker areas indicate high frequency while lighter areas indicate low frequency. No reports are associated with white areas.

terms is reported in Table II.

In both the groups, the most popular term is "number". This turns out to be negligible because, in almost every transcribed conversation, the submitters refer to "the number" they received the call from. It is worth to mention that the word "press" has higher ranking in the auto-responders group than it has on human operators. This suggests that calls made with auto-responders typically ask the victim to press some keys on the phone keyboard (e.g., to be transferred to someone else).

On the other hand, it is interesting to note that calls made with auto-responders often refer to "accounts" (the second most popular term), while this term is not found in any of the calls by human operators. This gives an indication that most of the calls made with auto-responders attempt to elicit some account's credentials.

The term "credit" has high ranking on both the types of calls. This suggests that most of the vishing activity focuses on credit cards. This is confirmed also by the fact that, except for the word "number", "credit" is the most popular term overall according to our measurements.

V. LIMITATIONS

The PhonePhishing.info project has some limitations.

First, forged reports can be submitted, since the users are trusted. To alleviate this problem we adopted two measures. As detailed in Section III-B, we manually review and moderate every submission. This allows to avoid non-relevant content to be analyzed. In addition, to avoid mass submissions (and denial of service) through automated tools, we force every user to pass a CAPTCHA[3] test while submitting a report.

On one hand, the manual review ensures good quality of the reports. On the other hand, this causes overhead and requires a human moderator. Hence, the second limitation of our system is its scalability. Unfortunately, to solve this issue we have to trade-off between quality and scalability. Although heuristics may be designed to filter non-relevant reports, these could be easily circumvented. For this reason, we opted for a manual review process.

Last, while the effectiveness of e-mail phishing can be estimated (e.g., by analyzing data and logs taken from the websites used by phishers), it is nearly impossible to determine whether or not a vishing attempt succeeded. Clearly, this problem is very difficult to overcome. In fact,

[3] http://recaptcha.net

%-POPULARITY	STATE
12.56	California
9.04	Florida
7.53	New
7.53	District
7.03	non-US
5.52	Ontario
3.51	Texas
3.51	Georgia
3.01	New
3.01	Jamaica
2.51	Ohio
2.51	New
2.51	Missouri
2.01	Nebraska

Table I
GEOGRAPHICAL DISTRIBUTION OF THE PREFIXES (15 MOST POPULAR LOCATIONS).

AUTO-RESPONDERS			HUMAN OPERATORS
	%-rank		
number	22.39	19.81	number
account	11.96	13.83	credit
credit	8.68	7.54	person
press	7.52	7.23	interest
inform	5.79	6.28	claim
record	5.21	5.97	lower
debit	4.44	5.66	inform
servic	4.05	4.71	answer
expir	3.66	4.40	address
enter	3.08	3.77	servic
regard	2.89	3.45	bureau
activ	2.70	3.14	state
person	2.50	2.83	offer
autom	2.31	2.51	**press**
secur	2.12	2.20	later

Table II
RANKING OF THE 15 MOST POPULAR (STEMMED) TERMS FOUND IN THE TRANSCRIBED CONVERSATIONS.

while it is relatively easy to find the sites used for a phishing campaign (e.g., by extracting links from suspicious e-mails or by scanning for known phishing kits [25]) and to track stolen credentials, the same task cannot be performed on phone calls without wiretapping.

VI. RELATED WORK

This work is related to vishing, social engineering and, partly, phishing and cyber underground economy. Due to limited space, this section provides just a few references to the most recent and relevant works in these areas.

The first detailed description of the vishing phenomenon appeared in [14]. The author provides brief, clear definitions of the emerging *"*-ishing"* practices (e.g., smishing, vishing) and points out the characteristics of the vishing attack vectors. In addition, the report shows a few scripts of the typical ploys used by the scammers. Although this work focuses on voice phishing, it is worth to mention SMS Watchdog [16], a recent effort toward the automatic detection of smishing activity, i.e., scams through the short text-message system. As we mentioned early in this work, vishing is inherently more difficult to analyze and mitigate with respect to e-mail-based or text-based phishing for which effective countermeasures exist. However, this paper concentrates on analysis and not on protection mechanisms, thus we refer the interested reader to [26], [27] for a in-depth comparison of anti-phishing techniques and tools.

While the social engineering component of phishing websites and e-mails has been thoroughly analyzed, for instance in [7], no similar measurements have been done for vishing, yet. Moreover, an analysis of the traditional phishers' modus operandi has been published in [28], where two large repositories of phishing e-mails are analyzed. Interestingly, the authors were able to identify the types of hosting preferred by the scammers and also the practice of exploiting URL-shortening services to masquerade the malicious URLs and bypass filters. A quantitative analysis of e-mail-based phishing has been done in [25], which is the first systematic evaluation to determine to what extent the modern criminals exploit automated mechanisms (e.g., phishing kits, automated exploits) for streamlining phishing attacks to increase their profit in the new underground economy business model.

The role of phishing in the current cyber underground economy is discussed in [29], where the phishing campaign run by one criminal organization is analyzed in depth. In addition, the authors have estimated empirically the lifetime of a phishing campaign alongside with the techniques adopted by criminals to extend this time-frame. This once again supports the intuition that modern cyber criminals are well-organized and profit-driven, as opposed to old ones, which were more concerned with their reputation. This new scenario was discussed in [30], and in [31] from a purely economic perspective.

VII. CONCLUSION

Traditional, *a-lá-Mitnick*, phone scams mostly leveraged social engineering. Instead, modern voice phishing take advantage of the customers' familiarity with "new technologies" such as Internet-based telephony, text-messages, and automated telephone services. In addition, modern phone scammers, i.e., vishers, have learned to streamline their phishing campaigns in a very effective way.

In this paper we analyzed the vishing phenomenon on a collection of detailed reports submitted by the victims through the PhonePhishing.info website, a public service that we built for the purpose of gathering data about vishing. In our measurements, we observed that vishing is popular in the U.S., with particular focus on some states. Also, we have observed that a good share of the vishers that we recorded resort to automated responders. Last, we have identified some recurring, popular terms such as "press", "credit", "account", that are more frequent on automated calls with respect to calls made by human operators. This paper is the first analysis of the modus operandi of the vishers based on the content of real vishing calls.

We have identified some limitations of our data collection system. In particular, as a future improvement, we plan to create a client application for mobile phones. This will allow us to take advantage of detailed information available in the call records of the phone. In general, having an automated mechanism to submit vishing data will increase the accuracy of the reports and avoid errors introduced by the sender. In addition, a future work is to investigate whether the nodes utilized for malicious VoIP activity belong, or are related to, the resources compromised and traded by the underground community.

REFERENCES

[1] G. Ollmann, "The phishing guide - understanding & preventing phishing attacks," http://www.ngssoftware.com/papers/NISR-WP-Phishing.pdf, NGSSoftware Insight Security Research, Tech. Rep., September 2004.

[2] R. Rasmussen and G. Aaron, "Global phishing survey: Trends and domain name use," http://www.antiphishing.org/reports/APWG_GlobalPhishingSurvey_1H2009.pdf, October 2009.

[3] S. Nambiar, S. Sainkar, D. Cowings, Y. Wee, Z. Raza, R. Shah, A. Raut, and R. Bagul, "The state of phishing," http://eval.symantec.com/mktginfo/enterprise/other_resources/b-state_of_phishing_report_05-2009.en-us.pdf, Symantec Corporation, Tech. Rep., May 2009.

[4] B. Stone-Gross, M. Cova, L. Cavallaro, B. Gilbert, M. Szydlowski, R. Kemmerer, C. Kruegel, and G. Vigna, "Your botnet is my botnet: analysis of a botnet takeover," in *CCS '09: Proceedings of the 16th ACM conference on Computer and communications security*. New York, NY, USA: ACM, 2009, pp. 635–647.

[5] M. Ahamad, D. Amster, M. Barrett, T. Cross, G. Heron, D. Jackson, J. King, W. Lee, R. Naraine, G. Ollmann *et al.*, "Emerging cyber threats report for 2009," http://hdl.handle.net/1853/26301, Georgia Institute of Technology, Tech. Rep., October 2008.

[6] D. Turner, M. Fossi, E. Johnson, T. Mark, J. Blackbird, S. Entwise, M. K. Low, D. McKinney, and C. Wueest, "Symantec global internet security threat report – trends for 2008," http://eval.symantec.com/mktginfo/enterprise/white_papers/b-whitepaper_internet_security_threat_report_xiv_04-2009.en-us.pdf, Symantec Corporation, Tech. Rep. XIV, April 2009.

[7] R. Dhamija, J. D. Tygar, and M. Hearst, "Why phishing works," in *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*. New York, NY, USA: ACM, 2006, pp. 581–590.

[8] C. Pettey and H. Stevens, "Gartner says number of phishing attacks on u.s. consumers increased 40 percent in 2008," http://www.gartner.com/it/page.jsp?id=936913, April 2009.

[9] T. McCall, "Gartner survey shows phishing attacks escalated in 2007; more than $3 billion lost to these attacks," http://www.gartner.com/it/page.jsp?id=565125, December 2007.

[10] M. Kotadia, "Phishers target yahoo instant messenger," http://www.zdnet.com.au/news/security/soa/Phishers-target-Yahoo-Instant-Messenger/0,130061744,139185847,00.htm, March 2005.

[11] J. Evers, "Phishers hijack im accounts," http://news.cnet.com/Phishers-hijack-IM-accounts/2100-7349_3-6126367.html, October 2006.

[12] M. Hofman, "There is some smishing going on in the eu," http://isc.sans.org/diary.html?storyid=6076, March 2009.

[13] J. Shah, "School of smish," http://www.avertlabs.com/research/blog/?p=75, August 2006.

[14] G. Ollmann, "The vishing guide," http://www.infosecwriters.com/text_resources/pdf/IBM_ISS_vishing_guide_GOllmann.pdf, IBM, Tech. Rep., May 2007.

[15] K. D. Mitnick and W. L. Simon, *The Art of Deception: Controlling the Human Element of Security*. New York, NY, USA: John Wiley & Sons, Inc., 2002.

[16] G. Yan, S. Eidenbenz, and E. Galli, "Sms-watchdog: Profiling social behaviors of sms users for anomaly detection," in *RAID*, ser. Lecture Notes in Computer Science, E. Kirda, S. Jha, and D. Balzarotti, Eds., vol. 5758. Springer, 2009, pp. 202–223.

[17] S. McGann and D. Sicker, "An analysis of security threats and tools in sip-based voip systems," in *Second VoIP Security Workshop*, 2005.

[18] S. Garera, N. Provos, M. Chew, and A. D. Rubin, "A framework for detection and measurement of phishing attacks," in *WORM '07: Proceedings of the 2007 ACM workshop on Recurring malcode*. New York, NY, USA: ACM, 2007, pp. 1–8.

[19] OpenDNS, "The phishtank.com home site," http://phishtank.com, February 2010.

[20] G. Networks, "Phishtank sitechecker home page," https://addons.mozilla.org/en-US/firefox/addon/3840, June 2008.

[21] F. Maggi, "The phonephishing.info home site," http://phonephishing.info, February 2010.

[22] F. T. Commission, "Debt collection faqs: A guide for consumers," http://www.ftc.gov/bcp/edu/pubs/consumer/credit/cre18.shtm, February 2009.

[23] F. C. Commission, "What is a toll-free number and how does it work?" http://www.fcc.gov/consumerfacts/tollfree.html, October 2008.

[24] M. Porter, "An algorithm for suffix stripping," *Program: electronic library and information systems*, vol. 40, no. 3, pp. 211–218, 2006.

[25] M. Cova, C. Kruegel, and G. Vigna, "There is no free phish: An analysis of "free" and live phishing kits," in *WOOT'08: Proceedings of the 2nd conference on USENIX Workshop on offensive technologies*. Berkeley, CA, USA: USENIX Association, 2008, pp. 1–8.

[26] C. Ludl, S. Mcallister, E. Kirda, and C. Kruegel, "On the effectiveness of techniques to detect phishing sites," in *DIMVA '07: Proceedings of the 4th international conference on Detection of Intrusions and Malware, and Vulnerability Assessment*. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 20–39.

[27] L. F. Cranor, S. Egelman, J. I. Hong, and Y. Z. 0002, "Phinding phish: An evaluation of anti-phishing toolbars," in *NDSS*. The Internet Society, 2007.

[28] D. K. McGrath and M. Gupta, "Behind phishing: an examination of phisher modi operandi," in *LEET'08: Proceedings of the 1st Usenix Workshop on Large-Scale Exploits and Emergent Threats*. Berkeley, CA, USA: USENIX Association, 2008, pp. 1–8.

[29] T. Moore and R. Clayton, "An empirical analysis of the current state of phishing attack and defence," in *Workshop on the Economics of Information Security*, 2007.

[30] T. Cymru, "the underground economy: priceless," http://www.usenix.org/publications/login/2006-12/openpdfs/cymru.pdf, December 2006.

[31] T. Moore, R. Clayton, and R. Anderson, "The economics of online crime," *Journal of Economic Perspectives*, vol. 23, no. 3, pp. 3–20, 2009.