Phi.sh/$oCiaL: The Phishing Landscape through Short URLs

Sidharth Chhabra*, Anupama Aggarwal†, Fabricio Benevenuto‡, Ponnurangam Kumaraguru†

* Delhi College of Engineering, † IIIT-Delhi, ‡ Federal University of Ouro Preto

sidharth.chhabra2011@coe.dce.edu, anupama1002@iiitd.ac.in, fabricio@iceb.ufop.br, pk@iiitd.ac.in

ABSTRACT

Size, accessibility, and rate of growth of Online Social Media (OSM) has attracted cyber crimes through them. One form of cyber crime that has been increasing steadily is phishing, where the goal (for the phishers) is to steal personal information from users which can be used for fraudulent purposes. Although the research community and industry has been developing techniques to identify phishing attacks through emails and instant messaging (IM), there is very little research done, that provides a deeper understanding of phishing in online social media. Due to constraints of limited text space in social systems like Twitter, phishers have begun to use URL shortener services. In this study, we provide an overview of phishing attacks for this new scenario. One of our main conclusions is that phishers are using URL shorteners not only for reducing space but also to hide their identity. We observe that social media websites like Facebook, Habbo, Orkut are competing with e-commerce services like PayPal, eBay in terms of traffic and focus of phishers. Orkut, Habbo, and Facebook are amongst the top 5 brands targeted by phishers. We study the referrals from Twitter to understand the evolving phishing strategy. A staggering 89% of references from Twitter (users) are *inorganic* accounts which are sparsely connected amongst themselves, but have large number of followers and followees. We observe that most of the phishing tweets spread by extensive use of attractive words and multiple hashtags. To the best of our knowledge, this is the first study to connect the phishing landscape using blacklisted phishing URLs from PhishTank, URL statistics from bit.ly and cues from Twitter to track the impact of phishing in online social media.

1. INTRODUCTION

With the advent of Web 2.0 technologies, social services like Twitter, Facebook, and MySpace have emerged as popular media for information sharing. The number of users on these social media services are constantly surging. Twitter is witnessing half a million new accounts per day with 140 millions tweets posted everyday.[1] Facebook has more than 500 million active users and about half of them access their account daily.[2] Although appealing as mechanisms to promote social interactions, the ease of use and creating content has also resulted in proliferation of spam such as online scams, phishing and spread of malware in online social media.[3] Studies claim that 35% of Facebook users are at risk of being phished.[4] Twitter has been attacked multiple times by phishers and has seen a spurt in crime rate during 2010.[5] Orkut also had its share of phishing attacks. For instance, a recent attack offering "free mobile recharge" compromised a number of accounts and replicated itself far and wide.[6]

While phishing detection through traditional channels such as email has been extensively researched; these solutions may not be directly applicable in online social media. The landscape of phishing in email domain has changed since phishers stole AOL account information. Gradually, generic phishing emails targeting banking and other financial services evolved. Lately, there has been a trend of targeted (spear) phishing to scam executives and managers of organizations (whaling). Phishing has even found its way from the Internet to telephone and VoIP and phishers are now using fast-flux DNS techniques to evade blacklists [15]. By using online social media as channel for phishing, phishers have taken it to the next level of difficulty to detect their acts. A variety of strategies to protect people from phishing have been proposed in the literature and implemented – silently eliminating the threat, warning users about the threat, and training users not to fall for attacks [16].

In the recent past, researchers have studied Twitter for detecting spam. Twitter is one of the most popular microblogging services, where one gets to update their status by expressing it in 140 characters. Given this constraint of limited text, users have started using URL shortening services

CEAS '11, September 01– September 02, 2011, Perth, Australia.

[1] http://blog.Twitter.com/2011/03/numbers.html
[2] http://www.facebook.com/press/info.php?statistics
[3] http://blog.neworld.com/?p=790
[4] http://blog.zonealarm.com/2010/01/35-of-facebook-users-could-be-victims-of-phishing-scams.html
[5] http://www.barracudalabs.com/downloads/2010EndyearSecurityReportFINAL.pdf
[6] http://www.chotocheeta.com/2010/07/11/orkut-phishing-attack-free-mobile-recharge-scam/

heavily to add a URL in their statuses. Phishers are making use of the URL shorteners to obfuscate the phishing URLs to spread their phishing statuses (links). As an example, Figure 1 shows a real tweet containing a blacklisted phishing URL obfuscated by a bit.ly URL. By clicking the URL in the tweet, the users are routed to a webpage which looks like the Twitter login page. Tempted by the offer to see other user's profile, users might enter his / her credentials and lose them to phishers.

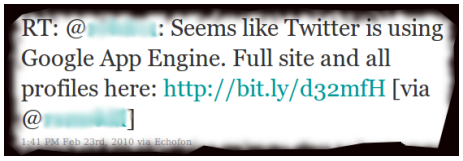

Figure 1: A tweet with a short phishing URL (using bit.ly). When a user clicks on the link, he is directed to http://wenbinginTwitter.appspot.com/, a phishing page.

Recent work has shown that spam URL blacklists work poorly for online social media because of the extensive use of URL shortening services [1]. It becomes more difficult for the users to make a judgment whether the URL is spam or not. Spammers hijack trending topics and prompt users to retweet the spam tweet so that it spreads in the compromised user's network [2, 19, 29]. Although there are techniques based on behavioral and network analysis of users to detect spam [2], there is little effort towards understanding and detecting phishing in online social media.

URL shorteners emerged as a bridge between needs of users to make their links shorter and terser, leading to obfuscating long URLs. URL shortening services were launched as early as 2001 and now there are hoards of them with bit.ly leading the pack.[7] [8] Some popular URL shorteners like bit.ly, goo.gl, is.gd, ow.ly have shortened in total billions of links till now. Bit.ly alone accounts for about half of all URL shorteners on Twitter and beats the second best by a large difference [1]. The extensive usage of URL shorteners has aided spammers to spread spam. In July 2009, 6.2% of all detected spam messages on the Internet contained short URLs.[9] Many URL shorteners also offer its users, statistics for the click-stream on their shortened links and may pose a privacy risk. Sensing danger for their survival with increased roar over these risks, involved with URL shorteners, many such services have added a phish filter in the redirection process. However, these phish filters depend on URL blacklists which are known to lag behind [11] and are not universal [24].

Our objective in this research is to track the evolution of phishing through the landscape of URL shorteners on online social media and attempt to answer some of the unexplored questions like - Which are the brands (traditional vs. online social media) targeted by phishers? Where (on the web) do these shortened phishing URLs originate from and what is the spread (across the globe) of the victims clicking on these shortened phishing URLs?

[7] 1,117 services. `http://urlshorteners.org/`
[8] `http://www.appappeal.com/the-most-popular-app-per-country/url-shortener`
[9] `http://www.certmag.com/read.php?in=3863`

Answering these questions is important for both technology developers and policy decision makers. However, to the best of our knowledge, there is no (very little) effort done in providing such overview of phishing related to online social media. In order to approach these questions, we used the blacklisted phishing URLs from PhishTank[10], linked these URLs to bit.ly and analyzed phishing shortened URLs originating from Twitter. Our main findings were:

- Phishers were using URL shorteners not only for reducing space but also to hide their identity.
- Online social media brands account for more than 70% clicks amongst the top 10 brands. Online social media brands like Facebook, Habbo, and Twitter are targeted by phishers more than traditional brands like eBay and HSBC.
- Phishing URLs which were referred from Twitter had an edge over the others with respect to attracting victims.
- Around 30% of the users turned *organic* (manual) to *inorganic* (automatic) in last 2000 tweets which is an indicator of spread of phishing (spam in general) campaigns.

The remainder of the paper is organized as follows: In the next section, we discuss related research work on phishing and URL shorteners. In Section 3, we present the data collection and explain the framework of our analysis. In Section 4, we discuss the results of our analysis, and discuss the new face of phishing through URL shorteners. We discuss the answers to the questions posed in detail in this section. In Section 5, we discuss implications of our findings and finally in Section 6, we discuss some future directions and limitations of the research presented in this paper.

2. RELATED WORK

With proliferation of spam and phishing in online social media, in recent years, many techniques have been developed for spam detection. In this section, we cover the closely related literature from spam domain as there is very little research work that has been done on detecting phishing in online social media.

There has been a number of recent papers reporting the existence of spam in different online social media services, including YouTube [3], Facebook [10], MySpace [18], and Twitter [19]. Recently, Benevenuto *et al.* [2] proposed an automatic spam detection technique using graph-based and content-based features to classify users in Twitter as spammers or non-spammers. Similarly, Wang [27] sampled Twitter users and used a classification approach to distinguish the suspicious behaviors from the normal ones. More recently, Castillo *et al.* [4] approached the problem of automatically assessing the credibility of topics defined by a set of tweets, classifying those topics as credible or not credible. Gao *et al.* [10] crawled Facebook and found that 6% of the wall posts were malicious and were posted mainly from compromised accounts. They found that URL shorteners were used to obfuscate the URL and hide the real (malicious) website. Even though URL shorteners accounted for 6%

[10] `http://www.phishtank.com/`

of URLs, each shortened URL was published more than 10 times. Finally, in a recent work, Thomas *et al.* [26] showed that spam targeting emails qualitatively differ in significant ways from spam campaigns targeting Twitter.

Most of the spam spread on Twitter is auto-generated by spammers [6]. In 2011, Zhang and Paxson proposed a method to detect automated activity on Twitter accounts using publicly available timeline of Twitter users. The fact that an *organic* user usually has a uniform distribution of tweets with respect to time, helps in detection of automated accounts. Automated accounts however exhibit a non-uniform behavior because of a fixed timing-distribution. Such non-uniform (referred as anomaly detection in classical security literature) behavior makes the automated activities detectable. The keywords associated with spam tweets have a much higher automation rate. An analysis of source of automated tweets reveals that automated tweets are sent using services like Twitterfeed and Twitter's REST API which provide automation and scheduling, as opposed to organic users who use the usual Twitter's web interface [29].

Phishing (one form of spam), is to fool gullible users out of their essentials for gaining monetary benefits, is a $2.8 billion "industry" in US.[11] It can be understood as a two step process. First lure the customer to click on the link and second fool him / her to divulge their credentials by spoofed webpage. Phishing has been extensively studied in email, phones, and websites. Researchers have studied why it works [8], how it works [13] and where it works [12]. Various solutions have been proposed like Dynamic Security Skins [7], Web Wallet [20], trusted devices [22], and PhishGuru [15]. We cannot do justice to the entire literature, as there are lot of them. However, we have covered a few to show the spectrum of work in this area.

Though rule based features succeed in phishing detection to some extent, the use of URL shortening services allow the phishers to evade phish blacklists hence significantly reducing their detection effectiveness. Antoniades *et al.* present a characterization of short URLs. The popularity of short URLs has hiked because of the character limit in IM systems and micro-blogging websites like Twitter. They show that the maximum access to short URLs come from emails and online social media. Twitter is among the top 5 referrers of bit.ly short URLs and comprises 11.77% of the total accesses. It is shown that the click distribution of shortened URLs is approximately a log-normal curve. A temporal analysis shows that some fraction of short URLs are not short lived and exist for more than three months [1]. Grier *et al.* used bit.ly API to collect click-through information about the spam URLs which were obfuscated using bit.ly URL shortening service and found that Twitter spam is more effective than email spam with a high click-through rate of 0.13% [11]. Some phishing filters are based on domain reputation [23] and some on heuristics of the phishing emails and websites [9, 30]. Kumaraguru *et al.* have proposed educating users on basics of domain names, URLs, etc as a solution to the problem of phishing [16]. These approaches would be of little use with shortened phishing URLs. Some services such as LongUrl provide a Firefox add-on which can give you a peek into the link even before you view the actual webpage.[12] A preview may be useful to differentiate between a phishing campaign and non-campaign but phishing websites usually resemble the original websites and therefore even a preview will not help with phishing campaigns.

After analyzing the literature, we found that not much work has been done on studying the true positive (close to ground truth) phishing datasets and characterizing the landscape of phishing in online social media. This lead us to use PhishTank (an openly available phish database) to analyze phishing that is done through short URLs (bit.ly). We also used Twitter to study the use and impact of the micro-blogging site in phishing using bit.ly.

3. METHODOLOGY AND DATASETS

In this section, we describe the architecture that we built to gather data from PhishTank, bit.ly and Twitter; we used the data for performing the analysis that would answer the questions that we described in Section 1.

3.1 Data Collection

We used a three-phase system in our framework (see Figure 2) for analysis of phishing through online social media. The first step (Dataset collection in Figure 2) was to fetch the PhishTank [13] database for the year 2010 and filter those which were voted 'yes'.[14] We obtained 1,18,119 such URLs. In the second step (Filtering in Figure 2), we query "LookUp" endpoint of bit.ly API for each URL from step 1. "LookUp" endpoint returns the global shortened URL (`http://bit.ly/[hash]`) for a given long URL (`http://www.abcdef.ghi/jkl/mno`) if there is any. A long URL may have many short URLs, shortened by different users (after logging into the bit.ly system) but the global shortened URL keeps cumulative count of statistics for every click on the long URL through bit.ly. It is also possible that the phisher shortened not the phishing page but the domain name and then lured the user to the phishing page. Therefore, we also queried the domain name for every phishing URL. One can also query every file structure inside the URL, but it would increase the number of queries exponentially. Bit.ly API has rate limits, therefore, we chose to look up only the long URL and its domain name. There are cases when phishing pages are hosted inside famous and trustworthy domain names for example there were a few (in our dataset) on Google Spreadsheets. To filter such false negatives (popular domains), we removed URLs with exceptionally high number of clicks.[15] At the end of this process, we had 6,474 short URLs for "phishing" URLs with 3,692 exact matches and rest domain matches. In the third step (Analysis in Figure 2), for every short URL we query different API end-points namely clicks, clicks-by-day, countries and referrers from bit.ly. For referrals from Twitter, we used Twitter API to fetch additional information.

3.2 Dataset

Our dataset comprises of phishing URLs collected from PhishTank for the period Jan 1, 2010 to Dec 31, 2010.

[11] `http://www.brandprotect.com/catching-a-phish.html`
[12] `http://longurl.org/`
[13] `www.phishtank.com`
[14] The number of people required to verify a phish depends on the history of those users who are voting. It will always be more than one. `http://www.phishtank.com/faq.php`
[15] Hundred and thirty eight domain names with average clicks 1,00,430.3 were removed

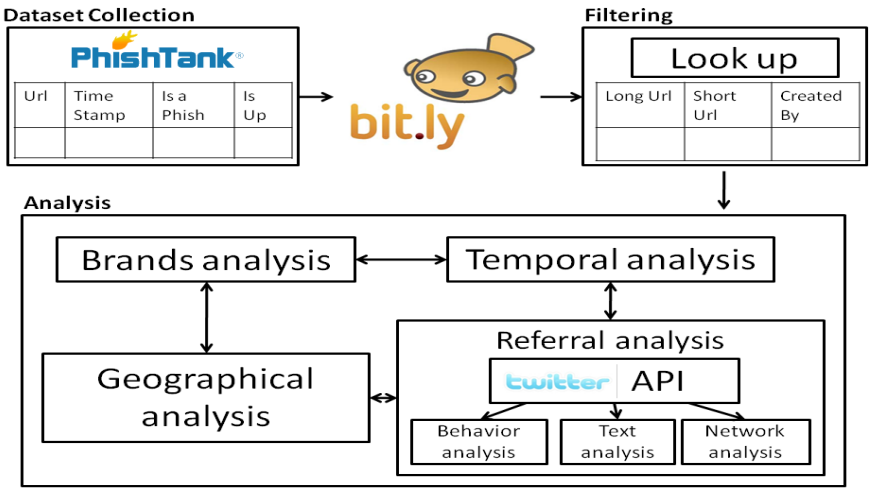

Figure 2: Architecture that we built to collect true positive phishing URLs from PhishTank. We used these URLs to retrieve the statistics from bit.ly and then do different analysis on the combined information.

PhishTank *is a collaborative clearinghouse for data and information about phishing on the Internet.* Anyone can contribute to PhishTank by voting / submitting URLs. They use an adaptive cut-off for number of votes required for a submitted URL to be declared verified (referred as yes in our dataset). Figure 3 shows distribution of these URLs. June and July have the highest number of 'yes' URLs. On average, number of unknown URLs decreased with time. The dataset had total 1,96,442 URLs out of which 1,18,119 were voted 'yes' by users as phishing URLs.

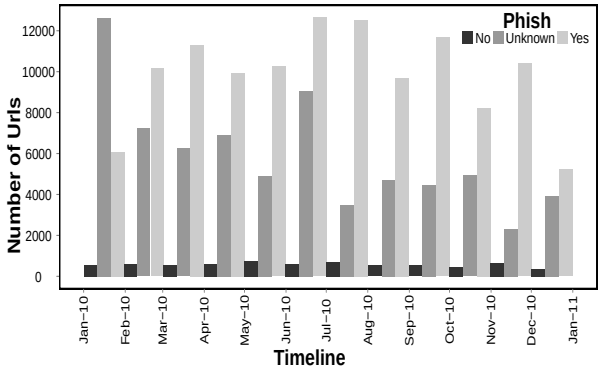

Figure 3: Number of URLs in our dataset; shows number of unknown URLs is decreasing over time.

Among the 1,18,119 URLs, there were 63,175 unique primary domain names with distribution as shown in Figure 4. The figure shows that number of URLs per domain accounts for a power law distribution. Some top primary domain names were *altervista, tennesse, timexsenegal, 110mb* and *atspace.*

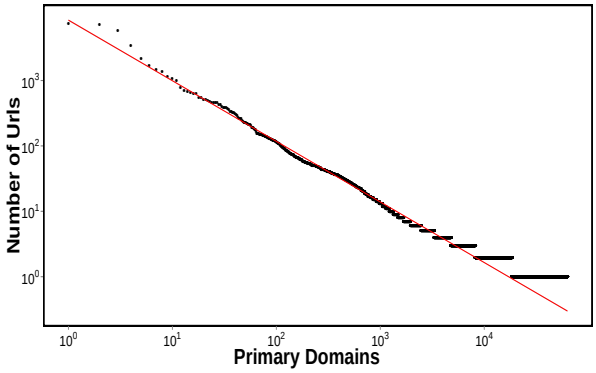

Figure 4: Distribution of number of primary domain names with number of URLs; shows a straight line fit on log-log scale.

Table 1: PhishTank dataset for the period of 1 Jan, 2010 until 31 Dec, 2010.

		Vote		
		Yes	No	Unknown
Online	Yes	11,081	392	1,234
	No	1,02,175	5,991	68,731
	Unknown	4,863	523	795

Table 1 shows the break up of URLs in PhishTank dataset for their final voting status (verified i.e. yes, rejected i.e. no and inconclusive i.e. unknown) and their online presence (online i.e. yes, offline i.e. no and can't determine i.e. unknown). We see majority (87.2%) of online URLs were voted yes whereas only about half of offline URLs were verified. Most of inconclusive URLs were offline.

4. RESULTS

In this section, we present the detailed analysis that we did on the data to answer the questions that we raised in

Section 1. We discuss the impact of using URL shorteners by analyzing the space gain, change in landscape of the target brands attacked by phishers, referral analysis for the bit.ly URLs pointing to phishing URLs and geographical spread of the victims clicking on the phishing URLs.

4.1 Space Gain

A bit.ly URL has two parts, the domain name, *bit.ly* and a hash $HaSh10$ (a case sensitive alphanumeric code). Bit.ly hashes can be of variable sizes; with time, the length of hashes have increased from 4 to 6 (to accommodate increasing demand). Our dataset (6,474 URLs) had 24 hashes of length 4,462 of length 5, and the rest 5,988 had length 6. Bit.ly started using length of 6 as default from mid of 2010. Majority of bit.ly hashes in our dataset are of length 6 which may imply that most of these shortened URLs were created in later half of 2010.

To ascertain if bit.ly has really helped phishers, we calculate the *space gain* for each URL. By space gain, we mean the fraction of space saved by using bit.ly URL instead of the actual long URL. We find average space gain to be 39%; Figure 5 shows the cumulative *space gain* for PhishTank URLs found in bit.ly. For 50% of the phishing URLs, we observe a 37% or less space gain; for generic URLs, researchers have shown 91% space gain [1]. This implies that URLs shortened by phishers are not as long as generic URLs which favors the hypothesis that URL shorteners are not only used to make communication terse but to hide the real link behind bland hash so as to escape any scrutiny which is based on only URL text. Also Antoniades *et al.* found URLs with space gain ≤ 0 to be negligible [1], whereas we find 379 (5.9%) URLs which are shorter or equal to the shortened URL. We believe, phishers are taking up this loss in trade for new identity and an extra hop. Adding multiple hops, i.e., the use of URL redirection can break apart many spam filters and gain trust of users. URL redirection is a common technique used by phishers and it has been studied in past by analyzing vulnerabilities of open redirects [25].

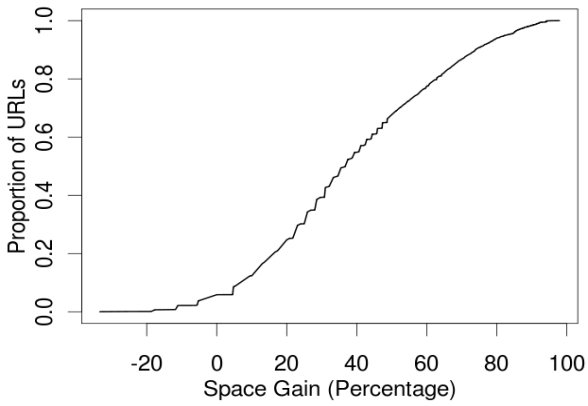

Figure 5: Shows the cumulative space gain in percentage against the proportion of URLs. We find that about 37% or less of space gain for half the URLs in our dataset.

4.2 Target Brands

In this section, we present the evolution of phishing targets from e-commerce services / financial institutions to online social media brands like Facebook, Orkut. To investigate the brands (or companies) targeted by phishers using URL shortener, we followed two steps: first, we used the brands listing created by PhishTank; the administrators of PhishTank maintain a list of popular 104 brands.[16] When tagging or annotating a URL as phishing, PhishTank suggests the contributor (one who adds or annotates the phishing link in PhishTank) to select the brand from the previously generated list. We compared the URLs that we had with the URLs from the 104 PhishTank brands, we found 50 popular brands amongst 2,349 URLs. We could not confirm the target for most of them as about 95% of these pages were down at the time of analysis.

Figure 6 shows the frequencies for top 10 brands with boxplot for number of clicks they received during the period of our analysis. The bottom and top of the box are 25^{th} and 75^{th} percentiles, and the band near the middle of the box is 50^{th} percentile. Four of the top 10 are online social media brands (Facebook, Orkut, Habbo and Zynga); five are e-commerce services (Bradesco, eBay, HSBC, IRS and PayPal) and one an email service provider (Live). Online social media brands account for more than 70% clicks amongst the top 10 brands. Though PayPal accounts for a third of "branded" URLs but median number of clicks (50^{th} percentile) is 1, whereas Bradesco[17] has about 100 URLs but median number of clicks around 200. Habbo[18] is a recent entry to online social media and is a prime target amongst all online social media brands, in-fact, it has almost the same number of shortened phishing URLs as PayPal. Fast growth coupled with large user base supported by open architecture of online social media is making it more lucrative and easy to attack for phishers day-by-day. We see that there are many brands where the number of URLs are low, but the median clicks are high and some where the number of URLs are high and the median clicks are low which negates the silent assumption that large number of phishing URLs trap large number of victims.

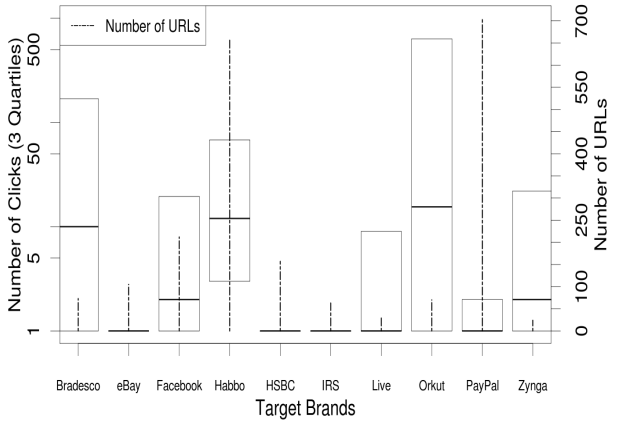

Figure 6: Shows the frequencies for top 10 brands with number of clicks they received during the period of our analysis. Four of the top 10 are online social media brands.

[16] http://www.phishtank.com/target_search.php
[17] http://www.bradesco.com.br
[18] http://www.habbo.com.br

To observe the changing focus of phishers, we draw the temporal distribution of clicks for branded URLs. Figure 7 shows weekly average of clicks for top 5 brands. We look at only the top 5 brands because they constitute 78.37% clicks out of all branded URLs. We see that Habbo's average number of clicks increased heavily after Sept 2010; there seems to be a large difference between average clicks for Habbo and the next hit brand PayPal after Sept 2010. We also observe that on average PayPal's clicks are increasing with time and follows a cyclical pattern whereas Facebook achieved peak during July and August. HSBC received some traffic during February and eBay got spikes of clicks sparsely through the year. This indicates about the change in focus of phishers, from financial institutions / e-commerce websites to online social media. It brings to fore the need to shift focus of phish detectors to online social media.

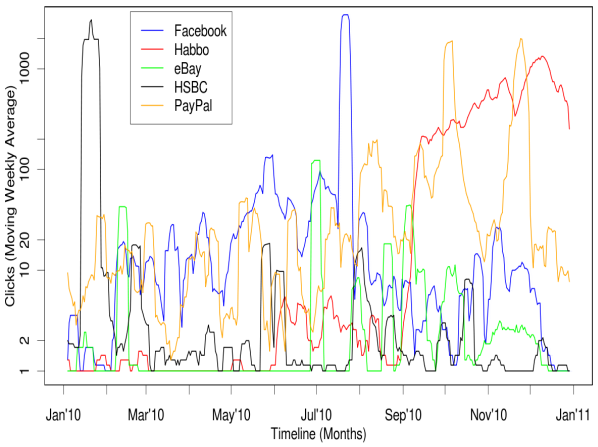

Figure 7: Shows the weekly average of clicks for top 5 brands. We see that Habbo has become primary target of phishers in last quarter (October, November, December).

4.3 Referral Analysis

Until now, we have been focusing on victims but now we examine the breeding zones, the places which are actively used to spread these phishing links. We query bit.ly API to fetch referral for clicks on every URL and get cumulative click counts for URLs which lead to the bit.ly URL. From the bit.ly statistics, we found that websites other than Twitter, which impose no limits on text length are also referrals for short phishing URLs. This affirms the belief that short URLs are used for the purpose of hiding suspicious URLs behind cover rather than shortening them. Antoniades et al. found that Facebook accounted for only 1.72% referrals (in general short URLs) [1]. However, here we found that Facebook accounted for 11.13% referrals of phishing short URLs and Orkut accounted for 31.48%. Twitter has risen from 12% in general URLs to sizeable 23% of all referrals in phishing URLs and has become a phisher's paradise. This significant increase clearly shows that online social media is the new target of phishers. Emails play a lesser role (18%) here which makes URL shorteners more suspicious for phishing in online social media.

To characterize more about the phishers, we queried the Twitter API for the referrals we got from the bit.ly. Orkut and Facebook don't allow crawling user profiles, so, we could not get more information from them. Though Twitter accounts for 23% of referrals, a significant portion of referral came from protected / private accounts.[19] We found 990 Twitter users who were public when they tweeted or retweeted one or many "phishing" status (tweet). At the time of analysis, only 864 among them were present, rest were suspended or deleted.[20] Table 2 reports some of the characteristics of these users who tweeted a phishing URL. Large number of followers, followees and statuses are indicators of automated accounts [10].

To study whether Twitter acts as a strong push for a phishing URL's publicity or not, we look at the clickthrough and geographical statistics from bit.ly. Table 3 compares summary statistics for number of clicks, geographical spread (number of countries), temporal spread (lifetime) and web popularity (number of referral) among URLs which were referred and not referred by Twitter. It clearly shows that phishing URLs which were referred from Twitter had an edge over the others. In-fact mean lifetime was seven times higher and mean number of clicks quadrupled. Therefore, building technologies around curbing phishing in Twitter (online social media) can largely reduce the effect of phishing on the Internet.

4.3.1 Behavior Analysis

Next, we classified user profiles into *organic* and *inorganic*. An *organic* account is one of a legitimate Twitter user who posts her tweets manually. Hence, an organic user usually has a uniform distribution of tweets with respect to time. Whereas, an *inorganic* account would exhibit detectable non-uniformity in timing pattern [29]. In order to detect behavior, we collected recent status updates (max. 200) and used Pearson's χ^2 test to determine if the timing distribution is uniform (human). We were able to obtain timeline (status updates) for 820 users as others were protected. Zhang and Paxson used this method and concluded that 16% of profiles were found to be using *inorganic* means [29]. We used same parameters i.e. $p < 0.001$ as threshold and if either of minutes or seconds distribution had p-value less than 0.001 then it was tagged *inorganic*. We found that 89% of the profiles were using their accounts inorganically. Inorganic accounts exhibit a robotic pattern of status updates as shown in Figure 8 (Present). A point in the plot is the time-stamp of tweets by the user. The user updates were concentrated around mid of the hour. With the advent of tools like tweetdeck, powertwitter etc., one can schedule tweets and thus create a mirage of a trend. We observed here two things, firstly, the probability of using inorganic means for wrong motive is significantly higher than for general purpose and secondly, phishers still employ the good old personalized message which is why a 11% of users update their timeline manually (*organically*).

To find if the users' behavior has changed over the time we went back 2000 status messages in their timeline (we were able to download tweets in the past only for 516 users), retrieved 200 tweets (from 2001 until 2200) and again did the same behavioral analysis. We found only 213 (41.3%) Twitter users were behaving inorganically. Around 153 (29.7%)

[19] `http://support.twitter.com/entries/14016-about-public-and-protected-accounts`

[20] `http://support.twitter.com/articles/15790-my-account-is-suspended`

Table 2: Some descriptive statistics of users who tweeted a phishing URL.

Statistic	Friends	Followers	Status Messages
Minimum	0	0	0
Median	597	302.5	2,138
Average	5,724.3	1,189	6,369.1
Maximum	12,49,000	1,27,291	3,40,113

Table 3: Summary statistics for URLs with / without Twitter referral. All values are mean with median in bracket. Shows that phishing URLs which were referred from Twitter has an edge over the others.

Twitter Referral	Clicks	Countries	Lifetime	Referrals
No	70.2(0.0)	2.7(0.0)	12.8(0.0)	2.5(1.0)
Yes	95.1(14.0)	7.2(4.0)	84.0(8.0)	7.2(4.0)

users turned *inorganic* to *organic* in last 2000 tweets which is an indicator of spread of phishing (spam in general) campaigns. As an example of change in behavior, Figure 8 shows the message-posting pattern of a user in present and past (200 tweets back).

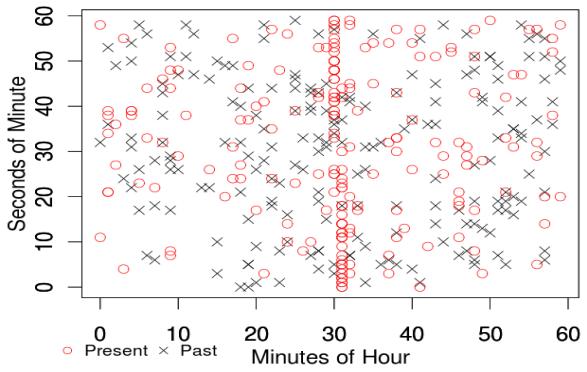

Figure 8: Temporal pattern for status updates of a user. A point on the plot is time-stamp for a tweet. Past (black X) is the posting pattern for user in the past (2000 tweets back) and Present (red circle) is the posting pattern in the present for 200 tweets. Shows change from organic (manual) to inorganic (automatic).

4.3.2 Network Analysis

In order to understand the network properties, we queried Twitter API for "Friends and Followers resources." We found that the friend-follower network is sparse with 254 nodes having 356 links among them as shown in Figure 9. Even though this is a small sample, we found $1/3^{rd}$ of nodes were connected. The network density is 0.01 and reciprocity is 56%, which is significantly higher than the 22% that has been observed in general population of Twitter [17]. Spammers increase their influence by following various strategies to increase the number of followers. One of the strategy is to follow others in hope of getting followees (return favor)

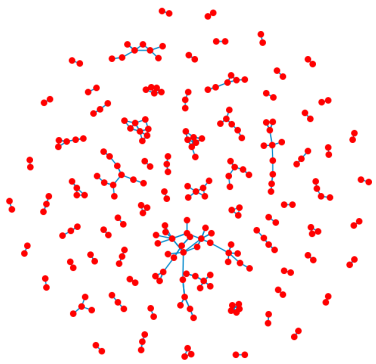

Figure 9: Network of the people tweeting the phishing URLs; shows sparse network with high reciprocity. We used ORA from CASOS Center at CMU to create this figure.

and later discontinue being follower.[21] A high reciprocity and sparseness points towards to a similar strategy being followed amongst phishers. Also $1/5^{th}$ of links are Simmellian ties [14] and there are 26 cliques of sizes 3, 4 and 5 which points to presence of strong ties amongst the nodes. Simmellian ties are measure of cooperation amongst nodes and longevity of the network. This implies that phishers communicate and network effectively to achieve higher gains (trap victims).

4.3.3 Text Analysis

The references from bit.ly API also contains few direct links to tweets. We did text analysis of these tweets to infer the properties of phishing tweets. There were 120 tweets of which 67 were in English rest were in Brazilian, Dutch, Russian and others. The average length of English tweet was 95.7 characters (min. 33, max. 140, var. 32.5) which is close to the limit (140 characters). Since most of the references for phishing URLs come from private / protected Twitter accounts, we were not able to obtain a larger set of tweets through this architecture. By reversing the way we collected the data (as shown in Figure 2) i.e. using tweets collected during 2008-2009 [2, 5], we filtered 3,692 tweets containing shortened phishing URLs using the PhishTank data 2008-2009. Two tweets: "Allows you to schedule your Twitter messages for future" and "How to Send Custom Tweet List about your product on recurring basis" accounted for $2/3^{rd}$ of tweets dataset. Use of third party Twitter applications to schedule their tweets is popular among phishers. Tweetdeck is the most popular app used in 48.5% tweets followed by Twitrobot at 21.6%. Phishers use these services to spread phishing URLs at specific intervals with least effort. The average length of tweets containing the phishing URLs was found to be 83.3 (min. 29, max. 140, var. 60.9) characters. Figure 10 is the tag cloud for these tweets. The tag cloud shows the most popular words were "Product", "Twitter" and "Friends." Though there were 3,692 tweets containing phishing URLs, there were only 89 distinct sentences which appeared in these tweets. This shows that phishers

[21] `http://www.barracudalabs.com/downloads/2010EndyearSecurityReportFINAL.pdf`

spread the same phishing tweets by rewetting and re-posting in their network.

We also observe that the language used in the phishing tweets was proper English and was devoid of use of any shorthand language which is a common practice amongst genuine Twitter users. Similar results were obtained in mobile spam literature too – spam SMSes contain proper words like free, sms, call, get etc., while legitimate SMSes contain mostly shorthand and local language words like, ur, ki, hai, kya, and emoticons [28]. To test this, we crawled Twitter user profiles for the tweets belonging to similar period (2008-09). We began with a seed of 10 public Twitter users[22] and crawled their friends and followers recursively until we had 3400 tweets containing short URLs (URL's length $\leq$ 22, which is the length of short URLs). The average length of these tweets was found to 60.1 (min. 25, max. 140, var. 47.8) which is significantly less than observed in phishing tweets. Phishers make efficient use of the space gained by shortening the URLs and lure the viewer by fascinating words. We also checked the words in the tweets against a dictionary[23] and found that 93.2% words in phishing tweets were present in dictionary whereas in the crawled tweets only 72.02%. We also found that maximum number of hashtags in the crawled tweets was 4 whereas in phishing tweets dataset, it was 16. Hence, length, hashtags and language of tweets can act as discriminating features in a phishing filter.

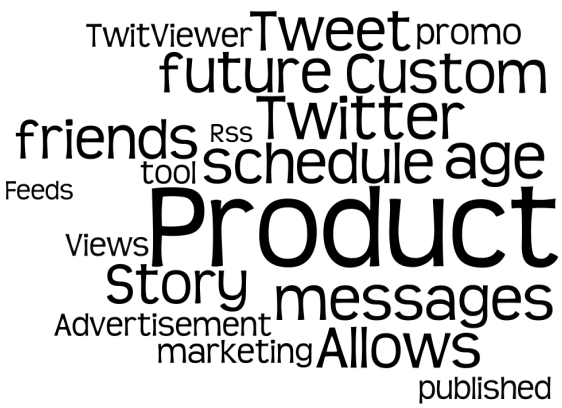

Figure 10: Tag Cloud for the words from tweets containing phishing URL. Some frequent words are: "product", "story", "allows" and "schedule."

4.4 Locational analysis

To find the reach (geographical) of these bit.ly shortened phishing URLs across the globe, we fetched geographical tags associated with clicks on these URLs using the "countries" endpoint of bit.ly API. For each URL, we got a list with number of clicks and ISO code for the country (e.g. IN = India). In total, these URLs were clicked from as many as 140 different countries across the globe. Figure 11 shows countries colored proportional to number of clicks on the phishing URLs. For every country, we divided the number of clicks by respective Internet population.[24] Brazil takes the lead here, closely followed by the US and Canada. It may be because of the emergence of Habbo as the prime target and its major audience being in Brazil. Count of URLs is a kind of measure of intent of phisher and clicks on that are measure of phisher's success rate. USA, India and Brazil are the prime target destinations for phishers as most phishing URLs were found in these countries. Table 4 enlists popular brands in some of the popular target destinations. There are financial institutions (type of organizations that got phished heavily in the last few years) like HSBC and Bradesco in the listing. Except for PayPal in Russia, all other countries has online social media at the top of the list. In-fact except for Russia, all other countries have two of three brands as some online social media brand. This shows that online social media has become a favorite target for phishing attacks.

Table 4: Top brands in different countries. Shows that online social media services has been the biggest target across countries. Brands presented in the order of first, second, third.

Country	Brands
USA	Habbo, PayPal, Facebook
India	Orkut, PayPal, Facebook
Brazil	Habbo, Bradesco, Orkut
Great Britan	Habbo, PayPal, Facebook
Russia	PayPal, HSBC, Habbo
Australia	Habbo, PayPal, Orkut

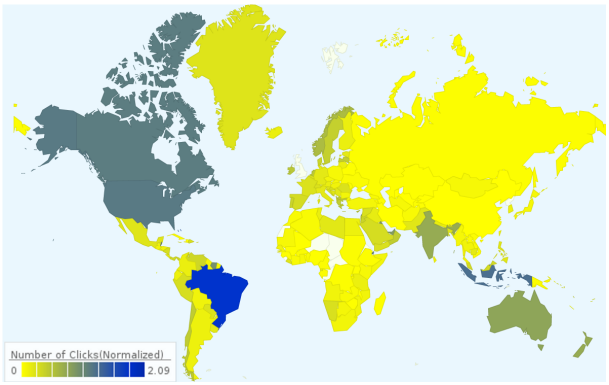

Figure 11: Location information for the clicks, normalized for different Internet population; shows Brazilians have been the most gullible followed by US and Canada.

To find if any correlation exists between a URL's geographical spread and its lifetime (days between first and lask click during the period 1/1/2010 - 28/2/2011), we draw Figure 12 where every point is a URL whose x, y and size is determined by its lifetime, geographical span and number of clicks respectively. In the figure, we see that there are URLs which had short lifetime but spanned more than 80 countries with more than 1000 clicks. And there are URLs which have lifetime of more than 400 days but few clicks and not spread in more than 20 countries. We observe large number of clicks for URLs which are evenly spread geographically and temporally.

[22] 10 userids from `http://www.public.asu.edu/~mdechoud/datasets.html`

[23] Created using `http://norvig.com/big.txt`

[24] `http://www.internetworldstats.com/stats.htm`

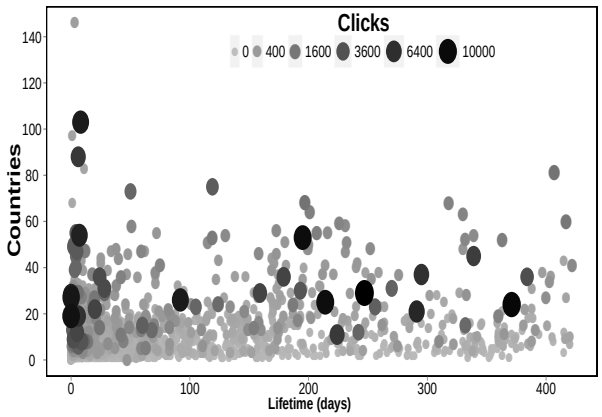

Figure 12: Scatter plot of phishing URLs for geographical and temporal spread.

5. DISCUSSION

With newer technologies and deeper Internet penetration, phishing strategies and targets are evolving. We analyzed and discussed phishing attacks on online social media using URL shorteners. A shortened URL makes the length of actual URL shorter and hides the long URL behind it. We found that space gain for half of phishing URLs in our dataset was 37% which was significantly less than space gain in general URLs on the Internet. Also the tweets which quoted these URLs had length more than the general. This points to malicious intent of phisher for hiding their URLs behind bland hashes and entice viewers through words. The new identity in form of bit.ly domain name gives a sense of trust to the viewer and takes him / her to the trap page. These results show that phishers have been continuously changing their channels (emails, instant messaging, targeted emails, mobile phones, etc.) and finding ways (URLs shorteners) to lure victims by various means.

The next question, we had was, who were the targets of this new strategy. We found that online social media brands namely Facebook, Habbo and Orkut were amongst the top targets competing with e-commerce services. This shift in phishers' approach can be attributed to rapid growth, wider and gullible audience and open platform of online social media. We also found that geographically, USA, India and Brazil were the target countries of phishers. Phishers have realized that developing countries are fast catching up with the developed and are untapped "markets" for luring victims. In-fact, Brazilians were most victimsed which can be also be attributed to popularity (amongst phishers) of Habbo, a south-American social media brand. The evolved phishing tactics requires that systems be developed to catch URLs which are embedded in the social media, strong regulatory provisions by governments in developing countries and pooling of the international collaboration to run behind "global" phishers.

And finally, we traced the footprints of phishers. Most referrals for these phishing URLs arose from online social media. Facebook, Orkut and Twitter combined accounted for 2/3rd of all referrals. Most of the Twitter accounts publishing these tweets were *inorganic* (automated). Third party applications were extensively used for this purpose. On analyzing the tweet text we found, usage of proper english, longer tweets and more hashtags. A filter based on both semantic and syntactic text features and network properties can be effective for detection of such Twitter accounts.

6. LIMITATIONS AND FUTURE WORK

Even though we use PhishTank and bit.ly for our analysis, we are aware of certain limitations of both. Our results should be considered keeping these limitations in mind. PhishTank dataset has two issues. Firstly, out of 1,18,119 'yes' URLs, only 11,081 URLs were online at the time of collection as shown in Table 1. Around 90% of the URLs were offline when users voted on it, which poses a question on the validity of it being really a phish. Secondly, PhishTank is dominated by active users and few users make large number of votes, which makes it susceptible to manipulations. A few users can go hand-in-hand to add false positives (a legitimate site marked as phishing site) to the data that is being collected by PhishTank. Nevertheless, the accuracy of PhishTank is 97% [21].

As far as our knowledge goes, this is one of the first research work to get PhishTank, bit.ly, and online social media together to characterize phishing. We foresee a lot of future research questions that can be posed in this direction. As the first step, one can look at enhancing the PhishTank dataset to a more comprehensive and complete dataset (e.g. from Antiphishing Working Group) of phishing URLs.

Acknowledgments

The authors would like to thank JW Lim, moderator at PhishTank who shared the PhishTank URLs with us. We would also like to thank The National Council for Scientific and Technological Development (CNPq) for supporting the workshop, Privacy and Security in Social Media: An International Perspective, 2011 which brought two of the authors to work together on this research problem. We also thank all members of PreCog research group at IIIT-Delhi for their valuable feedback and suggestions. Authors thank Aditi Gupta for her feedback on initial drafts of this paper.

7. REFERENCES

[1] D. Antoniades, I. Polakis, G. Kontaxis, E. Athanasopoulos, S. Ioannidis, E. Markatos, and T. Karagiannis. we.b: The web of short URLs. *20th International World Wide Web Conference (WWW)*, 2011.

[2] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida. Detecting spammers on twitter. In *Proceedings of the 7th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference (CEAS)*, 2010.

[3] F. Benevenuto, T. Rodrigues, V. Almeida, J. Almeida, and M. Gonçalves. Detecting spammers and content promoters in online video social networks. In *Int'l ACM Conference on Research and Development in Information Retrieval (SIGIR)*, pages 620–627, 2009.

[4] C. Castillo, M. Mendoza, and B. Poblete. Information credibility on twitter. In *20th International Conference on World Wide Web (WWW)*, WWW '11, pages 675–684, New York, NY, USA, 2011. ACM.

[5] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi. Measuring user influence in twitter: The

million follower fallacy. In *4th International AAAI Conference on Weblogs and Social Media (ICWSM)*, Washington DC, USA, May 2010.

[6] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia. Who is tweeting on twitter: human, bot, or cyborg? In *Proceedings of the 26th Annual Computer Security Applications Conference*, ACSAC '10, pages 21–30, New York, NY, USA, 2010. ACM.

[7] R. Dhamija and J. Tygar. The Battle Against Phishing: Dynamic Security Skins. In *SOUPS '05: Proceedings of the 2005 symposium on Usable privacy and security*, pages 77–88, New York, NY, USA, 2005. ACM Press, New York, NY.

[8] R. Dhamija, J. D. Tygar, and M. Hearst. Why Phishing Works. *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, 2006.

[9] I. Fette, N. Sadeh, and A. Tomasic. Learning to detect phishing emails. *16th International conference on World Wide Web*, June 2006. Retrieved Sept 2, 2006, http://reports-archive.adm.cs.cmu.edu/anon/isri2006/CMU-ISRI-06-112.pdf.

[10] H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Y. Zhao. Detecting and characterizing social spam campaigns. In *Proceedings of the 10th annual conference on Internet measurement*, IMC '10, pages 35–47, New York, NY, USA, 2010. ACM.

[11] C. Grier, K. Thomas, V. Paxson, and M. Zhang. @spam: the underground on 140 characters or less. In *Proceedings of the 17th ACM conference on Computer and communications security*, CCS '10, pages 27–37, New York, NY, USA, 2010. ACM.

[12] T. Jagatic, N. Johnson, M. Jakobsson, and F. Menczer. Social phishing. *Communications of the ACM*, 50(10):94–100, October 2007.

[13] L. James. *Phishing Exposed*. Syngress Publishing, Canada, November 2005.

[14] D. Krackhardt. Simmelian ties: super strong and sticky. In R. M. Kramer and M. A. Neale, editors, *Power and Influence in Organizations*, pages 21–38. Sage, Thousand Oaks, CA, USA, 2008.

[15] P. Kumaraguru. *PhishGuru: A System for Educating Users about Semantic Attacks*. PhD thesis, School of Computer Science, Carnegie Mellon University, 2009. http://reports-archive.adm.cs.cmu.edu/anon/isr2009/abstracts/09-106.html.

[16] P. Kumaraguru, S. Sheng, A. Acquisti, L. F. Cranor, and J. Hong. Teaching johnny not to fall for phish. *ACM Trans. Internet Technol.*, 10:7:1–7:31, June 2010.

[17] H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 591–600, New York, NY, USA, 2010. ACM.

[18] K. Lee, J. Caverlee, and S. Webb. Uncovering social spammers: social honeypots + machine learning. In *Int'l ACM Conference on Research and Development in Information Retrieval (SIGIR)*, pages 435–442, 2010.

[19] K. Lee, B. Eoff, and J. Caverlee. Seven months with the devils: A long-term study of content polluters on twitter. In *Int'l AAAI Conference on Weblogs and Social Media (ICWSM)*, 2011.

[20] R. C. Miller and M. Wu. Fighting Phishing at the User Interface. *O'Reilly*, August 2005. In Lorrie Cranor and Simson Garfinkel (Eds.) Security and Usability: Designing Secure Systems that People Can Use.

[21] T. Moore and R. Clayton. Evaluating the wisdom of crowds in assessing phishing websites. In G. Tsudik, editor, *Financial Cryptography and Data Security*, volume 5143 of *Lecture Notes in Computer Science*, pages 16–30. Springer Berlin / Heidelberg, 2008.

[22] B. Parno, C. Kuo, and A. Perrig. Phoolproof phishing prevention. *Proceedings of the Financial Cryptography and Data Security 10th International Conference*, March 2006. Retrieved Nov 5, 2006, http://sparrow.ece.cmu.edu/ parno/pubs/phishing.pdf.

[23] P. Prakash, M. Kumar, R. Kompella, and M. Gupta. Phishnet: Predictive blacklisting to detect phishing attacks. In *INFOCOM, 2010 Proceedings IEEE*, pages 1–5, March 2010.

[24] S. Sheng, B. Wardman, G. Warner, L. F. Cranor, J. Hong, and C. Zhang. An empirical analysis of phishing blacklists. *Proceedings of the 6th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference (CEAS)*, 2010.

[25] C. A. Shue, A. J. Kalafut, and M. Gupta. Exploitable redirects on the web: Identification, prevalence, and defense. In *WOOT*, 2008.

[26] K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song. Design and evaluation of a real-time url spam filtering service. In *IEEE Symposium on Security and Privacy*, 2011.

[27] A. Wang. Don't follow me: Spam detection in twitter. In *Int'l Conference on Security and Cryptography (SECRYPT)*, 2010.

[28] K. Yadav, P. Kumaraguru, A. Goyal, A. Gupta, and V. Naik. SMSAssassin : Crowdsourcing Driven Mobile-based System for SMS Spam Filtering. *Accepted at HotMobile*, 2011.

[29] C. M. Zhang and V. Paxson. Detecting and Analyzing Automated Activity on Twitter. *Passive and Active Measurement Conference*, 2011.

[30] Y. Zhang, J. Hong, and L. Cranor. Cantina: A content-based approach to detecting phishing web sites. In *Proceedings of the 16th International conference on World Wide Web*, 2007.