Classification of Phishers

Ram Dantu, *University of North Texas*
Srikanth Palla, *University of North Texas*
Joao Cangussu, *University of Texas at Dallas*

Classification of Phishers

Ram Dantu, Srikanth Palla, and Joao Cangussu

Abstract

Phishing attackers masquerade as genuine senders and try to steal consumers' personal identity data and financial account credentials. In spite of aggressive efforts, technology companies have had limited success in restricting phishing attacks. Unfortunately the nature of phishing attacks changed over time from passive, such as password guessing and eavesdropping to active attacks, such as employing Trojans to intercept traffic and adopting social engineering techniques. No matter how many authentication techniques we develop, phishers always adapt. However, phishers cannot become part of the recipient's social network without consent. Though they can forge certain fields in an email header, phishers do not have access to the complete header. In this paper, we describe techniques for detecting phishers based on their traffic paths, traffic patterns, and on the receivers' social network. Considering such issues, we based our solution on the trustworthiness of the relays participating in routing the emails. We examine the email's header rather than the content. We designed our classifier to perform the following analyses in four steps: i) DNS-header analysis, ii) Social network analysis, iii) Wantedness analysis, and iv) Proactive classification. We classify phishers into: i) Serial phishers, ii) Recent phishers, iii) Prospective phishers, and iv) Suspects. Finally, our classifier appends an alert level or label to the email's "subject" before adding the email to the inbox.

1 Introduction

Phishing is a form of online identity theft; it uses social engineering techniques to lure consumers into revealing confidential information such as social security numbers, user names, passwords and financial credentials. Phishers typically use inexpensive communication channels such as e-mail, instant messaging, and VoIP phone calls to launch phishing attacks. They hijack the brand names of reputable organizations involved in on-line trade and commerce such as banks, e-retailers, and credit card companies to create a plausible premise and convince recipients to release confidential information that can be used for malicious purpose.

The Anti-phishing Working Group [1], a global consortium of technology firms, and law enforcement organizations reports that it received 20,109 unique phishing reports in May 2006. They identified 11,976 unique phishing websites in May 2006 and the majority of these phishing websites are hosted in the United States. According to their May 2006 phishing activity trends report [2], 137 reputable brands were hijacked between May 2005 and May 2006. Phishing attacks escalated to a double-digit rate by the end of 2005. A majority of Internet users regard with suspicion or distrust e-mail messages from companies or individuals whom they do not know from prior experience. E-commerce companies such as PayPal and eBay are on the verge of losing consumers' trust due to phishing attacks [21]. If this continues, banks and other online trading companies may not rely on e-mail messages (which decrease their marketing and communication costs) to reach their customers.

Phishing attacks are accomplished in several ways. Many of these attacks use an amalgam of technologies. For example, phishers can initiate an attack by sending e-mail messages that contain URLs of compromised and or counterfeit websites in bulk. These websites, in turn lure those who respond into revealing confidential information. In phishing attacks, the compromised websites may also prompt the user to install software (for example, missing plugins.) During the installation process, the phisher's website installs malware on the user's computer.

Phishers can also record keystrokes and monitor the respondent's display by using malware such as keyloggers and screenloggers. Another commonly increasing attack is content-injection phishing. Phishers insert malicious content into a legitimate website which in turn redirects the data to a phishing server. Apart from these methods, attacks such as session hijacking, hosts file [3] poisoning, DNS-based (Domain Name System) phishing and search engine phishing are also popular among phishers.

2   Lack of Success

In spite of aggressive efforts (refer to Figure 1), technology companies have had limited success in restricting phishing attacks. Unfortunately the nature of phishing attacks changed over time from passive such as password guessing and eavesdropping to active attacks such as employing Trojans to intercept traffic and adopting social engineering techniques. The real threat is the fraud that occurs when phishers masquerade as genuine senders. The tactics used in masquerading change according to the anti-phishing techniques being developed.
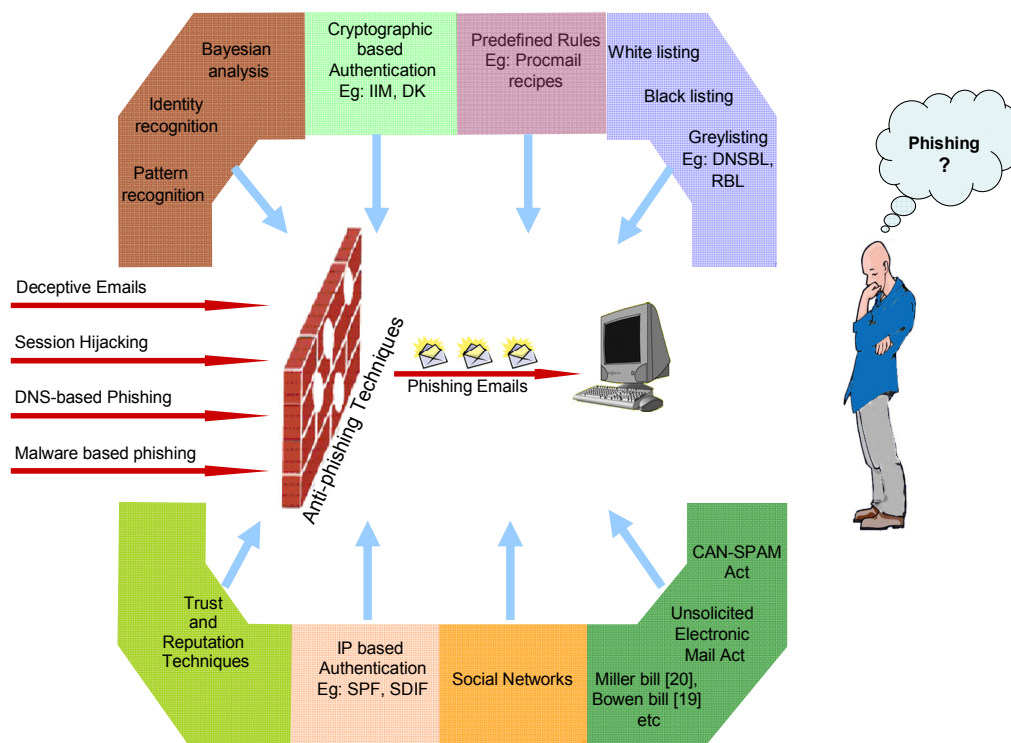
Fig. 1. Though a large number of anti-phishing solutions exist, there has been no recess in the number of phishing attacks. Phishing attacks have already reached a double-digit rate by the end of 2005.

According to a survey [11] conducted by Cipher Trust Inc., an e-mail security company, spammers adopted authentication techniques such as Sender Policy Framework (SPF) ("whereby a domain may explicitly authorize the hosts that are allowed to use its domain name, and a receiving host may check such authorization" [5]) or Sender ID Framework (SIDF) ("a mechanism by which servers can determine what e-mail address is allegedly responsible for most proximately introducing a message into the Internet mail system, and whether that

introduction is authorized by the owner of the domain contained in that e-mail address" [6]) at a faster rate than legitimate users. Out of two million e-mails sent to the Cipher Trust customers, only 5 percent of the incoming e-mails had valid SPF or SIDF records. Within this 5 percent the percentage of spammers having SPF or SIDF records is slightly more than the percentage of the legitimate e-mail senders. Of the legitimate e-mails, 2.8 percent passed the SPF or SIDF checks as compared to 3.8 percent of spam e-mails. Phishers use innovative methods to counter the authentication techniques. As of now it is unclear how many spammers with valid SPF or SIDF records phish. However, if spammers can get through these authentication techniques, phishers may also obtain valid SPF or SIDF records. Moreover the average life span [2] of a phishing website is 5.0 days or even shorter, making any anti-phishing solutions based on blacklists very ineffective.

Phishers are commonly misinterpreted as being amateurs. On the contrary, phishers are technically innovative and are professional criminals [22]. Their phishing attacks blend technologies with a high degree of sophistication. Most phishing e-mails do not contain enough significant content for content-based filters to work on. Sometimes the e-mail content can just be the URL of a counterfeit website. A majority of the solutions proposed for phishing to date have offered back-end solutions. For example, investigators use digital fingerprinting techniques to extract and maintain a database of tokens from purported phishing websites. They then use these databases of tokens in generating a series of signatures and hashes to identify unique portions of the content. The authors of these techniques claim that a pattern exists for every phisher and try to identify a pattern of phishing in suspicious websites. When the technique detects a phishing pattern, the website receives a phishing label. This might provide a good back-end solution, but such techniques not suffice for real time detection. *We need real-time solutions. Phishing attacks should be restricted before they cause any damage. We need to warn users before they open their e-mails.*

We based our solution on the trustworthiness of the relays participating in the relaying of the e-mails. We examine the e-mail's header rather than the content. We further classify phishers into serial phishers, prospective phishers, recent phishers and suspects based on their phishing intensity and distrust value.

3   Proposed Methodology

E-mail provides the primary vector for the phishers; the majority of the phishing attacks are initiated using e-mails. Though phishers try to counterfeit the websites, they do not have access to all the fields in the e-mail headers. They can forge certain fields such as, inserting spurious "Received:" header lines before

dispatching e-mails but cannot spoof the complete header. Our classification method examines the header of an e-mail, the social network of the recipient, *wantedness* and *unwantedness* of the e-mail's source. In particular, we analyze the trust and reputation of the contents in the header.

Normally recipients associate their priorities by performing certain actions on the e-mails. For example, reading and deleting, deleting without reading and archiving. These actions generate an implicit feedback such as storage time of read and unread e-mails of any given sender. In addition to learning from implicit feedback, we also base our decision on the explicit feedback from the recipient.

The major factors in any phishing attack are **forgery** and **social engineering**. No matter how many authentication techniques we develop, phishers always adapt. However, phishers cannot counterfeit or become part of the recipient's social network without consent. Considering such issues, we designed our classifier to perform the following analyses in four steps: i) DNS-header analysis, ii) Social network analysis, iii) *Wantedness* analysis and iv) Proactive classification of phishers.

4   Traffic Profiles of the Corpus

Although there are some benchmark corpuses available, two reasons make them inadequate for the analysis conducted here. *First, all the available e-mail corpuses do not contain original headers. This was done to anonymize the recipients. Second, the corpuses in the existing benchmarks such as TREC [15] spam filter evaluation tool kit contains only spam and non-spam e-mails. We need a mix of phishing, and legitimate e-mails for our analyses.* Hence, we tested our methodology on live e-mail corpuses with the users' permissions. Corpus used in this paper consists of 13,843 e-mails collected over 2.5 years. This corpus has a mix of legitimate, phishing, telemarketing, and opt-in e-mails. In addition, our methods were tested with other corpuses from real-people and arrived at similar results.

5   Architecture Model of the Proposed Classifier

As seen in Figure 2 our classifier performs three analyses on the incoming e-mails. As a first step, we undertake, DNS-header analysis, where the e-mail headers are verified for spoofing by performing DNS-lookups. In this step, our classifier determines whether a sender is a phishing suspect or non-phisher. The outputs from this step (phishing suspects and DNS-lookup passed e-mails) are passed to the social network analysis (Step 2). We use social network analysis to categorize the e-mails into i) socially close and ii) socially distinct. In Step 3, we

identify the falsely classified senders. The outputs from Step 3 are passed onto Step 4. Our classification engine classifies phishers into i) Recent Phishers, ii) Serial Phishers, iii) Prospective Phishers and iv) Suspects.
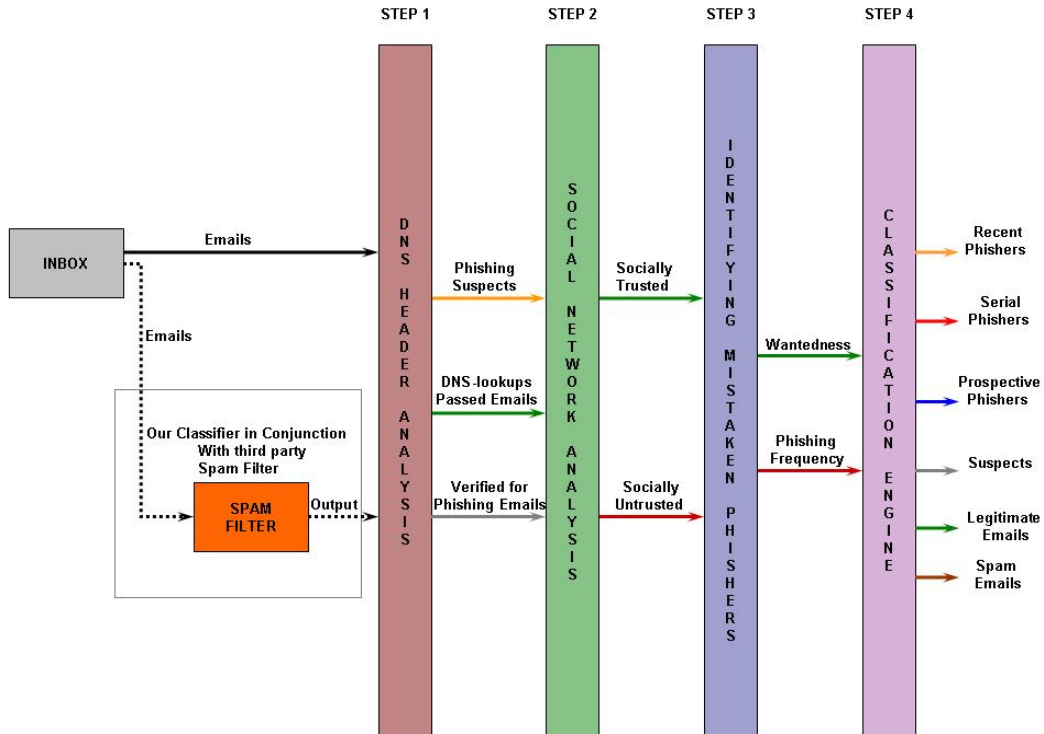

Fig 2. Architecture for phishers classification.

Our classifier can be used in conjunction with the third party spam filters such as, SpamAssassin for filtering spam and phishing e-mails (see Figure 2). First, the outputs from the spam filter i) spam and ii) non-spam e-mails are verified for phishing in Step 1.

5.1 Step 1: DNS-Header Analysis

This analysis identifies spoofing in the e-mail headers in two stages: i) Header validation and ii) Authenticity verification. We validate information such as, hostnames of the i) senders, ii) mail servers, and iii) relays recorded in the header during the SMTP authorization process. From our experiences with phishing e-mails, we found a majority of phishers spoof the hostname or domain name in the "Received:" lines in the e-mail headers. For instance, any phishing e-mail posing to have originated from PayPal will specify the e-mail ID in "Return-path:" field

as somename@paypal.com. Phishers commonly do this to circumvent any filtering process based on the sender's e-mail IDs and to make their phishing e-mails more convincing to the receiver, thus achieving their initial goal of making the receiver read their e-mails. Our assumption is, if an e-mail claims to be from a particular domain, at least the sender's or the mail server's IP address should belong to that domain. Any mismatch of the IP addresses will result in suspecting that e-mail as phishing. We performed DNS lookups on the host/domain names specified in the "Received:" header lines and match the returned IP address with the IP addresses recorded by the relays during the SMTP authorization process. We label senders who fail the DNS lookups as phishing suspects and move to the next step.

5.2 Step 2: Proximity Based on Social Network

Global software companies such as Microsoft, and Lotus research group developed products that can perform social network analysis and use this information to identify the messages from socially important people. In the following paragraphs we present a brief description of such products and compare them with our social network analysis.

In this analysis we utilized the information obtained by analyzing the recipient's *"sent"* e-mail folder. We calculated social proximity between senders and receivers and use this information to filter false positives and false negatives from DNS-header analysis [18]. Proximity depends on the number of transactions between sender and receiver. For example, the higher the number of outgoing e-mails, the higher the proximity between sender and the receiver. Table I lists various parameters used by different methods for quantifying the relationship between senders and receivers. Proximity information was used for classifying the senders into i) socially close and ii) socially distinct categories.

Table I. Information used by different methods for finding the social relationship

Social Information	A$_1$	A$_2$	A$_3$	A$_4$
Number of outgoing e-mail transactions to a sender	✓			✓
Number of replies to a sender	✓	✓		✓
Number of read e-mails from a sender	✓			✓
Number of unread e-mails from a sender	✓			✓
Number of e-mail conversations initiated by the sender	✓			✓
Number of e-mail conversations initiated by a sender in which the recipient is an active participant	✓			✓
Number of e-mail conversations initiated by a sender in which the recipient is not an active participant	✓			✓
Number of e-mails sent to a sender	✓		✓	✓
Most or least number of replies from the user	✓		✓	✓
Correspondents included in the most or least e-mail conversations by the user	✓			✓
The number of mailing lists the user has subscribed to				✓
The number of carbon-copied messages		✓		
The importance of a message is inversely proportional to the number of recipients it is being dispatched to		✓		✓
E-mail arrival time				✓
Storage time of read and unread e-mails				✓
Interval between successive e-mails from a sender				✓
Time taken by the user to respond to a specific sender			✓	
Time spent in composing messages			✓	
Time spent in reading messages			✓	
E-mail reciprocity			✓	✓
Familiarity to the user's community				✓
Path traversed by the e-mails (example: trusted/untrusted paths)				✓

A$_1$: SNARF (Social Network Relationship finder) [12] A$_2$: Bifrost (Inbox organizer) [13]
A$_3$: DriftCatcher (E-mail client) [14] A$_4$: Our Classifier

5.3 Step 3: Trustworthiness Based on Recent E-mails

This step detects falsely classified e-mails. For this purpose, we measure parameters associated with the senders and recipients such as senders trust, and recipient's *wantedness* of the senders' e-mails. For example, we believe a sender's trustworthiness depends upon the nature of recent transactions. In this step, we analyze the nature (spam/legitimate) of the e-mails recently sent to the recipient. A sender's trustworthiness is proportional to the length of the interval between successive legitimate e-mails. That is, the longer the time interval between a sender's legitimate e-mails, the less the recipient interested in that

sender. If the previous e-mails are fraudulent, the new incoming e-mails from that sender are also less trustable and hence, *distrust* is high. Where as, if the recent e-mails are legitimate, then new e-mails from the sender are more trustable. We measured *trustworthiness* of e-mails in socially close and socially distinct categories obtained from the social network analysis.
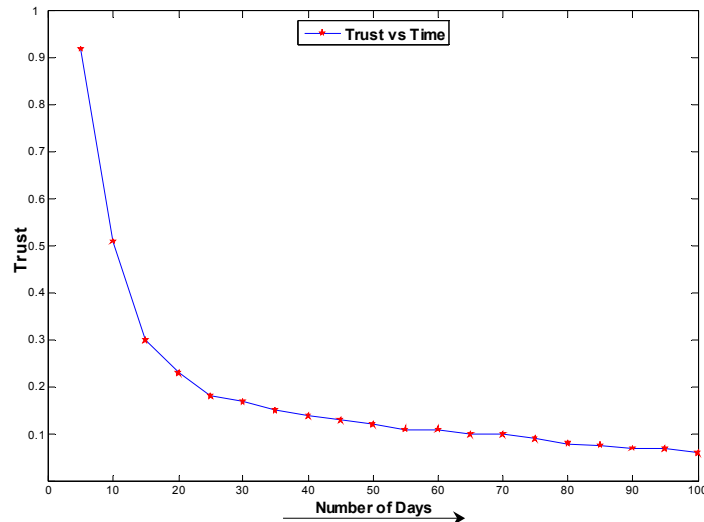

Figure 3: Decrease in trust of a sender (ebay.com) over time. The average time interval $\Delta T_{legitimate\_e\text{-}mails}$ between the most recent e-mails (spam/legitimate) and the last legitimate e-mail indicates the periodicity of legitimate e-mails. As time lapses, $\Delta T_{legitimate\_e\text{-}mails}$ increases, and there is a higher chance of sender's recent e-mails being spam or fraudulent in nature. Hence trust on a sender decreases as $\Delta T_{legitimate\_e\text{-}mails}$ increases. This decrease in trust initially may be small but as $\Delta T_{legitimate\_e\text{-}mails}$ increases, trust decreases exponentially.

Figure 3 shows a decrease in trust of a phisher whose e-mails were disguised as legitimate e-mails from ebay.com (based on the Corpus). In this group of e-mails, the header of the first e-mail was not spoofed. Due to this, DNS header analysis classified this e-mail as legitimate, thus initially giving high trust to this e-mail. As time progressed, more and more spoofed e-mails were received from this sender. We computed trust for all the suspects and resolved the false positives.

5.4 Step 4: Classification of Phishers

We classified phishers using their severity of their actions. Based on the distrust value and phishing frequency[*], phishers from the socially untrusted category are further classified into i) serial phishers (the most severe case), ii) recent phishers, iii) prospective phishers, and iv) suspects. Figure 4 illustrates this classification.
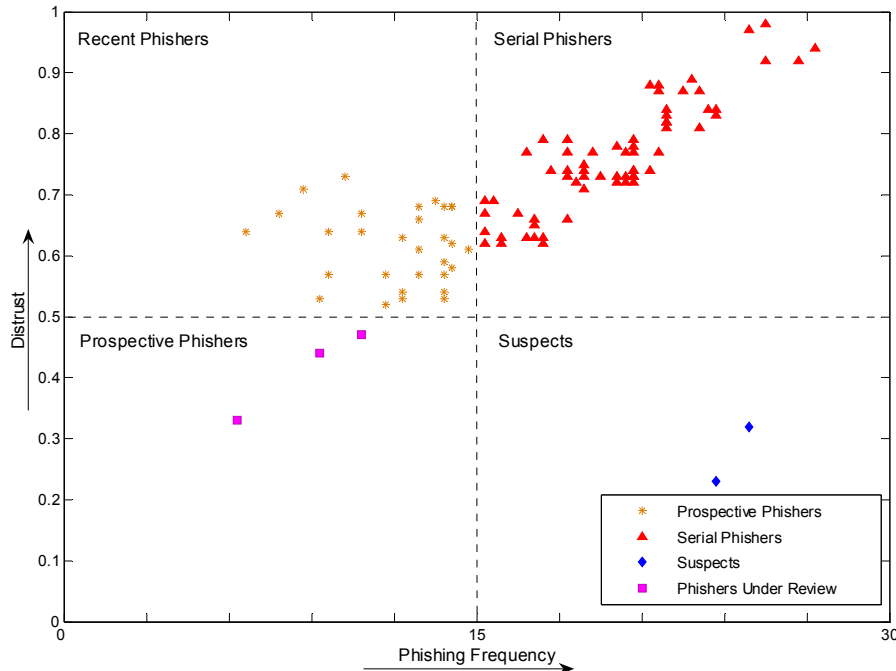

Fig. 4. Distrust value vs. Phishing frequency

Calculation of Precision

We calculated the hit rate, also known as the true positive rate "tp$_{rate}$", as the ratio of true positive instances to the total number of positive instances that were originally categorized by handlabeling. We achieved high hit rates for categorizing phishers into serial (98.37%), recent (89.87%), prospective (96.54%) and suspects (92.13%).

---

[*] The phishing frequency is the number of phishing e-mails originated from the phisher or his domain.

6   Comparison with the Existing Techniques

While each of the current solutions (Two-factor authentication, Mutual authentication, E-mail authentication, Anti-phishing browsers) address a portion of the problem that phishing creates, none of them addresses a sufficient number of factors to effectively combat phishing. Any solution must address issues such as identifying spoofing, providing accurate identification of the severity of the risk, interoperability and cost effective. In Table III, we provide a comparison of proposed classifier with the existing anti-phishing solutions. A list of factors that have a high impact on phishing classification is presented in Table III. As it can be seen, the proposed classifier considers the majority of these factors while the other existing phishing techniques consider only a small subset of these factors.

It should be noticed that a quantitative comparison between our approach and the other techniques is not possible. First, authentication techniques (A_1, A_2, and A_3) do not prevent or warn the user of a phishing e-mail. They try to prevent the phishers to extract the confidential information by means of some authentication. The importance of authentication would decrease if we can prevent the recipient from receiving the phishing or at least warn him/her that the e-mail is potentially a phishing message. Also anti-phishing browsers are based on black lists and as stated before phishing sites have a short life span (around 5 days) that considerably decreases the effectiveness of such anti-phishing techniques.

Table II Comparison of our classifier with existing anti-phishing techniques

Anti-phishing Techniques	A$_1$	A$_2$	A$_3$	A$_4$	A$_5$
URL Analysis				✓	
E-mail headers Analysis			✓		✓
Content Analysis			✓		
Two-way authentication between client and server		✓			
Timestamps and sequence numbers	✓	✓			✓
Public keys	✓	✓			
Two factors, something you know for example, passwords and something you have such as phrase and authenticating picture	✓	✓			
DNS lookups			✓		✓
Multi-variate Analysis					✓
Social context of the sender					✓
Authentication services with reputation service			✓		✓
Classifies phishers proactively				✓	✓
Adaptive filtering					✓
Blacklists and whitelists				✓	✓
Interoperability with third party authentication mechanisms					✓
Cost effective implementation					✓

A$_1$: Two-factor authentication [4]　　　　A$_2$: Mutual authentication
A$_3$: E-mail authentication [5][6][7][8]　　A$_4$: Anti-phishing browsers [16]
A$_5$: Our Classifier

7   Conclusions and Future Phishing Venues

We applied our methodology on a live corpus of 13,843 e-mails collected over 2.5 years. After analyzing the corpuses, we were able to separate the phishing e-mails from legitimate e-mails. Our classifier performed comparatively better than existing anti-phishing solutions by detecting 99% of the non-legitimate traffic accurately in both the corpuses. In addition, we tested our methods with other corpuses from real-people and arrived at similar results. We introduced the concept of *trustworthiness* of the senders/sending domains. Majority of the phishers use special softwares for sending phishing e-mails which record only two IP addresses (a spoofed sender's IP address and a legitimate destination IP

address) in the header of the message. This makes it appear as if an e-mail reached its destination in a single hop. These e-mails are successfully classified during DNS-header and social network analyses.

Our classification resulted in significantly fewer false positive and false negative rates. However, there are a few cases where the header information is not sufficient to classify e-mails. Cases like, spammers using compromised relays for mailing and phishers mailing from trusted domains. Normally it is not possible to compromise the entire path traversed by the e-mails. Only those relays which lie in the beginning of the path are susceptible to being commandeered. (This is also true in cases of zombies or botnets.) In situations like these we give more importance to the senders having a significant prior communication history. In both cases, initially few phishing e-mails may get through. Since our classifier has a continuous-learning mechanism from both implicit & explicit user feedbacks, it overcomes these issues by learning from the inferred reputation information. Our classifier can be used in conjunction with the existing spam filtering techniques to restrict spam and phishing e-mails from reaching the recipient's inbox.

Text, voice, and video are three modes in human communication. Each mode conveys an added level of trust to consumers. With the advent of technologies like voice over IP and video over IP, the differences between online and face-to-face communications are reduced. E-mails have already fallen victim to phishing attacks. It won't be long before phishers pose as a serious threat to VoIP and Video over IP. In fact, there have been reports confirming phishing attacks (loosely termed as vishing [10]) using VoIP already. By the time VoIP and Video over IP are widely deployed, technology companies must make them as impervious as practicable to spam and phishing attacks.

Internet is highly anonymous in nature. In such an environment, we cannot rely on the technologies that cannot provide a high level of security. Phishing e-mails may not be successful, ultimately, in luring the majority of the Internet users, but, imagine the degree of believability a phishing video call delivers. In the case of video phishing, phishers may more effectively convince the users by displaying forged credentials such as fake IDs of target companies. The added problem with phishing attacks is that they can happen in forms we least expect and there is no one–size-fits-all solution. We need real-time solutions for phishing. But, real-time detection of phishing requires more than content processing and checking the backend databases for forged URLs. We need to proactively detect phishers based on their traffic paths, traffic patterns, and on receivers' social interaction with the senders.

References

1. Anti-Phishing Working Group. http://www.antiphishing.org/
2. Phishing Trends Report Available.
   http://www.antiphishing.org/reports/apwg_report_apr_06.pdf
3. Dave Houde and Tim Hofamn, "TCP/IP for Windows", Prentice-Hall, 2001.
4. Schneier, B. 2005. Two-factor authentication: too little, too late. Commun. ACM 48, 4 (Apr. 2005), 136.
5. Wong M and Schlitt, W, "Sender Policy Framework (SPF) for Authorizing Use of Domains in E-Mail, Version 1"
   http://www.ietf.org/rfc/rfc4408.txt.
6. Sender ID: Authenticating E-Mail. http://xml.coverpages.org/draft-ietf-marid-core-03.txt.
7. Fenton, J. and Thomas, M, "Identified Internet Mail", Cisco Systems Inc, May 13, 2005.
   http://www.identifiedmail.com/draft-fenton-identified-mail.txt.
8. Delany, M, "Method and system for authenticating a message sender using domain keys", US Patent, 6,986,049, January 10, 2006.
9. H. Schulzrinne, Jonathan D. Rosenberg, "Signaling for Internet Telephony," icnp, p. 0298, Sixth International Conference on Network Protocols (ICNP'98), 1998.
10. Kelly, J, "Vishing' Attacks use VoIP",
    http://www.darkreading.com/document.asp?doc_id=98787
11. Roberts, P, "Spammers use sender authentication too, study says they've adopted the technology faster than legitimate e-mail senders", IDG News Service, 2004, August 31.
    http://www.computerworld.com/printthis/2004/0,4814,95617,00.html
12. Carman N, Bernheim B, Smith, A. J, Marc A. and Fisher, D, "Social network Relationship Finder", Proceedings of the Second Conference on E-mail and Anti-Spam (CEAS), 2005. http://www.ceas.cc/papers-2005/149.pdf
13. Bälter O, Sidner C L, "Bifrost Inbox Organizer: Giving users control over the inbox", Proceedings of the second Nordic conference on Human-computer interaction Aarhus, Denmark Year of Publication: 2002
14. Lockerd A and Selker Ted, "DriftCatcher: The Implicit Social Context of E-mail", Proceedings of Human-Computer Interaction INTERACT 2003. Text Retrieval Conference (TREC). http://trec.nist.gov/
15. Festa, P, "Netscape readies anti-phishing browser".
    http://news.com.com/Netscape+readies+antiphishing+browser/2100-7355_3-5558006.html

16. POBox, http://www.pobox.com/
17. Srikanth Palla, "E-mail Spam Detection using SMTP Path and Relay Analysis", MS Thesis, Department of Computer Science, University of North Texas, 2006.
18. California Assembly Bill 1676 –
    http://ca.rand.org/statebulls/bulletins/statebull109j.html
19. California Assembly Bill 1629 –
    http://www.elite.net/elite/services/ca1629-2.html
20. Jeremy Kirk, "PayPal phishing plea to email providers; eBay's service ask email tools to block messages", PC Advisor,
    http://www.pcadvisor.co.uk/news/index.cfm?NewsID=8827
21. Michael Cohn, "Phishing Attacks Linked To Organized Crime", Bank Systems & Technology, July 20, 2004,
     http://www.banktech.com/news/showArticle.jhtml?articleID=23902354