Detecting Phishing in Emails

Srikanth Palla and Ram Dantu, *University of North Texas, Denton, TX*

Abstract— Phishing attackers misrepresent the true sender and steal consumers' personal identity data and financial account credentials. Though phishers try to counterfeit the websites in the content, they do not have access to all the fields in the email header. Our classification method is based on the information provided in the email header (rather than the content of the email). We believe the phisher cannot modify the complete header, though he can forge certain fields. We based our classification on three kinds of analyses on the header: DNS-based header analysis, Social Network analysis and Wantedness analysis. In the DNS-based header analysis, we classified the corpus into 8 buckets and used social network analysis to further reduce the false positives. We introduced a concept of wantedness and credibility, and derived equations to calculate the wantedness values of the email senders. Finally, we used credibility and all the three analyses to classify the phishing emails. Overall, our method resulted in far less false positives. We plan to perform two more analyses on the incoming email traffic, i) End to End path Analysis, by which we try to establish the legitimacy of the path taken by an email; ii) Relay Analysis, by which we verify the trustworthiness and reputation of the relays participating in the relaying of emails. We like to combine all the methods i) Path analysis, ii) Relay analysis, iii) DNS-based header analysis and iv) Social Network analysis for developing a stand alone email classifier, which classifies the incoming email traffic as i) Phishing emails ii) Socially Wanted emails (Legitimate emails) iii) Socially Unwanted emails (Spam emails). We believe this research can be extended for VoIP spam analysis.

Index Terms— **DNS-lookups, SMTP-authorization, Tolerance, Credibility, Wantedness.**

1 INTRODUCTION

Email spammers can be categorized based on their intent. Some spammers are telemarketers, who broadcast unsolicited emails to several hundreds/thousands of email users. They do not have a specific target, but blindly send the broadcast and expect a very limited rate of return. The next category of spammers comprise the opt-in spammers, who keep sending unsolicited messages though you have little or no interest in them. In some cases, they spam you with unrelated topics or marketing material. Some of the examples are conference notices, professional news or meeting announcements. The third category of spammers is called phishers. Phishing attackers misrepresent the true sender and steal the consumers' personal identity data and financial account credentials. These spammers send spoofed emails and lead consumers to counterfeit websites designed to trick recipients into divulging financial data such as credit card numbers, account usernames, passwords and social security numbers. By hijacking brand names of banks, e-retailers and credit card companies, phishers often convince the recipients to respond. Legislation can not help since a majority number of phishers do not belong to the United States. In this paper we present a new method for recognizing phishing attacks so that the consumers can be vigilant and not fall prey to these counterfeit websites. Based on the relation between credibility and phishing frequency, we classify the phishers into i) Prospective Phishers ii) Suspects iii) Recent Phishers iv) Serial Phishers.

1.2 PROBLEM DEFINITION

Current spam filters focus on analyzing the content of the email and marks the spam emails as *BULK* and expects the recipient to make a decision[1]-[16], [20]. *There are two problems with this approach. First, it is very difficult for the not-technical-savvy consumers to verify the authenticity of the source and second, the phishers always devise new content to bypass the spam filters. Moreover, many filters available in the market have high degree of false positives (this means labeling legitimate emails as spam).* Hence, consumers are worried about losing a legitimate email (e.g.,missed job opportunity, sale or some other important transaction). We need a solution for phishing which should not depend on content analysis but consider other aspects of the email. Our classification method is based on the information provided in the email header (rather than the content of the email). We believe the phisher cannot modify the complete header, though he can forge certain fields.

1.3 BACKGROUND WORK

Research on header analysis was recently reported by Microsoft, IBM, and Cornell University in the 2004 and 2005 anti-spam conferences [17][18]. Goodman [17] outlines that IP addresses present in the email headers are the most important tools in fighting spam. Barry Lieba's [18] filtering technique make use of the path traveled by an email, which is extracted from the email header. They have analyzed the end units in the path but not the end-to-end path. They claim their approach complements the existing filters but does not work as a stand alone mechanism. Papers [17] [18] have used the information provided in the email header for classifying the emails as spam

and non spam. Christine E. Drake [19] in his paper, “Anatomy of Phishing Email”, discusses the tricks employed by the email scammers in phishing emails. Much research is being done in spam filtering, but only a few papers [19] have discussed about the threat of phishing attacks. We believe there is a lot of research yet to be done in this area.

The rest of the paper is organized as follows. In section 2, we describe how an email is processed and relayed to the destination. We also discuss the profile of the email traffic used in our experiments, followed by our methodology and the deployment of our classifier. In section 3, we present our architecture model and describe the DNS-based header analysis, Social Network analysis and Wantedness analysis, *used to categorize an email.* Section 4 contains the classification of phishers and a summary of our results. Section 5 contains the related work done in non content based analysis for spam filtering. Section 6 concludes the paper with a summary of the major contributions of our paper.

2 EXPERIMENTAL SETUP

There are three main phases through which an email passes before reaching the recipient, i) MUA (Mail User Agent) ii) MTA (Mail Transfer Agent) and iii) MDA (Mail Distribution Agent). These are the three main parts in the email environment. MUA is the program used by an email user to send and receive emails. Examples of an MUA are Thunderbird, Microsoft Outlook etc. MUA reads the incoming emails that have been delivered to the recipient’s mailbox, and passes the outgoing messages to an MTA for dispatching. The MTA acts as a "mail router". It accepts a message passed to it either by an MUA or another MTA, and passes it to the appropriate MDA for delivery. MTA is normally a mail server software like sendmail, postfix, qmail etc. The MDA accepts emails from an MTA and performs the actual delivery. Examples of MDA are mail.local, procmail etc.

MTA is the most important of the three agents. It is responsible for making intelligent decisions about an email transfer. It may not actually perform the delivery itself; it is the part which tells the other agents how to interact and what to do. Figure 4 illustrates the whole process of email transfer from the sender to the recipient. Sender's MUA (Thunderbird, Outlook, etc) on the sender's machine passes the message to the MTA (sendmail) on the local host. The MTA (sendmail) notices that the message is addressed to the recipient who is at another domain. It will reach the recipient domain via SMTP and passes the message to the local MDA. The local MDA connects to the MTA on the recipients’ domain and hands over the message. The MTA on the recipient’s domain notices that the message is addressed to a user on the local host, so it passes the message to the local MDA. The local MDA saves the message in the recipient's mailbox. The next time the recipient logs on to his machine and runs his MUA, the message is there waiting for him to read.

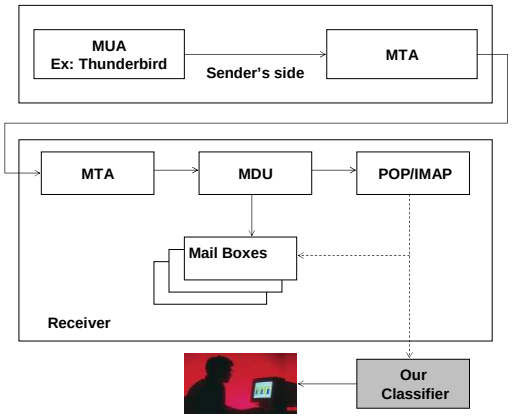

Figure 1 Various phases of email transactions

We deployed our classifier on the recipient’s local machine running an IMAP proxy and Thunderbird (MUA). All the recipient’s emails were fed directly to our classifier by the proxy. Our classifier periodically scans the user’s mailbox files for any new incoming emails. DNS-based header analysis, Social Network analysis and Wantedness analysis were performed on each of the emails. The end result is the tagging of emails as Phishing, Opt-outs, Socially Close or Socially Distinct.

2.1 TRAFFIC PROFILE

The existing corpuses do not contain original headers (this is done to hide the identity of the email recipients) and since our methodology is based on email headers, it is not possible to use the existing benchmarks for our analysis. Moreover, a majority of the benchmarks have spam emails but do not contain legitimate emails. So, we used live corpuses with the user’s permission. Figure 2 describes a live corpus of 13,843 emails (collected over 2.5 years). This corpus has a mix of legitimate, spam and phishing emails. Different categories of the emails are shown in the Figure 3.

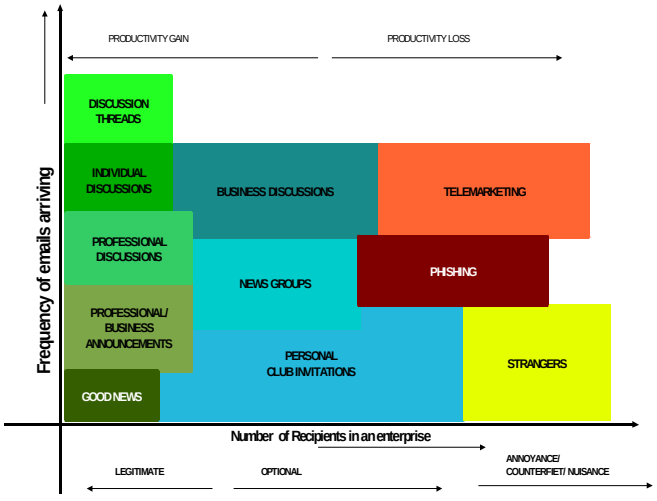

Figure 2 Email traffic profile

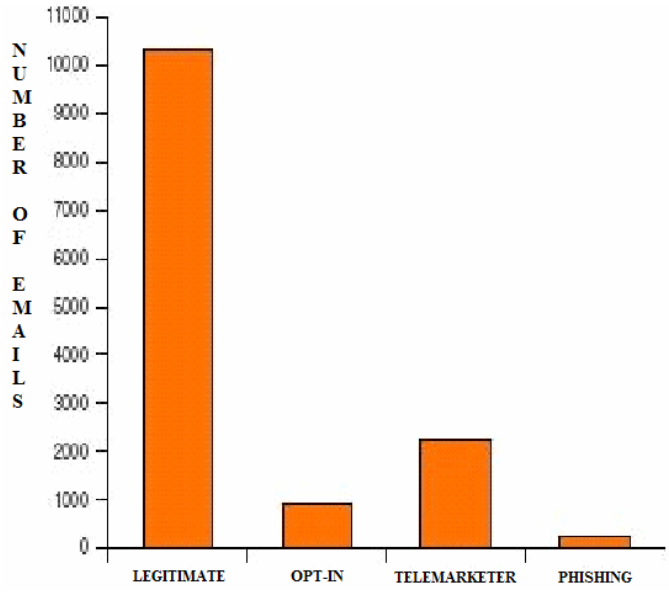

Figure 3 Corpus traffic profile

In this paper, we classify the phishing emails based on four steps. Step 1: DNS-based header analysis for verifying the legitimacy of the email source. Step 2: Social Network analysis to reduce the false positives. Step 3: Wantedness analysis to calculate the *credibility* and *wantedness* of the senders. Step 4: We used the relation between fraudulency and the phishing frequency for classifying the phishers. We used credibility and wantedness of the senders in trusted domains for classification. *One advantage of our technique over existing techniques is that our analysis can be tailor made to a recipient's email activity*. In the sections to follow, we describe the four steps in detail.

2.2 METHODOLOGY

We are working on an innovative methodology for isolating the phishing emails. Our method examines the header of the email, the social network of the recipient, wantedness and unwantedness of the email's source. In particular, we analyze the trust and reputation of the contents of the header. In addition, we also base our decision based on the actions of the recipient (e.g. spam emails deleted without reading), *sent* folder, and implicit feedback from the recipient's social network (e.g., co-workers, friends).

3 ARCHITECTURE

The architecture model of our classifier consists of three analyses: DNS-based header analysis, Social Network analysis, and Wantedness analysis. In DNS based header analysis, we validate the header and perform DNS lookups on the hostnames provided in the *Received:* header lines and classifies the emails as trusted or untrusted. The resultant is again treated with social network rules for further classification. We calculate wantedness values for each sender. The outputs after these three analyses are sent to a *classifier*, where, based on the fraudulency of the senders and their phishing frequency (measure of the number of phishing emails that originated from the sender or his domain), the final classification is made. DNS-based header analysis, social network analysis, and wantedness analysis are provided with continuous user feed back, making them completely configurable as per the user settings.

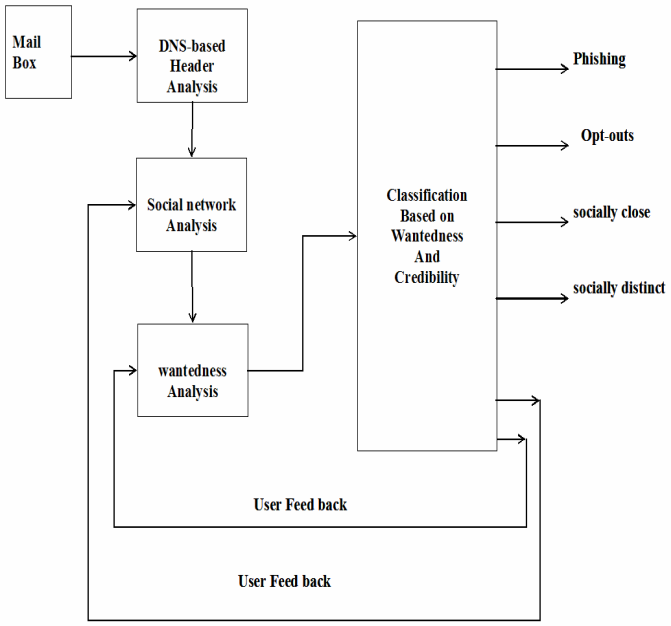

Figure 4 Architecture diagram

3.1 STEP 1: DNS-BASED HEADER ANALYSIS

Stage 1: Header Analysis: In this analysis, we validate the information provided in the email header for the following: i) the hostname, ii) position of the sender, iii) mail server and iv) the relays in the path. We divide the entire corpus into two buckets: i) emails, which can be validated by a DNS look up and ii) emails which cannot be validated by a DNS lookup (due to a lack of proper information in the header). Bucket1 contains all the emails which have proper hostname provided in the *Received:* lines in the email header and bucket2, contains all the emails which have their hostname either improperly configured or set to a format which is not suitable for a DNS lookup.

Stage 2: We manually verified many phishing emails, and found out that a majority of the phishers spoof the hostname or domain name in the *Received:* lines in the email headers. From our observations, in some cases, they do not even provide any name or IP address. So, our next step involves doing DNS lookup on the hostname provided in *Received:* lines in the header and matching the IP address returned, with the IP address which is stored next to the hostname by the relays during the SMTP authorization process. We performed these lookups on bucket1, resulting in a further subdivision of the bucket1 into two buckets. i) a trusted bucket and ii) an untrusted bucket. The trusted bucket contains all the non-spoofed emails and the untrusted bucket contains spoofed

emails along with some legitimate emails which failed in the lookup process. We forward bucket2 and both the trusted and untrusted buckets to the next step for a Social Network analysis.

3.2 STEP 2: SOCIAL NETWORK ANALYSIS

Each of the three buckets bucket2, untrusted bucket and trusted bucket are treated with the rules, built by analyzing the "sent" folder emails of the receiver. *These rules can be built as per the recipients email filtering preferences.* For example, we used the following three rules in this analysis. We analyze the "*sent*" folder and create a list of trusted domains from the email ids of the people to whom the receiver has replied or ever sent an email. This list is used in formulating the following two rules for further analysis.

Example Rule 1: All the senders in the sent folder will be removed from the untrusted bucket*:** This rule matches the domain name constructed from the email id provided at the *Return-path:* field in the emails with the domain names in the list extracted from the "*sent*" folder of the receiver. When bucket2, untrusted bucket and trusted bucket from the DNS based analysis are treated with this rule, it results in a further subdivision of each of these three buckets into socially trusted and socially untrusted buckets. Socially trusted buckets contain all the emails filtered out due to the ***Rule 1. Socially untrusted buckets contain the emails which fail to comply with ***Rule 1***.

Example Rule 2: Familiarity with the sender's community: This rule uses the *Return-path:* field provided in the email header. It is derived from the domain name of the email id in the *Return-path:* field and tries to match with the domain names specified in the path of the emails. All the emails in the socially untrusted buckets from ***Rule 1*** are treated with this rule, resulting in a further subdivision of these socially untrusted buckets into socially trusted buckets and socially untrusted buckets. Socially trusted buckets contain the emails filtered out due to ***Rule 2***, where as, socially untrusted buckets contains the remaining emails which fail to comply with ***Rule 2***.

Example Rule 3: Familiarity to the path traversed: This rule filters out the emails which match the "*sent*" folder emails of the receiver. This rule further splits the socially untrusted bucket (after passing through the second rule) into socially trusted buckets and socially untrusted buckets. Socially trusted buckets contain all the emails filtered out due to this rule and socially untrusted buckets contain remaining emails which fail to comply with this rule.

Socially trusted buckets and socially untrusted buckets thus obtained are classified as trusted emails and untrusted emails, which are further, classified into prospective phishers, suspects, and serial phishers.

3.2.1 CLASSIFICATION OF TRUSTED AND UNTRUSTED DOMAINS

Figure 5 shows how the DNS based analysis and Social Network analysis are performed on the email corpuses. In DNS based analysis, the corpus is subdivided into two buckets, valid DNS-lookup bucket of size 11968 emails and invalid DNS-lookup bucket of size 1875 emails. Our hypothesis is that, in phishing emails, the phisher will always provide his email id in the Return-path: field in the header, at the domain of whom he claims to be mailing from. For instance, any phishing email posing to have originated from Paypal, will carry the email id at Return-path: field as somename@paypal.com. Phishers commonly do this to circumvent any filtering process based on the sender's email ids and also to make their phishing emails more plausible to the receiver, thus achieving their primary purpose of making the receiver read their emails.

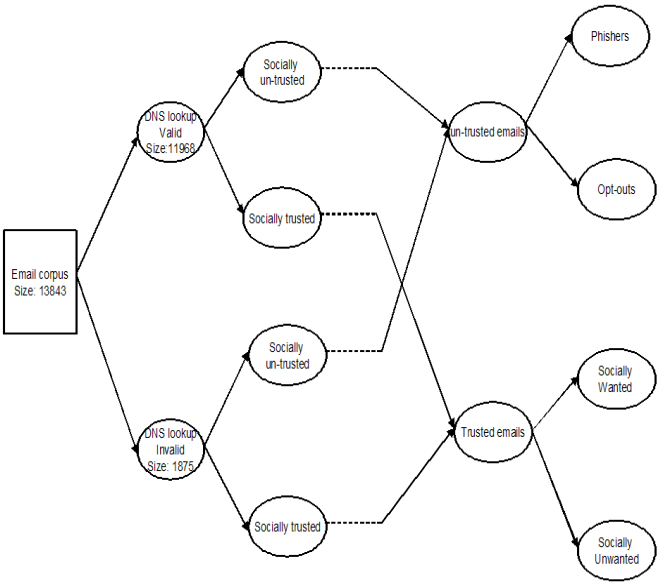

Figure 5 Email corpus classifications

We construct the domain name from the email ID provided at the *Return-path:* field, perform DNS lookup on that domain and try to match the returned IP address with the IP address present in the path. Our assumption is, if an email 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 marking that email as untrusted and that email will be placed in the untrusted bucket. All the emails that match their IP addresses will be placed in the trusted bucket. There are some legitimate emails in the untrusted bucket which fail to match their IP addresses; this may be due to the change in IP addresses of the domain at which they hold their email accounts. *This is possible in our case because we collected our corpus over a period of 2.5 years and the domain could have started using a dynamic IP addressing policy.* Such cases are taken care of by the Wantedness analysis. There are no phishing emails in the

trusted category and no legitimate emails in the untrusted category. The untrusted category now contains emails from phishers, opt-ins and telemarketers.

Based on the results from the DNS-lookup, we divided the corpus into trusted and untrusted categories. Hence, it can be seen that DNS-lookups constitute the core of our method. There is a possibility of these DNS-lookups resulting in failure for some email queries. This is possible because some senders might be using machines which are misconfigured in providing the identification information. Also due to security reasons some organizations deliberately conceal the IP addresses of their machines. If the sender's domain employs NAT policy, the machines behind the NAT provide private IP addresses which are not valid for DNS-lookups. In classifying these kind of senders we analyzed the nature of their recent emails sent to the recipient. If their recent emails were fraudulent, the less credible are the emails from those senders. If their recent emails are legitimate, the emails from them are more credible. We discussed more about the Wantedness analysis in the section 3.3. We computed wantedness values for the senders' domains in both the final trusted and untrusted buckets. Senders having their wantedness above a threshold value in the untrusted bucket are classified as Opt-outs and the senders whose wantedness values are below the threshold value are classified as Phishers. Senders who have a high wantedness value are marked as socially close, where as, senders with low wantedness values are classified as socially distinct in the final trusted bucket.

3.3 STEP 3: WANTEDNESS ANALYSIS

Measuring the sender's credibility (ρ): We believe the credibility of a sender depends upon the nature of his *recent* emails. If the recent emails sent by him are legitimate, his credibility increases where as, if the recent emails from the sender are fraudulent, his fraudulency $(\hat{\rho})$ increases. Credibility of a sender increases with a decrease in the time period $\mathbf{\Delta T}_{legitimate\ emails}$ between his legitimate emails. If the legitimate emails are distributed sparsely from each other in the time domain for a sender, his credibility decreases. This decrease in credibility initially may be small but as the length of the time period $\mathbf{\Delta T}_{legitimate\ emails}$ between his legitimate emails increases, it becomes exponential in nature. Suppose the most recent legitimate email for a sender occurs at time $\mathbf{T_j}$ in his time domain $\{\mathbf{T_i < T_j < T_n}\}$. The decrease in the credibility over the time period $\mathbf{\Delta T = [T_n - T_j]}$ can be estimated as $\rho(i,n) = \rho(i,j)e^{-\Delta T}$.

Larger the time interval between his legitimate emails, more is the decrease in the credibility of a sender. This can be observed in Figure 6.

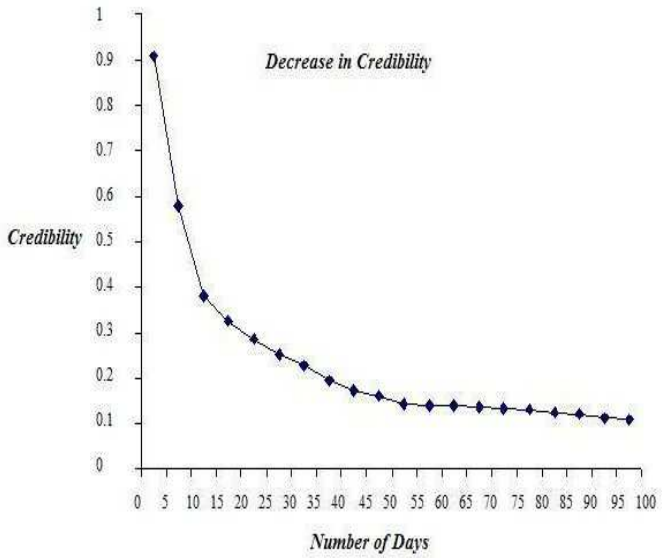

Figure 6 Credibility drop over the time period for a phishing domain

Figure 6 shows a plot of all the phishing emails disguised as legitimate emails from ebay.com. In this group of phishing emails, the header information of the first most email is not spoofed. Though it is a phishing email it resulted in marking of this email as legitimate by DNS-based header analysis, thus giving high credibility to this phisher initially. As time progressed, more and more spoofed emails were received from this phisher. The time period between his legitimate email (first email) and the recent email sent by him increases causing an exponential drop in his credibility. Over a period of time, our classifier will overcome these kinds of issues by learning from the computed credibility drop from the senders' past history and also from the recipients' feed back.

$$\text{Belief}(\tau) \propto \left\{\frac{1}{\Delta T_{\text{legitimate emails}}}\right\} \quad \text{.....(1)}$$

$$\text{Disbelief}(\hat{\tau}) \propto \left\{\frac{1}{\Delta T_{\text{fraudulent emails}}}\right\} \quad \text{.....(2)}$$

$$\text{Credibility}(\rho) = \left[\frac{\tau}{\hat{\tau}}\right] = \left\{\frac{\Delta T_{\text{fraudulent emails}}}{\Delta T_{\text{legitimate emails}}}\right\} \quad \text{.....(3)}$$

$$\text{Fraudulency}(\hat{\rho}) = \{1-\rho\} \quad \text{.....(4)}$$

Equations 1-4 are used to calculate the credibility and fraudulency of the senders. $\mathbf{\Delta T}_{legitimate\ emails}$ in equation (1) is the average age of all the legitimate emails with respect to the most recent email sent by a particular sender. $\mathbf{\Delta T}_{fraudulent\ emails}$ in equation (2) is the average age of all the fraudulent emails with respect to the most recent email sent by a particular sender.

We computed the credibility values of the senders' domains in the final untrusted bucket (obtained from the DNS-based and social network analysis) and used them to further filter out the false positives and false negatives.

3.3.1 CREDIBILITY OF UNTRUSTED SENDERS

The final untrusted bucket contains emails from Phishers, Opt-ins and Telemarketers, resulted from DNS based and social network analysis on corpus-I. We computed the credibility for the senders' domains of these emails and used them for resolving false positives. Figure 7 is a credibility (ρ) plot for the senders' domains in the untrusted bucket. X-axis is the number of domains of the senders. The star dotted curve represents credibility of these domains.

After grouping all the senders in the final untrusted bucket, we ended up with a total of 105 domains. As majority of these domains are untrusted, upon experimenting with various values for threshold, we used 0.70 as a threshold value. A total of 102 domains fall below the threshold except for 3 domains marked as ***O₁, O₂ and O₃.***

In the Figure 7, from the 1st domain through 102nd domain, the senders' credibility falls below the threshold. Their fraudulency when compared to credibility is more except for 103rd (0.68), 104th (0.72) and 105th (0.73) domains senders, which are marked as ***O₁***, ***O₂*** and ***O₃*** in the Figure 7. Upon verifying with recipient's feed back, it appears that all the senders from the 1st domain through 103rd domain are either; Opt-ins, Phishers or Telemarketers and emails from these domains are fraudulent in nature for instance Nigerian scam emails, walmart scam emails etc. Senders from ***O₂*** and ***O₃*** domains are false negatives; they are Opt-ins to the recipient and their wantedness according to the recipient's feed back is very low. Sender from ***O₁*** domain is a false positive; he is a trusted sender to the recipient and according to the recipient's feed back, he should be classified as trusted sender and from the plot in Figure 7 one can observe that his computed credibility value is above the threshold.

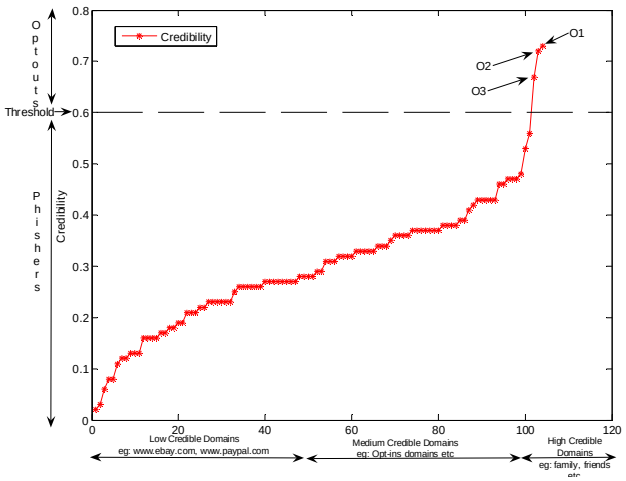

Figure 7 Credibility for the the untrusted bucket.

All the emails from ***O₁*** should be marked as credible and he should be included in the trusted category. On further verifying the reasons for his inclusion in the untrusted bucket by the DNS-based header analysis, we found that his mail servers are using private IP addresses where as they provide a valid host name (which has a public IP assigned) at the SMTP authorization. This mismatch in IP addresses during DNS-based header analysis resulted in classifying this sender as un-trusted. Over a period of time our classifier will overcome these kind of issues by learning from the computed wantedness of the senders and also from the recipients feed back. Thus finally the emails from these senders belonging to ***O₁***, ***O₂*** and ***O₃*** domains are marked as Opt-outs and removed from the Phishing category.

3.3.2 MEASURING THE RECIPIENT'S WANTEDNESS

In this section we introduce two parameters i) Tolerance (α_+), and ii) Intolerance (β_-). Using these parameters we calculate the wantedness (χ_R) of the senders' emails with respect to recipient. Normally the storage time of an email in recipient's inbox depends on its importance to the recipient. The tolerance value of a sender is directly proportional to the amount of time $\{T_{rd}\}$ his *"read"* emails are stored in the recipient's inbox. As the significance of a sender's emails increases, the recipient's tolerance towards that sender increases. The intolerance value is directly proportional to the amount of time $\{T_{urd}\}$ the sender's *"unread"* emails are stored in the recipient's inbox.

The tolerance value also depends on the frequency at which a sender is sending legitimate emails. The average time period $[\Delta T_{\text{legitimate emails}}]$ between his most recent email and his legitimate emails over the time domain indicates the how frequently the sender is mailing legitimate emails to the recipient. If this time period $[\Delta T_{\text{legitimate emails}}]$ increases as time progresses, the majority of emails from this sender are either fraudulent or less significant to the recipient. Tolerance is inversely proportional to the time period $[\Delta T_{\text{legitimate emails}}]$.

$$\text{Recipient's Tolerance } (\alpha_+) \prec \{T_{rd}\} \quad \cdots (1)$$

$$\text{Recipient's Tolerance } (\alpha_+) \prec \left\{\frac{1}{\Delta t_{\text{legitimate emails}}}\right\} \cdots (2)$$

$$\text{Recipient's Intolerance } (\beta_-) \prec \{T_{urd}\} \quad \cdots (3)$$

$$\text{Recipient's Intolerance } (\beta_-) \prec \left\{\frac{1}{\Delta t_{\text{fraudulent emails}}}\right\} \cdots (4)$$

$$\text{Wantedness } (\chi_R) = \left\{\frac{\text{Tolerance}(\alpha_+)}{\text{Intolerance}(\beta_-)}\right\} \quad \cdots (5)$$

$$\chi_R = \left\{\left[\frac{T_{rd}}{T_{urd}}\right] \times \left[\frac{\Delta t_{\text{fraudulent emails}}}{\Delta t_{\text{legitimate emails}}}\right]\right\} \quad \cdots (6)$$

$$\text{Unwantedness } (\gamma_R) = \{1 - \chi_R\} \quad \cdots (7)$$

The intolerance value of a sender depends on how often he sends fraudulent emails. The average time period [$\Delta T_{\text{fraudulent emails}}$] between his most recent email and his fraudulent emails over a time domain indicates, how frequently the sender is mailing fraudulent emails. Intolerance is inversely proportional to the time period [$\Delta T_{\text{fraudulent emails}}$]. We computed wantedness of the sender's emails as the ratio of the tolerance to the intolerance.

3.3.3 WANTEDNESS OF TRUSTED SENDERS

The final trusted category contains non spoofed emails. Based on the recipient's socially activities these emails can further classified as i) socially close and ii) socially distinct. Emails from family members, friends etc are classified as socially close. Emails from strangers, opt-ins etc are categorized as socially distinct.

We computed the wantedness values for the senders' domains of these emails and used them for classifying these trusted emails into socially close and socially distinct. Figure 8 is a wantedness (χ_R) plot for the senders' domains in the trusted bucket. X-axis is the number of domains of the senders. The curve in the Figure 8 represents the wantedness of these domains.

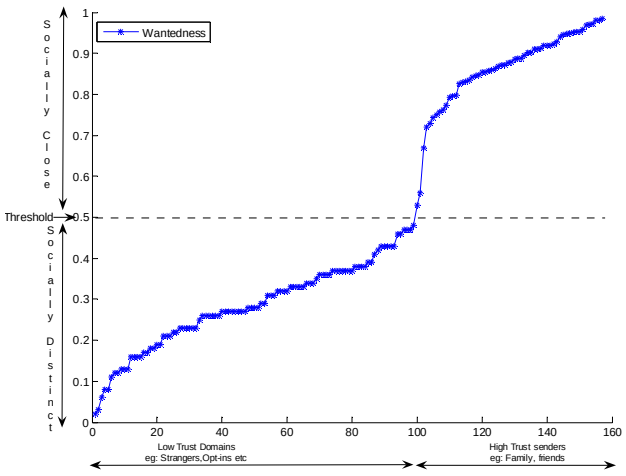

Figure 8 Wantedness of trusted senders

After grouping all the senders in the final trusted bucket, we ended up with a total of 157 domains. As majority of these domains were trusted, we used 0.5 as an initial threshold value. All the senders whose domains fall below the threshold are low trust domains. Majority of emails from these domains are from Opt-ins, Telemarketers etc. All the domains above the threshold are high trust domains. Majority of the senders from these domains are considered socially close to the recipient. The senders whose domains are below the threshold are classified as socially distinct and senders whose domains are above the threshold are classified as socially close.

4 STEP 4: CLASSIFICATION

Classification of Phishers: We calculated the fraudulency values using equation (3) during wantedness analysis and used these values for classifying phishers into i) Prospective Phishers, ii) Suspects, iii) Recent Phishers, and iv) Serial Phishers. Figure 9 shows the final classification of phishers. We calculated the phishing frequency for each sender in the final phishing bucket. Phishing frequency is the number of phishing emails which originated from the phisher or his domain.

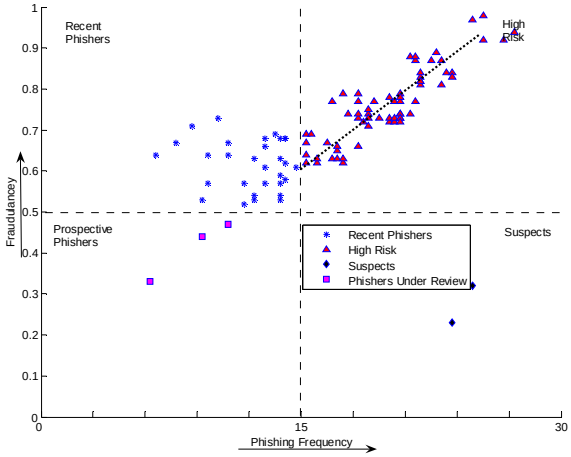

Figure 9 Phishers and Fraudulency

We plotted phishers fraudulency values and phishing frequencies to make the final classification of the phishers. Phishers who are having low fraudulency and low phishing frequency are classified as ***"Prospective Phishers",*** where as those who are having high phishing frequency and low fraudulency are classified as ***"Suspects".*** Phishers whose phishing frequency is low but their fraudulency is high are marked as ***"Recent Phishers".*** Phishers having both high phishing frequency and high fraudulency are classified as ***"Serial Phishers".*** Utmost caution should be taken while opening emails from these senders.

Classification of Trusted Senders: Figure 10 shows the final classification of domains in trusted category. We calculated the wantedness and their respective credibility for each sender in the final trusted bucket. We plotted domains credibility and wantedness for making the final classification of the senders. Senders who have a low credibility and low wantedness are classified as ***"High Risk",*** where as, those who have a high wantedness and low credibility are classified as ***"Strangers".*** Senders whose wantedness is low but credibility is high are marked as ***"Opt-Ins".*** Senders having a high wantedness and high credibility are classified as ***"Socially close".***

A summary of our results can be seen in the Table 1. *We calculated precision as the percentage of messages that were classified as phishing that actually were phishing.*

Table 1. Summary of Results

	# of emails	False Positives	False Negatives	Precision
Corpus-I				
DNS Analysis	11968	260	0	85%
{[DNS Analysis] + [Social Network Analysis]}	2548	03	05	95.6%
{[DNS Analysis] + [Social Network Analysis]+ [Wantedness Analysis]}	563 (Domains)	03	01	98.4%
Corpus-II				
DNS Analysis	756	5	0	90.4%
{[DNS Analysis] + [Social Network Analysis]}	59	0	0	93.75%
{[DNS Analysis] + [Social Network Analysis]+ [Wantedness Analysis]}	148	1	0	99.2%

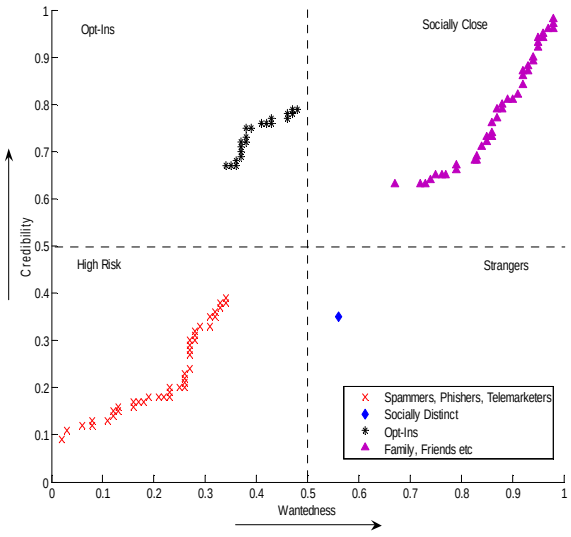

Figure 10 Classification of trusted domains

5 RELATED WORK

We applied our methodology on two live corpuses of 13,843 emails (collected over 2.5 years) and 764 emails collected over 8 months. After analyzing the corpuses, we were able to separate the phishing emails from legitimate emails. We also classify the legitimate email traffic as socially unwanted emails and socially wanted emails. Our classifier performs respectable, classifying 99% of the non legitimate traffic in both the corpuses. The classifier suggested by Microsoft, IBM, and Cornell University in 2005 anti-spam conference [17][18], uses header information for labeling the incoming emails as spam and non spam email and expects the users to make a decision on the authenticity of the source of the email.

Our classifier is accurate in classifying non legitimate traffic. For instance, there is one email in our corpus where its sender has cleverly spoofed his IP addresses with receivers, trusted domain IP addresses, in such cases SMTP Path Analysis [18], will end up in marking such emails as legitimate. Majority of the phishing emails lack, much information in the path. We found out that 99% of the phishers use a special software for sending phishing emails, which makes the emails look like, as if they reach the destination in one hop. That is, there are only two IP addresses present in the path, a spoofed senders IP address and legitimate destination IP addresses. SMTP Path Analysis [18] cannot classify these emails and require a third party classifier. We are successful in classifying these kind of emails with DNS based analysis and Social Network analysis.

6 CONCLUSION AND FURTHER WORK

From our observation of the corpus we found that if someone spoofs the IP address of the host, then mail server will tag the valid one with the spoofed one. In fact from the DNS-based header analysis and social network analysis, we can observe that the final un-trusted bucket (Figure 3) has no legitimate traffic. This encourages us that our method can result in very low false positives. This is made possible because of SMPT authorization. Our classifier can be used in conjunction with any existing spam filtering techniques for restricting spam and phishing emails. Currently we are working on developing an aggregate email classifier combining existing classifier with an innovative spam filtering technique. It classifies the incoming emails as i) Opt-outs ii) Phishing iii) Socially wanted (legitimate emails from recipient's social network) and iv) Socially unwanted (unsolicited emails from spammers and Telemarketers), rather labeling an email as spam or non spam.

Recently, we acquired a corpus from an enterprise (300K emails) and plan to research for the email correlation between their employees (i.e., social network). This corpus has got several senders and receivers. We are analyzing the sender and receiver similarities from this corpus. Our hypothesis is that spammers broadcast to large number of recipients and there is a similarity between senders (e.g., spammers' share the email addresses). For example, we would like to use the temporal and spatial correlation of the incoming emails for classification.

REFERENCES

[1] M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz. 1998, "A Bayesian Approach to Filtering Junk E-Mail", Learning for Text Categorization – Papers from the AAAI Workshop, pages 55–62, Madison Wisconsin. AAAI Technical Report WS-98-05

[2] N. Soonthornphisaj, K. Chaikulseriwat, P Tang-On, "Anti-Spam Filtering: A Centroid Based Classification Approach", Proceedings of IEEE proceedings ICSP 02.

[3] Spam Filtering Using Contextual Networking Graphs www.cs.tcd.ie/courses/csll/dkellehe0304.pdf

[4] W.W. Cohen, "Learning Rules that Classify e-mail", In Proceedings of the AAAI Spring Symposium on Machine Learning in Information Access, 1996

[5] G. Sakkis, I. Androutsopoulos, G. Paliouras, V. Karkaletsis, C.D. Spyropoulos, P. Stamatopoulos, "A memory based approach to anti-spam filtering for mailing lists", Information Retrieval 2003

[6] Luiz H. Gomes* Fernando D. O. Castro , Rodrigo B. Almeida ,Luis M. A. Bettencourt Virg´lio A. F. Almeida, Jussara M. Almeida. "Improving Spam Detection Based on Structural Similarity", Proceedings of USNIX, SRUTI (Steps for reducing unwanted traffic on the internet) workshop, July 2005.

[7] Aron Culotta, Ron Bekkerman, Andrew McCallum."Extracting social networks and contact information from email and the Web" First Conference on Email and Anti-Spam (CEAS) ***2004*** Proceedings.

[8] Yifen Huang, Dinesh Govindaraju, Tom Mitchell, Vitor Rocha de Carvalho, William
Cohen".Infe rring Ongoing Activities of Workstation Users by Clustering Email" Proceedings of First Conference on Email and Anti-Spam (CEAS), July ***2004***.

[9] Simon Corston-Oliver, Eric Ringger, Michael Gamon, Richard Campbell,"Integration of Email and Task Lists", First Conference on Email and Anti-Spam (CEAS), . July ***2004***.

[10] Vitor Carvalho, William Cohen," Learning to Extract Signature and Reply Lines from Email" Proceedings of First Conference on Email and Anti-Spam (CEAS), July ***2004***.

[11] Isidore Rigoutsos, Tien Huynh. "Chung-Kwei: a Pattern-discovery-based System for the Automatic Identification of Unsolicited E- mail Messages (SPAM)" Proceedings of First Conference on Email and Anti-Spam (CEAS), July ***2004***.

[12] Aleksander Kolcz, Abdur Chowdhury, Joshua Alspector, "The Impact of Feature Selection on Signature-Driven Spam Detection", Proceedings of First Conference on Email and Anti-Spam (CEAS), July ***2004***.

[13] Barry Leiba, Nathaniel Borenstein,"A Multifaceted Approach to Spam Reduction" Proceedings of First Conference on Email and Anti-Spam (CEAS), July ***2004***.

[15] Jennifer Golbeck, James Hendler, "Reputation Network Analysis for Email Filtering" Proceedings of First Conference on Email and Anti-Spam (CEAS) ***2004***.

[16] T.A Meyer and B Whateley"SpamBayes: Effective open-source, Bayesian based, email classification system." , Proceedings of First Conference on Email and Anti-Spam (CEAS), July ***2004***.

[17] Joshua Goodman."IP Addresses in Email Clients", First Conference on Email and Anti-Spam (CEAS) ***2004*** Proceedings.

[18] Barry Leiba, Joel Ossher, V.T. Rajan, Richard Segal, Mark Wegman, "SMTP Path Analysis"First Conference on Email and Anti-Spam (CEAS) ***2004*** Proceedings.

[19] Christine E. Drake, Jonathan J. Oliver, and Eugene J. Koontz, "Anatomy of a Phishing Email", Proceedings of First Conference on Email and Anti-Spam (CEAS), July ***2004***.

[20] P Oscar Boykin and Vwani Roychowdhury, Personal Email Networks: An Effective Anti-Spam Tool, IEEE COMPUTER, volume 38, 2004