A Sophisticated Approach to Detect Email Frauds

Sruthi Sree J S[1], Maniveena C[2]

[1, 2]CUSAT, College of Engineering Kallooppara, Kerala, India

Abstract: ***Electronic mail is a probably the most convenient way of transferring messages electronically from one person to another to any part of the world. Important features of e-mail such as its speed, reliability, efficient storage options and a huge number of added facilities make it highly popular among people from all sectors of business and society. Large amount of private and sensitive information is exchange using emails. But being popular has negative aspects too. Securing email has forever been an important issue. It can be used for both legitimate and illegitimate activities. Email frauds can be used for purposes such as spreading viruses, launch phishing attacks taking sensitive business data and other industrial espionage activities. Emails can be sent to a person from different address and it is not validated at any stage without being noticed by the receiver. Here presenting a new sophisticated approach to detect email frauds. This is the efficient method for providing integrity to emails.***

Keywords: Phishing, Botnet, Features, Data Mining, Naive Bayes

1. Introduction

Electronic mail, commonly referred to as email is a probably the most convenient method of transferring messages electronically from one person to another, emerging from and going to any part of the world. Nowadays email operates across the internet or other computer networks. Traditional email systems required that the sender and the recipient both be online at the same time.Todays email systems are based on store and forward model. Important features of E-mail such as its speed, reliability, efficient storage options and a large number of added facilities make it highly popular among peoples of business and society.

Large amount of private and sensitive information is exchange using emails. Emails are becoming largely popular has its negative aspects too. E-mails are the most targeted medium for a large number of attacks over the internet. Email Security has forever been an important issue because electronic mail is the basic internet services for communication. An email can be sent to many people within minutes. It can be used for both legitimate and illegitimate activities. Illegitimate emails include virus, Trojans may redirect to other illegitimate websites. Email frauds can be used for purposes such as spreading viruses, launch phishing attacks taking sensitive business data and other industrial espionage activities. When an email message appears to come from a legitimate source but actually it is from another illegitimate person. Emails can be sent to a person from different address and it is not validated at any stage without being noticed by the receiver.

There are thousands of emails are generating in a day such files, as physical document and conversation between the communicating parties. But nowadays email frauds are main issues in this kind of communication. Email related crimes can be categorized as email bombing, email spamming, header spoofing, email viruses and worms, phishing etc. It is thus essential to recognize and reduce users and machines misusing e-mail service. Fraud emails one which is unsolicited message; the receiver is not interested in. Some characteristics of such emails are:

- Greetings by offering prize.
- Containing financial term money, share, percent.
- Asks receiver to contact as soon as possible.
- May talk about death of some person and gives greed to receiver.

One of the major challenges of internet-fraud is the difficulty in discerning scam from spam and regular e-mail. In fact, scam messages differ from other types of spam for several reasons. First, the scam's major trait is its hidden criminal intent. In order to lure unsuspecting individuals, the text is engineered to read like regular e-mail, and thus pass successfully through spam filters. Second, messages from the same individual are not necessarily equivalent in text and story. Third, scam messages can be sent out over a longer time period than traditional bulk spam messages. Fourth, scam messages are not necessarily sent via the same physical routes as spam or via the same techniques, such as the commandeering of an open relay.

2. Challenges for Electronic Safety Spam, Fraud, and Email

The concept of spam is not a novelty to the internet world. In combination with other factor, with the increased implementation of e-mail as a direct marketing tool, the amount of spam sent over the Internet is continually growing. For this research, e-mail messages to be of three types, ham, spam, and scam. Figure depicts the relationships between e-mail types. Spam messages are unsolicited piece of piece of mail. The scam messages are a subset of spam messages which are intelligent in design, such that they attempt to coax the person to execute some action of illegal purpose beyond a simple click. Ham refers to genuine e-mail..It exist certain messages which are viewed as spam by some persons and ham by others.

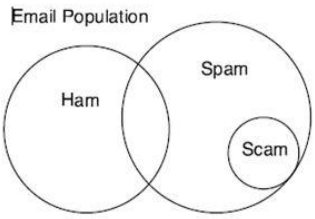


Figure 1.1: Email types and their relationships

2.1 Detecting Fraudulent Intent

Detecting the fraudulent intent can be stated as one of binary classification. Basically, design a message filter that discriminates between messages which contain patterns of fraudulent intent for the type of scam and other e-mail. That is trained to make a Boolean decision on a labeled dataset, where the labels are normal emails and fraud emails.

2.2 Popular email frauds by attackers

Emails navigate the globe daily. Like any other form of communication, email is also misused by cyber criminals. The ease the speed and relative anonymity of email has made it a powerful tool for criminals.

Some of the major email related crimes are:
1) Email spoofing
2) Sending malicious codes through email
3) Email bombing
4) Phishing
5) Spam Attack, email spamming
6) Denial of Service Attack
7) Sending threatening emails

Email spoofing is sending email messages with a forged sender address. It is easy to impersonate and forge emails because the core protocols do not have any method for authentication. It can be accomplished within a LAN or from an external environment using Trojan horses. Spam and phishing emails typically use such spoofing to mislead the recipient about the origin of the messaging. It is an e-mail activity in which the sender address and other parts of the e-mail header are altered to appear as though the e-mail originated from a different source.

Sending malicious codes through email is the fastest and easiest ways to propagate malicious code over the internet. Hackers bind Trojans, viruses, worms and other computer contaminants with e-greeting cards and then email them to unsuspecting persons.

Email bombing is one of the net abuses by sending huge volumes of email to an email address to overflow the mailbox or overwhelm the server where the email address is hosted in a denial-of-service attack.

Phishing attack is generally not harmful to the inner workings of your pc. It is designed to trick you into revealing your personal information, bank account information, passwords. For example, the phisher sends you a message that looks like it came from your bank. The message requests you to verify your account information to continue using your account. That tells you that if you do not verify the information your account will be closed. Someone that is unaware of phishing easily gets tricked into revealing their account information.

Spam attack and email spamming: Spam is unsolicited email or junk mail that you receive in your inbox and generally contain advertisements but it can also contain malicious files. When you click on spam emails the files are downloaded into your email client and into your pc. The similar thing can happen if you respond to spam in an attempt to get removed from the list.

Email spam has gradually increased since the early 1990s. Spammers collect email addresses from chat rooms, customer lists, websites, newsgroups, and viruses which harvest users' address books, and sold to other spammers. They also use a practice known as email appending in which they use known information about their target to search for the target's email address.

Denial of service attack occurs when the hacker sends bulks of email messages to your email client in an effort to block your email client or crashing your computer altogether. In an organization a denial of service attack on email can crash an entire network and prevent the users from responding to legitimate transfer.

These are types of email client exploits that can result in damage to your applications, data and even your computer's operating system. Additionally, an email client can cause you to expose private information and also expose your system to a finite list of malware that ranges from simply spying on your web browsing activity to literally destroying your pc components. If you aware of how these exploits occur, you can take preventive measures to protect emails.

Threatening emails are those made with intent to obtain a monetary advantage or to force a person to act against his or her will**.** It becomes quite easy for anybody with even a basic knowledge of computers to become a blackmailer by threatening someone via e-mail.

3. Problem Formulation

Peoples cannot imagine a day without internet. E-mail is one of the primary ways through which peoples communicate, not only for official communication but also to be in touch with our friends and relatives. E-mail plays a vital role in communication globally and sharing of data as well. The security issues also increased. The major problem nowadays is the attack on e-mail by the hackers. It is the right time to secure the communication over mail even on trusted network. Cyber criminals craft these emails to look convincing and sending them out to literally millions of people around the world. The cyber criminals do not have a specific target in mind nor do they know exactly who will fall victim.

These days attackers simply know the more emails they send out, the more people they may be able to fool. Whenever a network address is the source of large amount of spam that address is added to a blacklist and many internet service providers block all email from that address. Spammers are highly motivated to send mass email that is complicated to trace back to a particular network address. Again, a primary motivation for conceal their identities is that the email they send will not be automatically blocked. So we can't completely trace and filter the fraud emails using existing spam filters today. Presenting a new sophisticated approach to detect email frauds. This is the efficient method for providing integrity to emails.

4. Background and Literature Review

First step for concealing their identities, spammers embed false and misleading headers in spam email messages that they send.. For example, the from and reply to headers in spam email messages usually point to some random non-existent or innocent party. The first received header cannot be spoofed; it is not under the control of the sender. Header identifies the network address of the computer that initiated the connection to the recipient's mail server. To ensure the first received header does not reveal their identities spammers must somehow conceal the network addresses of their computers .Four identity concealment techniques currently used by attackers. Primary identity concealment techniques to send untraceable emails. The intention of these techniques is to hide the network addresses of the spammers' computers.

Bot network consists of tens of thousands of compromised machines running malicious software. This software propagates in the form of viruses and worms that infects the machines of innocent internet users. Once computer is infected by one of these programs a spammer can remotely take control the computer and send spam from it. For example, industry sources estimate the largest reported bot network to date typically desktop computers (many with high speed internet connections), and their users are usually unaware that their machines have been compromised.

The bot software hides itself on the drone machine and periodically checks for commands from the human bot network administrator .Bot networks are increasingly used for sending spam emails. Sending spams using a bot network, a spammer instructs all drones under his or her control to send email to addresses on the spammer's list. To the recipients, the emails sent by in a bot network appear to come from legitimate home or corporate users. The identity of the spammer is nowhere in these emails spammers manage to conceal their network addresses from spam recipients by employing bot networks

Open proxies are another method that spammers use to conceal their identities. A proxy server is a machine that helps two machines communicates with each other. Some organizations have lots of computers on their networks but have smaller number of proxy servers that are the only machines on the network that directly interact to the internet. This design provides more efficient web browsing for the users within that organization and secures the organization's network against unauthorized.

Open mail relays means email messages are hardly ever sent directly from the sender email server to the recipient's email server. Instead, email messages pass through a number of gateways called mail relays. Each time an email message passes through a mail relay, the relay inserts a received header at the front of the message that shows the address of the computer that connected to the mail relay. By the time the email message reaches its recipient it contains a number of received headers one for every relay server through which the email message has passed.

When not manipulated, the list of received headers identifies the entire chain of relays that processed the email message spammers can use open mail relays to conceal their identities. Instead of sending email to the recipient directly, the spammer sends email to the open relay and the open relay forwards the email to the recipient. The email messages travels from the spammer to the mail relay and then to the recipient. From the recipient's point of view, the email appears to come from the open relay, not the spammer. . A spam email message travels from the spammer to the email proxy, then to the mail relay, and finally to the specified recipients. The final emails contain the open proxy's network address, but reveal nothing about the spammer's identity.

Untraceable Internet connections are several ways to access the internet through a network address that cannot be linked to an individual or a physical location. Users who connect to the Internet through public Internet cafes, through free (or stolen)wireless connections, or through certain universities' on-campus networks need not identify themselves and can therefore send messages anonymously on the Internet. Spammers may also purchase ISP roaming access using false names and untraceable payment methods. There is no way to associate such network addresses with the spammers who use them. When using an untraceable Internet connection the spammer need not hide her network address and can send email directly to spam recipients (by directly connecting to the email port on the recipient's email server).

E-mails are the preferred medium for a large number of attacks over the internet. Some of the most popular attacks over the internet include increased scams, malicious attacks, spear phishing attack and targeted attack. Amar V Sabel and Prof. Vijay S Gulhane [1] presented some methods are actually in detection of spam related mails but they have higher false positives. Variety of filters such as Checksum-based filters, machine learning based filters and memory-based filters are usually used in order to identify spams. Spammers always try to find a way to avoid existing filters, new filters need to be developed to catch spam.. Ontology's allow for machine-understandable semantics of data.

Phishing pose a serious threat to end to end users and commercial institutions. Majority of the day phishing attacks employ email as their primary carrier. Recent defense mechanism focus on detection by validating the authenticity of the website. Very few approaches have been proposed which concentrate on detecting email based phishing attacks based on the structural properties inherently present in the phishing email.Madhusudhanan Chandreshekaran[2] present an approach consists of a novel technique to discriminate phishing emails from the legitimate emails using different structural features present in them. Specific difficulties of identifying a spammer by following the electronic trail embedded in spam email Dan Boneh [3]discusses describes how the email system works and how it is used by spammers and how they conceal their identities etc.

Cybercrime has grown voluminous in this hi-tech world. The flout towards cybercrime has become today's prime centric with developing countries frugality as well. Phishing, spams and email frauds are more equally exasperating. In this intellect learning Dr.P.S Jagadeesh Kumar [4] main attention is to make a healthy charge on phishing, spam and email fraud towards the wealthy personal information and realm.

Phishing emails are intended to fool the recipient into handing over personal information such as login names, passwords, credit card numbers, account credentials, social security numbers etc. Spam emails harm their victims through loss of funds and identity theft. They also affect internet business because people lose their trust in internet transactions for fear that they will become victims of fraud. Noor Ghazi[5] deals with the phishing detection problem and how to detect phishing emails. Phishing emails are intended to fool the recipient into handing over personal information such as login names, passwords, credit card numbers, account credentials, social security numbers etc.

Spam emails harm their victims through loss of funds and identity theft. They also affect internet business because people lose their trust in internet transactions for fear that they will become victims of fraud. Noor Ghazi[6] deals with the phishing detection problem and how to detect phishing emails.

5. Proposed Method

Here present a sophisticated approach to detect email frauds using advanced feature choice. Proposed solution is the fraudulent email detection model by employing various features, evaluating using well known classification algorithm. The extracted features and compared the performance of each category of features with the others in terms of the fraudulent email detection rate. The experiments performed on diverse feature sets and with the classification methods.

5.1 Feature set construction

Build feature sets for different kind of fraudulent emails. For example some emails are intended to deceive the receiver by tempting them and showing helplessness for getting their sympathies in order to get receivers personal information such as address and bank details and so on. Other emails may be more deceptive that look like these are sent by the receiver's side and require urgent action by the user and redirect them to some malicious website. Important features are extracted which have more capability to separate

Fraud emails differ from normal email and the special features have been added that contain specific words in the subject of the email and hyperlinks in the body in order to redirect the receiver to a certain website. Features are used to represent the author as his/her whiteprints. As fingerprints historically had been used by law enforcement experts to uniquely identify criminals whiteprints could be used to identify authors to their writing styles. Researchers can apply their results to cybercrimes especially to fraud emails by revealing the identity of the author. Like fingerprints each writer may have unique writing style. These unique writing style features are termed as stylometric other than there have many features to identify authors.

6. Proposed Algorithm

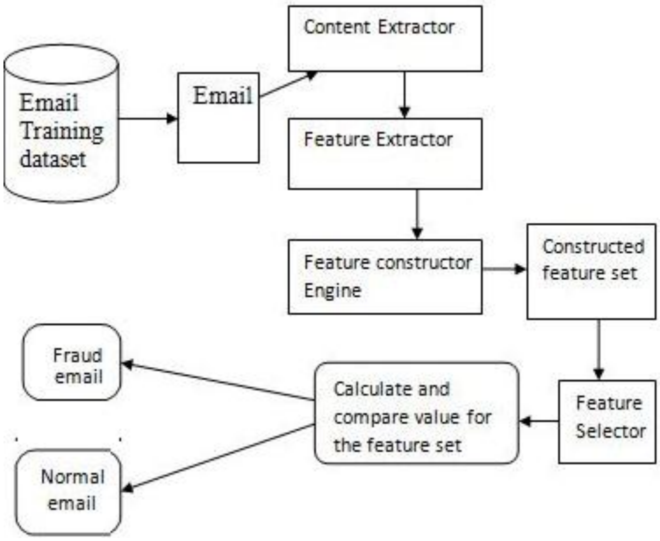
Here we have used classification process. For classification of data one efficient algorithm is naive bayes. Naive Bayes has a lot of advantageous properties than other algorithms due to their simplicity, linear computational complexity, and its accuracy. The majority of other approaches require iterated evaluation. Classification requires a single table lookup per token plus a final product or sum over each token. Storage space requirements are small in Naive Bayes because we need to store only the token counts rather than entire messages. The classifier is updated incrementally when new messages arrive. Each technique makes the independence assumption that the probability of tokens occurring in a message is independent.

6.1 Detection of fraudulent emails using Naive Bayes

Naive Bayes is very popular in commercial and open-source anti spam email filters. As a simplification, we focus on the textual content of the email messages. Operational filters would also consider information such as noun and contextual phrase, which can be added as additional attributes in the message representation discussed below. Alternatively, separate classifiers can be skilled for textual and other attributes and then form an ensemble.

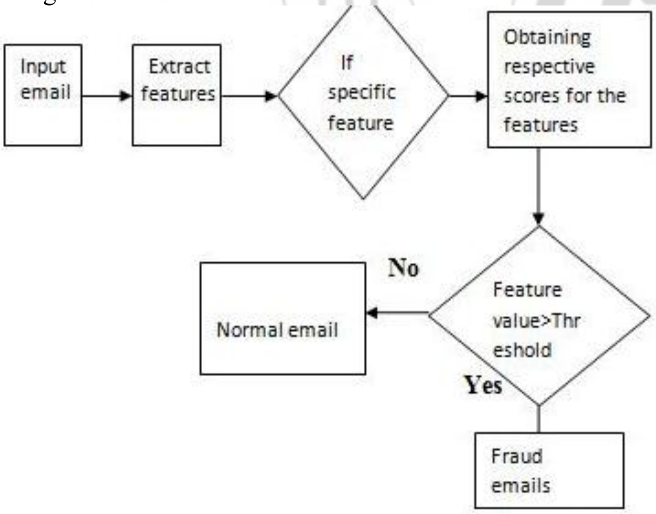
7. System Design

The Design of the fraudulent email detection is comprised of seven modules, which works as an assembly of tasks. The functions of each of the module are follows:



Feature construction engine builds up diverse feature sets which are designed according to the experience and are found in various kinds of the fraudulent emails. Feature sets are separated into different categories, depending on kinds of fraud being considered in the email. Although in the current work we classify emails into fraud or normal. Feature Selector select different feature sets are available not all features of worth considering for fraudulent email detection task.

This module used to select the features and compared with already defined datasets. Output Module produces the results based on the features and classification algorithm used. Here calculates and compare value for the data set and based on that values the output is produced using the accuracy of fraud email detection. We can classify the emails into two categories fraud emails and normal emails.



8. Conclusions

Email frauds have become problem in recent years. Fraud emails are real danger to internet communication. Detection of fraud emails are necessary there are techniques to detect spam emails and there have lots of limitations like accuracy is low, contents can be same as legitimate emails, detection rate is low etc. So some advanced method is required to overcome the limitations.. From this proposed model we can reduce the fraud activities using emails.

In various digital forensic investigations email data needs to be analyzed. Cybercriminals forge email headers or send it anonymously for illegitimate purposes which lead to several crimes and thus make e-mail forensic investigation crucial. However, this poses a threat to the privacy of the individual whose emails are being examined and in particular becomes a problem if the investigation clashes with fraud emails and suspicious activities using emails. That problem also can be overcome using this proposed method.

9. Acknowledgements

Many people have contributed to the success of this. Although a single sentence hardly suffices, I would like to thank Almighty God for blessing us with His grace. Last but not the least, I thank all others, and especially my Husband and classmates for their ultimate support and help in the successful completion of this work.

References

[1] Amar v Sablel and Prof. Vijay S. Gulhane(2015)"Email filter for spam email:Areviw"1-6.
[2] Madhusudhanan Chandrasekaran M, Narayanan K, Upadhyaya(2013),"Phishing email detection based on structural properties",2-5.
[3] Dan Boneh(2006)," The Difficulties of Tracing Spam Email",1-9.
[4] Dr.P.S.JagadeeshKumar,,Dr.S.MeenakshiSundaram (2015), Mr.Ranjeetkumar, "An intellectual learning on e-mail security and fraud,spam and phising".,49-64.
[5] Noor Ghazi M. Jamee , Loay E. George (2014), "Detection Phishing Emails Using Features Decisive Values",257-259.
[6] Gori Mohamed .J, M. Mohammed Mohideen, Mrs.ShahiraBanu,(2014),"Email phising an open threat to everyone",1-4.

Author Profile

Sruthi Sree J S obtained the Degree of Bachelor of Engineering in Computer Science and Engineering from Anna University Chennai in 2013. She is now pursuing her master degree in Computer Science with specialization in Cyber Forensics and Information Security at College of Engineering, Kallooppara under Cochin University of Science and Technology.

Maniveena C obtained B.E in Computer Science and Engineering from Anna University Chennai and M.Tech in Computer Science and Engineering from Anna University Chennai. She is currently working as Assistant Professor in College of Engineering, Kallooppara.