International Journal of Engineering Research & Management Technology ISSN: 2348-4039
May-2016 Volume 3, Issue-3
Email: editor@ijermt.org www.ijermt.org

A STUDY ON E-MAIL SPAM & THEIR FILTERS WITH PREVENTION METHODS

Deeksha Tomar
M.Tech Scholar
Radha Govind Group of Institution
Meerut

Mr. Rameshwar Tiwar
Assistant Professor
Radha Govind Group of Institution
Meerut

ABSTRACT

Electronic mail is used daily by millions of people to communicate around the whole world and is a mission to maintain a healthy relationship via E-Mail. From last years, unsolicited bulk email has become a major problem for email users to study as overwhelming amount of spam is flowing into users mailboxes daily. Spam frustrating for most email users. The necessity of effective spam filters increases. In this paper, we presented our how to get the solution to get rid of the fake E-Mails by phishing attack.

KEYWORDS: E-Mail, Spam, Spam Filter, Anti-Spam Filter, Spam Identification

INTRODUCTION

Phishing is an attempt by an individual or a group to steal personal confidential information from unsuspecting victims, financial gain for, identity theft and other fraudulent activities. In phishing, the attacker creates a replica of existing web pages to fool the users. Due to which the attacker gets personal, financial and password related information from users. Phishing emails are designed to look like legitimate messages from actual banks, businesses, and other organizations. In reality, though, criminals created the message, usually in an effort to steal your money, identity, or both. They want you to click links that will take you to a website that looks authentic but is really just there to capture your credit card or other personal information or perhaps to distribute malware

SPAM – UNSOLICITED BULK EMAIL

E-mail spam, known as unsolicited bulk Email (UBE), junk mail, or unsolicited commercial email (UCE), is the practice of sending unwanted e-mail messages, frequently with commercial content, in large quantities to an indiscriminate set of recipients.

The risks in filtering spam are sometimes legitimate mails may be rejected or denied and legitimate mails may be marked as spam. The risks of not filtering spam are the constant flood of spam clogs networks and adversely impacts user inboxes, but also drain valuable resources such as bandwidth and storage capacity, productivity loss and interfere with the expedient delivery of legitimate emails. General Advice to avoid spam is, Avoid giving your “real” email address to all but close associates, Setup web mail accounts (Google, hotmail etc.) for registering with web sites or for communicating with people you do not know, Educate your contacts to exercise caution with email address, Do not open junk email, just delete it, Never click to unsubscribe to a mailing unless you are sure it is a reputable entity.

LIST OF PHISHING TYPES PHISHING

A. SPEAR PHISHING

Phishing attempts directed at specific individuals or companies have been termed **spear phishing.** Attackers may gather personal information about their target to increase their probability of success. This technique is, by far, the most successful on the internet today, accounting for 91% of attacks

B. CLONE PHISHING

A type of phishing attack whereby a legitimate, and previously delivered, email containing an attachment or link has had its content and recipient address (es) taken and used to create an almost identical or cloned email. The attachment or link within the email is replaced with a malicious version and then sent from an email address spoofed to appear to come from the original sender. It may claim to be a resend of the original or an updated version to the original. This technique could be used to pivot (indirectly) from a previously infected machine and gain a foothold on another machine, by exploiting the social trust associated with the inferred connection due to both parties receiving the original email.

C. WHALING

Several recent phishing attacks have been directed specifically at senior executives and other high profile targets within businesses, and the term whaling has been coined for these kinds of attacks. In the case of whaling, the masquerading web page/email will take a more serious executive-level form. The content will be crafted to target an upper manager and the person's role in the company. The content of a whaling attack email is often written as a legal subpoena, customer complaint, or executive issue. Whaling scam emails are designed to masquerade as a critical business email, sent from a legitimate business authority. The content is meant to be tailored for upper management, and usually involves some kind of falsified company-wide concern. Whaling phishermen have also forged official-looking FBI subpoena emails, and claimed that the manager needs to click a link and install special software to view the subpoena

D. LINK MANIPULATION

Most methods of phishing use some form of technical deception designed to make a link in an email appear to belong to the spoofed organization. Misspelled URLs or the use of sub domains are the common tricks used by phishes. In the following example URL, http://www.yourbank.example.com/, it appears as though the URL will take you to the *example* section of your *bank* website; actually this URL points to the "*your bank*" (i.e. phishing) section of the *example* website. Another common trick is to make the displayed text for a link (the text between the <A> tags) suggest a reliable destination, when the link actually goes to the phishes' site. Many email clients or web browsers will show previews of where a link will take the user in the bottom left of the screen, while hovering the mouse cursor over a link.[46] This behavior, however, may in some circumstances be overridden by the phishes.

E. FILTER EVASION

Phishes have even started using images instead of text to make it harder for anti-phishing filters to detect text commonly used in phishing emails.] However, this has led to the evolution of more sophisticated anti-phishing filters that are able to recover hidden text in images. These filters use OCR (optical character recognition) to optically scan the image and filter it. Some anti-phishing filters have even used IWR (intelligent word recognition), which is not meant to completely replace OCR, but these filters can even detect cursive, hand-written, rotated (including upside-down text), or distorted (such as made wavy, stretched vertically or laterally, or in different directions) text, as well as text on colored backgrounds.

SPAM IDENTIFICATION METHODS

The several different methods to identify incoming messages as spam are, White list/Blacklist, Bayesian analysis, Mail header analysis, Keyword checking.

A. WHITELISTS/BLACKLISTS

The functionality of these filters is simple: a white list is a list, which includes all addresses from which we always wish to receive mail. we can add email addresses or entire domains, or functional domains. An interesting option is an automatic white list management tool that eliminates the need for administrators to manually input approved addresses on the white list and ensures that mail from particular senders or domains are never flagged as spam. The number of records can be configured. When an overflow occurs,

obsolete records are overwritten. A blacklist works similarly to competitive alternatives: this is a list of addresses from which we never want
to receive mail.

B. MAIL HEADER CHECKING

This is a fairly known method. Mail header checking consists of a set of rules that, if a mail header matches, triggers the mail server to return messages that have blank "From" field, that lists a lot of addresses in the "To" from the same source, that have too many digits in email addresses (a fairly popular method of generating false addresses). It also enables to return messages by matching the language code declared in the header.

C. BAYESIAN ANALYSIS

The word probabilities (also known as likelihood functions) are used to compute the probability that an email with a particular set of words in it belongs to either category. This contribution is called the posterior probability and is computed using Bayes' theorem. Then, the email's spam probability is computed over all words in the email, and if the total exceeds a certain threshold (say 95%), the filter will mark the email as aspam.

D. KEYWORD CHECKING

Another method widely used in filtering spam. It works by scanning both email subject and body. Using "conditions" i.e. combinations of keywords is a good solution to enhance filtering efficiency. We can specify combinations of words and update the list that must appear in the spam email. All messages that include these words will be blocked.

SPAM FILTERING TECHNIQUES

The various spam filtering techniques adopted to get rid of the problem of spam are discussed.

A. DISTRIBUTED ADAPTIVE BLACKLISTS

This technique can be used at the mail server. When a message is received by a MTA, a distributed blacklist filter is called to determine whether the message is a known spam. These tools use clever statistical techniques for creating digests.

Tools: Razor and Pyzor operate around servers that store digests of known spams.

B. RULE BASED FILTERING

Evaluate a large number of patterns--mostly regular expressions--against a candidate message. Some matched patterns add to a message's score, while others subtract from it. If a message's score exceeds a certain threshold, it is filtered as spam; otherwise it is considered as legitimate. Some ranking rules are fairly constant over time. Other rules need to be updated as the products and scam advanced by spammers evolves.

Tool: Spam Assassin is one of the popular rule based spam filtering tool.

C. BAYESIAN CLASSIFIER

Particular words have particular probabilities of occurring in spam email and in legitimate email. The filter doesn't know these probabilities in advance, and must first be trained so it can build them up. After training, the word probabilities (also known as likelihood functions) are used to compute the probability that an email with a particular set of words in it belongs to either category. Each word in the email contributes to the email's spam probability, or only the most interesting words. This contribution is called the posterior probability and is computed using Bayes' theorem. Then, the email's spam probability is computed over all words in the email, and if the total exceeds a certain threshold (say 95%), the filter will mark the email as a spam. Some spam filters combine the results of both Bayesian spam filtering and

other heuristics (predefined rules about the contents, looking at the message's envelope, etc.), resulting in even higher filtering accuracy, sometimes at the cost of addictiveness. Server-side email filters.

Tools: DSPAM, Spam Assassin, Spam Bayes, Bogofilter and ASSP, make use of Bayesian spam filtering techniques.

D. K NEAREST NEIGHBORS

If at least *t* messages in *k* neighbors of the message *m* are unsolicited, then *m* is unsolicited email, otherwise, it is legitimate.

Tool: TiMBL uses k nearest neighbor technique.

HOW TO PREVENT PHISHING SCAMS

A lot of phishing emails claim to come from legitimate sources or popular websites. The emails often ask the user to enter bank details or other personal information. There are also phishing scam websites which appear exactly like the original websites. Some of these fake websites are so well done that it's quite impossible to tell them apart unless you look at the URL. Most of these fake websites redirect users to pages with spaces where they have to fill in essential financial information usually used to access bank accounts. Once the phishers get a hold of the information, they can carry out fraudulent monetary transactions. Sometimes, the website may ask the user to fill in personal details like social security number, driver's license number, and other details which can be used to commit frauds in the user's name.

A. Check the email Carefully
B. Never Enter Financial or Personal Information
C. Identify a Fake Phone Call
D. Protection through Software
E. Never Send Personal Information through emails
F. Check Bank Details Regularly
G. Never Download Files from Unreliable Sources

ANTI-PHISHING SOFTWARE

There are many simple-to-use types of software out there that are specifically designed to address phishing threats. Using this type of software is not overkill; after all, if your information falls into the wrong hands, you could be in for a world of hurt. The best part is that there are plenty of free options out there, so you don't have to spend a bunch of money.

HOW ANTI-PHISHING SOFTWARE WORKS

Whether it is integrated into your web browser or operates in a standalone way, anti-phishing software works in a simple but very useful way. Information about known phishing scams and phishing sites are stored within these programs. That information is used to alert you if you stumble upon a potentially dangerous site. In the case of a browser-based program, an alert may pop up at the top of your browser's window; in the case of standalone software, an alert may pop up at the bottom of your screen. Some types of software automatically redirect you, while others give you the option of staying or leaving.

ANTI-PHISHING SOFTWARE

Phish Tank Site Checker and GFI Mail Essentials are two popular examples of anti-phishing software. There are also many popup-blocking programs that work well; many phishing attacks are conducted via popup windows. If you are in the market for an antivirus program, you could always choose one that includes anti-phishing features. Several of them are available, and it is nice to have everything integrated into one easy package.

IN-BROWSER OPTIONS

If you use Firefox or Chrome, you already enjoy a pretty decent amount of anti-phishing protection. Those two browsers are automatically equipped with Google Safe Browsing, which used to be available as a separate download. If simplicity is your top concern, you should consider using one of those browsers to avoid the majority of phishing attacks. There are also many different toolbars that can be installed on today's most popular browsers. Make sure to do a little research before you select one program. You can learn from the mishaps of others by reading reviews and testimonials. It pays to avoid anti-phishing software that does more harm than good. There is no such thing as being too protected against phishing scams. At the same time, you should never rely solely on software to keep yourself safe. Learn how to identify common phishing scams and how to recognize phishing emails and phishing sites. That knowledge will go a very long way towards keeping you safe.

CONCLUSION

Spam or unsolicited e-mail has become a major problem for companies and private users. This paper explored the various problems associated with spam and different methods and techniques attempting to deal with it. From the study we identified that, many of the filtering techniques are based on text categorization methods and there is no technique can claim to provide an ideal solution with 0% false positive and 0% false negative. There is lot of scope for research in classifying text messages as well as multimedia messages.

REFERENCES

1. Christina V "A Study on Email Spam Filtering Techniques" *International Journal of Computer Applications (0975 – 8887)*
2. http://www.zonealarm.com/blog/2014/07/7-ways-to-spot-phishing-scam/
3. http://phishme.com/top-10-phishing-attacks-2014/
4. http://resources.infosecinstitute.com/phishing-techniques-similarities-differences-and-trends-part-i-mass-phishing/