NATURE OF SPAMS IDENTIFICATION USING LAYOUT ABSTRACTION SCHEME

Roshan Jammer Shaik [1*], P. Suresh Babu [2*], Vennakula L S Saikumar [3*]

1. *M.Tech (CSE) Student, Dept of CSE, Koushik College of Engineering, Vishakapatnam.*
2. *Associate Professor, Dept of CSE, Koushik College of Engg, Vishakapatnam.*
3. *M.Tech (CSE) Student, Dept of CSE, Pydah College of Engg & Tech, Vishakapatnam.*

Keywords:
Robustness, Internet Security Systems, Data Center leverages, ISP (Internet service provider)

Abstract: *These days the communication medium is likely have expanded in a great manner. One of such is Email service which these days not only the end user but also the Industry people are likely to a mandatory service. Hence in order to provide effective email service vide of spam we have taken consideration of filtering the spammed data. Fighting spam is important. More than 50% of all email messages are unsolicited and usually unwelcome. Spam costs enterprises and individuals billions of dollars each year by consuming time and resources while delivering content of limited, if any, value. Effective Web and e-mail filtering relies on three components: a robust content analysis process, a vast amount of data and the means to analyze that data. The content analysis technology used in the Internet Security Systems (ISS) Global Data Center leverages all three of these components to provide the foundation for all content security products and solutions, for Mail Security solution. Again robustness is the main factor behind the mail of service of huge data and the problem is detection mechanism algorithm. Hence of, in this paper we try to give the stress on controlling the huge data mail with respect to preference of filtering mechanism based on user.*

1. INTRODUCTION

As it multiplies and spreads out to the recipients, this one email consumes resources from many Internet carriers, service providers, businesses and customers. It expends bandwidth, router backplanes, computer processing time and disk space, plus the time of Internet workers and email patrons. When multiplied by millions of messages from thousands of spammers each day the cost of unsolicited bulk email become a burden for all but the senders. E-mail communication is prevalent and indispensable nowadays. However, the threat of unsolicited junk emails, also known as spams, becomes more and more serious. Through an automatic e-mail, users can be notified periodically (daily, weekly, etc.) about the system in a way that causes them to act. For example, they may be notified that there are messages pending in their Spam Traps. Or, users may be notified that messages in their Spam Traps will be deleted within a certain time frame if they do not take action on them. Email spam targets individual users with direct mail messages. Email spam lists are often created by scanning Usenet postings, stealing Internet mailing lists, or searching the Web for addresses. Email spams typically cost users money out-of-pocket to receive. Many people - anyone with measured phone service - read or receive their mail while the meter is running, so to speak. Spam costs them additional money. On top of that, it costs money for if you need to fill out web registration forms or surveys at sites from which you don't want to receive further information, consider using secondary addresses to protect primary email accounts from spam abuse. Also, always look for a check-box that solicits future information/offers, and be sure to select or deselect as appropriate. Conscientious end users who follow these suggestions will ultimately play a significant

**** Roshan Jammer Shaik***
M.Tech (CSE) Student, Dept of CSE, Koushik College of Engineering, Vishakapatnam.

role in reducing the amount of spam that enters their organization's communications system, especially when automated spam filtering supplements their efforts. SPs and online services to transmit spam, and these costs are transmitted directly to subscribers. One particularly nasty variant of email spam is sending spam to mailing lists (public or private email discussion forums.) Because many mailing lists limit activity to their subscribers, spammers will use automated tools to subscribe to as many mailing lists as possible, so that they can grab the lists of addresses, or use the mailing list as a direct target for their attacks.

2. RELATED WORK

Spamming remains popular with direct email advertisers because there are no incremental charges. Reaching one million people costs the same as targeting one individual. Printed advertising, by comparison, has incremental price boosts. Printing and mailing costs increase when the intended audience grows from 1,000 to one million. Like the print advertisers, Internet carriers and service providers pay incrementally for what they use. The more bandwidth, hardware and personnel they require the more they pay. Some ISPs say as much as half the email they handle is resource-wasting spam. The price for handling this spam eventually gets passed along to the everyday customers who receive the unsolicited offerings. As a corporate tool, email provides many benefits, including ease of use and relative security, and in a business context, it is available to everyone for a reasonable price. Email has moved beyond its early forms to become an almost completely interoperable communication vehicle not only inside an organization but also across organizations and individuals. Interoperability is one of the most critical features of an effective communication tool, and email's interoperability has enabled email to be widely adopted both outside and inside the organization. As a mature tool, email Systems now include (at a minimum) interoperable email accessible from a desktop client or browser as well as integrated calendaring for group scheduling. Basic features such as delegation, search, filtering, flags, offline sync, out-of-office automated replies, attachments, and integrated contacts are standard. The systems can be traditional on-premise installations, appliances sold with a perpetual license, or some variation ranging from hosted by an outsourced provider to a full cloud-based subscription system.

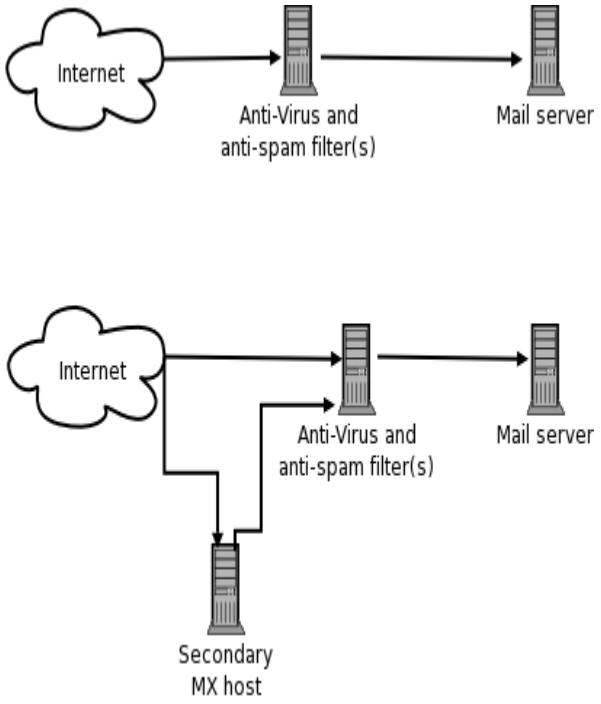

Fig. 2.1 Showing filtration of spammed mail from various networks

In corporate circles, email has become an indispensable tool for many, but with the explosive growth in the volume of emails and use cases, there is a growing sense of frustration among beleaguered users and concern from executives. All of the needs for a diverse set of business use cases. For many companies, email has become the default tool for much more than communication, and its use now extends to file sharing, peer-to-peer and team collaboration, project coordination, corporate broadcast communications, marketing, collecting feedback, and many more activities beyond its intended use. In fact, the culture in many companies has evolved to one of "copy everybody" and "reply all" just in case. Add to this the current pervasive nature of mobile email access, which in many organizations has created an unrealistic expectation of email as a real-time/synchronous tool, and it's easy to start to see the source of growing employee email fatigue.

3. METHODS

Spam exists because senders have a marketplace – a relatively small one, but profitable nonetheless. While direct mail marketing typically receives one to five percent response, email spammers sell to one in a thousand or one-tenth of one percent. Spammers can afford to sell in

such weak markets by using the bandwidth and hardware of other people. They pass on the incremental expenses to Internet carriers, service providers and users. If spammers paid more, they would sell less and possibly disappear. Various countries and states within countries have drafted legislation to make spamming more difficult, but few laws have passed. Besides, enforcement is difficult at best because of the international nature of the Internet. Being clever people, spammers can also be difficult to physically locate because all they need to operate is a notebook computer equipped with a modem. Moving around to avoid detection is no problem. So spamming and its related ills continue to grow. As far as we discussed in the related work tackling with email may follow various methods, out of them first of we likely to concentrate on first detecting followed by type of spam then delivering the solution. In order to cope with exception sometimes we like to lose many important mail as spam mail, hence of we use robustness mechanism as a high end solution. Coping with the Ranking algorithm , we use to efficiently provide the grade to corresponding domain based on such criteria we can filter the spammed to the end user, of course we also can use for grouped mail having high density of data. Next we will have to consider the Internet Security Systems of cipher data giving preference to hidden and cryptic data, for which we likely to use key based algorithm.

In the Web sense Advanced Classification Engine (ACE) detects blended threats, Web scripts, and dynamic attacks by combining a host of detection methods. Powered by the Threat seeker Network, it collects data from more than 900 million endpoints and analyzes up to 5 billion Web pages per day by leveraging technologies such as:

- Real-time security analysis
- Real-time content classification
- Real-time Data Classification (RTDC)
- Antispam
- Reputation services
- URL filtering
- Multiple malware solutions

Hence of data leverage in the sense of analyzing and based on report if same type of contextual data found, it detects and filters the mail.

An ISP presents one of the most complex environments for managing spam because of the high volume of e-mail, wide variety of users and high level of service demanded by customers. This white paper reviews the most common approaches to spam management in an ISP environment and details how a suitable solution must address particular challenges for ISP administrators.

ISPs generally consider three types of anti-spam solutions:

- Outsourced filtering services relay your mail through a third party system housed off site. These systems can sometimes be too costly for the typical ISP. They can also present concerns about a loss of control or security over ISP e-mail.
- Home grown solutions are typically based on open-source software such as SpamAssassinTM and Roaring. In most cases, these solutions were sufficient until about 2002, when the volume of spam and spammer's ever-evolving techniques began to render homegrown solutions unmanageable.
- Third-party in-house solutions offer a balance between the above two options. While benefiting from the experience of a third party provider, you keep the flow of your email traffic on your own network

Algorithm or procedure:

Input:Bytes of mails presented in array for the test

//the test bad have to in server.

Output:

1. Detection of all possibility,
2. Matching the data of fixed domain,
3. If (occurs)
4. Send as spam,
5. Else
6. For (if other than, go for length)
7. set the key,
8. If (key found)
9. Then go for (key++)
10. Key==3;

11. Spammed
12. Else not spam;
13. Exit;

A solution that deals with spam at the mail server is preferable for ISPs. Such a solution is centrally manageable and deals with the problem before spam and viruses reach end-users or consume network resources.

4. CONCLUSION

Informing users about the solution once is not generally enough. Users must form new habits to make the system effective, and their use of the system should be encouraged. As people who live in cold climates know well, layering provides the best protection against the pernicious and invasive threats to health and welfare. IT departments again facing the need to address rising spam volumes will do well to look at network level solutions such as the Symantec Mail Security nearly about 1 Million appliance to deflect a significant portion of spam so that other antispam defenses are better able to do their jobs in protecting server and PC performance and IT staff and user productivity at enterprise and service provider organizations. Attacking a spam filter can be thought of as an example of how machine learning algorithms can be defeated when an adversary has control of the workload. Thus, this paper is similar in spirit to, in which the authors used game theory to model attacks on spam filters. In that paper, the emphasis was placed on the game theoretic modeling; however, we showed that in the domain of spam filtering, a spammer can successfully attack current filters without using sophisticated game theoretic methods. As detection follows some algorithm, so of we are not worried about the matter of detection. Hence keeping the research of prevention is better than cure, lastly hoping to helpful to maintain spammed mail in research.

5. REFERENCE

1. Jöran Beel and Bela Gipp. Google Scholar's Ranking Algorithm: The Impact of Citation Counts (An Empirical Study). In André Flory and Martine Collard, editors, *Proceedings of the 3rd IEEE International Conference on Research Challenges in Information Science (RCIS'09)*, pages 439–446, Fez (Morocco), April 2009. IEEE. doi: 10.1109/RCIS.2009.5089308. ISBN 978-1-4244-2865-6. Available on http://www.sciplore.org.

2. Jöran Beel and Bela Gipp. Google Scholar's Ranking Algorithm: An Introductory Overview. In Birger Larsen and Jacqueline Leta, editors, *Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI'09)*, volume 1, pages 230–241, Rio de Janeiro (Brazil), July 2009. International Society for Scientometrics and Informetrics. ISSN 2175-1935. Available on http://www.sciplore.org.

3. Jöran Beel and Bela Gipp. Google Scholar's Ranking Algorithm: The Impact of Articles' Age (An Empirical Study). In Shahram Latifi, editor, *Proceedings of the 6th International Conference on Information Technology: New Generations (ITNG'09)*, pages 160–164, Las Vegas (USA), April 2009. IEEE. doi: 10.1109/ITNG.2009.317. ISBN 978-1424437702. Available on http://www.sciplore.org.

4. Jöran Beel, Bela Gipp, and Erik Wilde. Academic Search Engine Optimization (ASEO): Optimizing Scholarly Literature for Google Scholar and Co. *Journal of Scholarly Publishing*, 41 (2): 176–190, January 2010. doi: 10.3138/jsp.41.2.176. University of Toronto Press. Available on http://www.sciplore.org.

5. Z. Gyöngyi and H. Garcia-Molina. Link spam alliances. In *Proceedings of the 31st international conference on Very large data bases*, page 528. VLDB Endowment, 2005.

6. A.C. Cosoi, "A False Positive Safe Neural Network; The Followers of the Anatrim Waves," Proc. MIT Spam Conf., 2008.

7. E. Damiani, S.D.C. di Vimercati, S. Paraboschi, and P. Samarati, "An Open Digest-Based Technique for Spam Detection," Proc. Int'l Workshop Security in Parallel and Distributed Systems, pp. 559-564, 2004.

8. E. Damiani, S.D.C. di Vimercati, S. Paraboschi, and P. Samarati, "P2P-Based Collaborative Spam Detection and Filtering," Proc. Fourth IEEE Int'l Conf. Peer-to-Peer Computing, pp. 176-183, 2004.

9. P. Desikan and J. Srivastava, "Analyzing Network Traffic to Detect E-Mail Spamming Machines," Proc. ICDM Workshop Privacy and Security Aspects of Data Mining, pp. 67-76, 2004.

10. H. Drucker, D. Wu, and V.N. Vapnik, "Support Vector Machines for Spam Categorization," Proc. IEEE Trans. Neural Networks, pp. 1048-1054, 1999.

11. D. Evett, "Spam Statistics," http://spam-filter-review.topten reviews.com/spam-statistics.html, 2006.

12. A Gray and M. Haahr, "Personalised, Collaborative Spam Filtering," Proc. First Conf. Email and Anti-Spam (CEAS), 2004.

13. S. Hershkop and S.J. Stolfo, "Combining Email Models for False Positive Reduction," Proc. 11th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD), pp. 98-107, 2005.

14. AA Benczur, K Csalogány, T Sarlós, and M Uher. SpamRank – Fully Automatic Link Spam Detection. In *Adversarial Information Retrieval on the Web (AiRWEB'05)*, 2005.

15. A Kolcz and J. Alspector, "SVM-Based Filtering of Email Spam with Content-Specific Misclassification Costs," Proc. ICDM Workshop Text Mining, 2001.

16. A Kolcz, A. Chowdhury, and J. Alspector, "The Impact of Feature Selection on Signature-Driven Spam Detection," Proc. First Conf. Email and Anti-Spam (CEAS), 2004.

17. J.S. Kong, P.O. Boykin, B.A. Rezaei, N. Sarshar, and V.P. Roychowdhury, "Scalable and Reliable Collaborative Spam Filters: Harnessing the Global Social Email Networks," Proc. Second Conf. Email and Anti-Spam (CEAS), 2005.

18. T.R. Lynam and G.V. Cormack, "On-Line Spam Filter Fusion," Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), pp. 123-130, 2006.

19. Mehta, S. Nangia, M. Gupta, and W. Nejdl, "Detecting Image Spam Using Visual Features and Near Duplicate Detection," Proc. 17th Int'l Conf. World Wide Web (WWW), pp. 497-506, 2008.

20. V. Metsis, I. Androutsopoulos, and G. Paliouras, "Spam Filtering with Naive Bayes—Which Naive Bayes?" Proc. Third Conf. Email and Anti-Spam (CEAS), 2006.

21. M.S. Pera and Y.-K. Ng, "Using Word Similarity to Eradicate Junk Emails," Proc. 16th ACM Int'l Conf. Information and Knowledge Management (CIKM), pp. 943-946, 2007.

22. I Rigoutsos and T. Huynh, "Chung-Kwei: A Pattern-Discovery- Based System for the Automatic Identification of Unsolicited Email Messages (SPAM)," Proc. First Conf. Email and Anti-Spam (CEAS), 2004.

23. S. Sarafijanovic and J.-Y.L. Boudec, "Artificial Immune System for Collaborative Spam Filtering," Proc. Second Workshop Nature Inspired Cooperative Strategies for Optimization (NICSO), 2007.