

Identification of Factors Affecting Spamming – A Survey based Approach

Laxmi Ahuja
Amity Institute of Information Technology
Amity University, UP

Ela Kumar
School of information Technology
Gautam Buddha University
Greater Noida

ABSTRACT

The process of reducing spam has become much more difficult than before and this is because more than 85% of all email receive today constitutes as spam the large quantity of unsolicited email has also become a huge problem for IT relations in several universities. As spammers become more advanced, many more innocent students are becoming free in their web of fraud. Most IT-related spam messages revolve around the Phishing technique, where bank accounts and credit card information is targeted. On the other hand, spammers that target universities are more intent on gaining access to the universities' networks, a strategy that will enable them to spam more staff members and students. Present paper performs a survey on various web users to identify the factors affecting spam. Analysis shows that high keyword density content, excessive anchor text at the bottom of the web page, excessive reciprocal linking and keyword stuffing are the major factors that affect spam.

Keywords

Anti-spamming, web search engine

1. INTRODUCTION

SPAM, has become one of the biggest problems facing the Internet today. Spam has turned email into an unreliable Internet tool today. If you talk to some end users, they do not see much difference between Spam and the ordinary junk mail that mail carriers have delivered for years. In a recent survey, 93% of respondents reported dissatisfaction with the large volume of unsolicited email (spam) they receive. The problem has grown to the point where nearly 50% of the world's email is spam, yet only a few hundred groups are responsible. Many anti-spam solutions have been proposed and a few have been implemented [1,2]. Unfortunately, these solutions do not prevent spam as much as they interfere with every-day email communications.

The problems posed by spam have grown from simple annoyances to significant security issues. The deluge of spam costs up to an estimated \$20 billion each year in lost productivity -- according to the same document, spam within a company can cost between \$600 and \$1,000 per year for every user [3].

To prevent spam, both end users and administrators use various anti-spam techniques. An example of blocking spam includes the, use of filters. Despite this, dubious assumptions as those that specifically filter spam are least likely to

purchase from the Spammer. Another approach to block spam is to make use of blacklists that contain addresses of known spammers or compromised hosts. However, these lists have to be constantly updated because spammers have learned to counteract this by rapidly changing the origin of spam [4]. Some of these techniques have been embedded in products, services and software to ease the burden on users and administrators.

However in spite of a lot of research, no one technique is a complete solution to the spam problem, and each have trade-offs between incorrectly rejecting legitimate e-mail vs. not rejecting all spam, and the associated costs in time and effort [5].

2. PROPOSED SURVEY

To identify the factors responsible for spamming, we conducted a detailed Survey asking questions from the web search engine users regarding his/her preferences on some important issues. The survey consists of two parts, in first part there is some information about the respondent and in second section, total nine questions are there. The sample questionnaire is given here. The questionnaire has been developed to examine the views of search engine spamming techniques in technical educational institution in India.

Questionnaire consists of questions on the importance and perceived benefits of spamming intervention in technical educational institutes in India as well as the barriers and threats faced.

The prospective respondents of the questionnaire are section/department heads, faculty, students and other web users. Respondents were requested to give their rating based on their perception on the ideas listed as per the following scales.

S.No	Question Statement
	IMPORTANCE OF SPAMMING AND ANTI SPAMMING
1	What does user want from anti spamming software?
i)	High detection rate of spam
ii)	Low false positives
2	What method does user use to get protection from search engine spam?
i)	Using anti spamming software

ii)	Using need specific search engine (torrentz for searching movies)
3	What kind of web spam is most common?
i)	Content Spam
ii)	Link spam
iii)	Mirror Website
4	How do you combat web spam?
i)	Trustrank or HillTop
ii)	BadRank or SpamRank
iii)	SandBox
iv)	Anti Spam
5	How can one identify a Spam?
i)	If the receipt's personal identity and context are irrelevant.
ii)	The recipient has not granted permission to receive any information.
iii)	The transmission and reception of the message gives benefit to the sender.
6	How does a spam affect the user?
i)	Install viruses, Trojan, spyware.
ii)	Collect personal information
iii)	Compromises your network
iv)	Black listing your email system
7	Why is spam used?
i)	To sell a product
ii)	Promote fake product
iii)	Circulate computer viruses
iv)	Connect to illegal activities
8	Which search engine is safer?
i)	Google
ii)	Yahoo
iii)	MSN
iv)	Bing
9	What are your expectations from us in terms of deliverables? How would you qualify success?
i)	Higher Rankings in Search engines like Google, Yahoo, MSN, ASK etc.
ii)	Increased number of unique and targeted visitors
iii)	Sales and Registration Conversions

This survey was filled by 59 respondents from various domains and expertise. The results are compiled and analyzed and are put in the following section:

3. ANALYSIS

Q1 What does user want from anti spamming software?

Ans: 82% of the people want high detection rate of spam and 18% want Low false positives

Q2 What method does user use to get protection from search engine spam?

Ans: It means that user perform both the techniques to get protection from spam

Q3 What kind of web spam is most common?

Ans: It shows that majority of people feel that content spam is the most common web spam.

Q4 How do you combat web spam?

Ans: Result shows that Anti spam and spamRank are the two main web spam.

Q5 How can one identify a Spam?

Ans: Spam may be identified mainly if the recipient's personal identity and content are irrelevant. Also, a good number of people also think that the spam is used for the transmission and reception of the message give benefits to the sender.

Q6 How does a spam affect the user?

Ans: Viruses, Trojan and spyware are the main threat to the spamming. It has been the major factor responded by the majority of the users

Q7 Why is spam used?

Ans: Viruses and other illegal activities are the main sources for spam.

Q8 Which search engine is safer?

Ans: Google is undoubtedly the most safe search engine voted by majority of people

Q9 What are your expectations from us in terms of deliverables? How would you qualify success?

Ans: Lastly, higher ranking has been the preferred choice for maximum of the respondents

Based on the above survey, its analysis and literature available, we extracted some factors which are responsible for spamming. These factors are described in the following section.

4. FACTORS AFFECTING SPAMMING

Based on the above survey and further analyzing several facts in details, following factors are emerged, affecting spamming:

(a) High keyword density content – Google easily finds out contents that are containing number of keywords in the content beyond from the limit. The ideal keyword density is 5% to 9%. In order to avoid spamming even if you are repeating keywords, it is better to use synonyms of the keywords that you use.

(b) Excessive anchor text at the bottom of the web page – Putting anchor text at the bottom of the page is often recommended by most of the SEO experts. It is because they believe that it is easily seen by the search engine crawlers. However, a lot of ridiculous web marketing people overdo this technique that results to spamming. One or two links at the bottom of the page would be enough [8].

(c) Excessive reciprocal linking – Reciprocal linking is another effective SEO technique. This can only be done right when you are exchanging links with a website that bears the relevant content with your own website. Some of the

search engine optimizers are just out of their mind to exchange links with other websites, even buy them for an amount, not thinking if the website can help in boosting their own site [7]. The only way in making this tactic works is to link only with websites, that aside from being relevant, which are of the same page rank as yours. If your website has page rank of 3, then you have to link exchange with related websites that also possess 3 PR.

If you happen to link with websites that possess lower PR than yours, then your PR will be distributed to the site you link to. This means that your PR will, in a way, be reduced. Of course, if you link with websites with higher PR than yours, then it's a big luck for you if they accept your link exchange request.

(d) Keyword stuffing in the web menu- Search engine spiders hates seeing keyword the are overly stuffed anywhere on the page. Your website menu serves as the guide for your visitors in order to understand what your site is all about. It also serves as the giver of information on what services or product you offer.

(e) Directory submission to thousands of sites- This internet marketing technique is not really working for results that you are expecting. You might be expecting to get more traffic or higher ranking in the search engine. The fact about this technique is that it gives nothing but only asks more effort from you. It is recommended to submit your site only to site submission directories that are rarely visited or no visitors at all. You have to work for big guys on the site such as Google, Yahoo and MSN [6]. They have site directories. Open Directory Project is also one of the famous web directory submission sites. Only in famous sites, that you can draw more traffic or increased ranking in the search engine results page.

To be summarized

The five factors that spam the site are namely

- (a) high keyword density content,
- (b) excessive anchor text at the bottom of the web page,
- (c) Excessive reciprocal linking,
- (d) Keyword stuffing in the web menu, and
- (e) Directory submission to thousands of sites. These are just some of the several other tactics that often commits by careless search engine optimizers.

Spam becomes more "genuine"

Spamming in general has become a gainful business, and the scams that were once easily identified now appear more genuine. Spammers are using legitimate online businesses, such as PayPal or eBay, to trick recipients into giving up important account information. While most people would never respond to messages from senders that they don't know, this is a totally different situation when the request is being made from a sender within the IT network. To truly understand the characteristics of spam targeted at information technology, all students and members of the IT staff must become aware of all the spamming patterns and other key factors affecting the network.

Impact of the anti-spam on the email

Spam system classify the e-mails into wanted and unwanted messages, wanted messages are found in the inbox of the e-mails, while the unwanted messages are directed to the spam box. In many cases, anti-spams have several wrong estimations, although, the great roll of the anti-spam systems in term of privacy, security and protecting e-mails from any attacking cannot be ignored. From another view, spam cost

the companies too much by assuming the advertisement messages as spam [9]. Recently, most of the researchers are suffering from the spam false estimations; they tried to move toward approaches that minimizes the false positive estimations. Some of the analysis shows that anti-spam has directly affected the e-mail [2]

Issues and problems associated with spam email

There is no "silver bullet" for killing Spam. If that magic bullet existed, ISP's and corporations worldwide would have already pulled the trigger! Many customers think their Internet Service Provider (ISP) should be able to fix the problem. But Spam is a world-wide problem, and email systems around the world are not setup in a consistent manner.

5. CONCLUSION

In the present work, a test was conducted to determine the influence of Spam on user's satisfaction with his/her email system, but statistically significant evidence was not found that Spam affects overall user satisfaction with their email system. Improvements in user satisfaction need to be done as spam can be a nasty and often source of irritation. Individual user need to be protected from falling victims to frauds that can lead to financial loss, identify theft or even worse. The flood of Spam has threatened to make email unattractive to the extent of discarding it as a communications medium. It has also resulted in a decrease of user satisfaction towards the search results. On the positive side, an unexpected and fortunate side effect has also developed. It is believed that by using certain techniques, implementations could be developed that are specifically customized to a new approach of dealing with the threat of Spam.

6. REFERENCES

- [1] Sergey Brin and Lawrence Page, The Anatomy of a Large-Scale Hypertextual Web Search Engine, available at <http://infolab.stanford.edu/~backrub/google.html>.
- [2] Message Sniffer Anti Spam Filter - High Performance Scanner, available at <http://www.armresearch.com/products /index.jsp>
- [3] <http://www.symantec.com/connect/articles/anti-spam-solutions-and-security>, Anti Spam Solutions and Security
- [4] McAfee Anti-Spam: Protecting Your Organization from Spam, Phishing, and Other Unsolicited Messages (2006), available at http://www.mcafee.com/us/local_content /white_papers/anti_spam.pdf
- [5] Anti Spam Techniques, available at http://en.wikipedia.org/w/index.php?title=Anti-spam_techniques www.gfi.com
- [6] Impact of spam advertisement through e-mail: A study to assess the influence of the anti-spam on the e-mail marketing available at: <http://academicjournals.org/ajbm/PDF/pdf2010/4Sept/Raad%20et%20al.pdf>,
- [7] <http://www.windowsecurity.com/whitepapers/Impact-Reducing-SPAM-Part1.html>, SPAM - The Issues, Impact and Reducing SPAM (Part 1)
- [8] <http://internetmarketingxpert.com/2009/05/5-factors-of-website-spamming-webmarketing-expertscomau-part-3>.
- [9] Mostafa Raad, Norizan Mohd Yeassen, Gazi Mahabubul Alam, B. B. Zaidan and A. A. Zaidan, Impact of spam advertisement through e-mail: A study to assess the influence of the anti-spam on the e-mail marketing, African Journal of Business Management Vol. 4(11), pp. 2362-2367, 4 September, 2010