Dragomir V. Dimitrov. Spam - An experiment exploring the relation between the amount of spam received after registering in different websites. A Master's Paper for the M.S. in I.S. degree. April, 2005. 28 pages. Advisor: Robert Losee

Some web sites require registration in order to use their free services. But is this really free? The goal of this paper is to describe an experiment that was conducted to explore the relation between registering in different websites and the amount of junk emails received. The whole experiment was divided into three periods and the collected data from them was compared. The paper contains also an analysis of the results and recommendations for the users on how to avoid or at least limit the spam in their mailboxes.

Headings:

Spam

Email marketing

World Wide Web

SPAM - AN EXPERIMENT EXPLORING THE RELATION BETWEEN THE AMOUNT OF SPAM RECEIVED AFTER REGISTERING IN DIFFERENT WEBSITES

by
Dragomir V. Dimitrov

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

April 2005

Approved by

_____

Robert Losee

Table of Contents

Introduction

It was the fall of 1971 when Ray Tomlinson, a Principal Scientist at BBN Technologies (http://www.bbn.com), sent the first email message in his Cambridge MA lab. He used a program called CPYNET to send the message between two physically connected computers. Tomlinson also used for the first time the "@" sign to separate username from host name in mail addresses. It took almost two decades to see the proliferation of this service. Nowadays email services are probably the most used form of communication. The people use emails to communicate with their friends, colleagues and relatives. The advantages of the email are well known [1]. This is usually a free service, very fast (emails usually reach their destination in a matter of minutes or seconds.) and accessible from everywhere and any time, even through cell phones. Although there are several email providers who offer options for additional payable services, such as larger amount of mailboxes, firewall protection and online storage space, the mass email services are usually free. The users can send messages at any time that they choose. The recipient reads them at a time of their

choice. In comparison to the telephone, this is a great way to not interrupt people especially these in other time zones. A user can send an email to multiple people or to a group with just a few clicks. An email message can contain not only text, but also pictures, audio, video or other types of files. Some email providers offer encryption of the emails in order to protect the content of the messages from malicious people. Emails also offer other useful features as forwarding messages, keeping address books, anti-virus scanning, check spelling and spam filtering.

Nevertheless the email is not perfect. Emails can compromise the security of an organization because sensitive information can be easily distributed accidentally or deliberately. There are also access constraints – hardware, software or Internet. Security is low in email communication and it can be used for distribution of viruses, Trojans, spyware or other malevolent programs. The email services provider often does not require identification when one wants to create an email account. Thus everyone can create one or more fake accounts and use them to abuse other people. But the most annoying thing in these days is the spam emails. There are several definitions of what exactly spam is. According to Miriam Webster spam is: "unwanted/unsolicited usually commercial e-mail sent to a large number of addresses" [2]. Another definition from an online dictionary is:

> "Unsolicited "junk" e-mail sent to large numbers of people to promote products or services. Sexually explicit unsolicited e-mail is called "porn spam." Also refers to inappropriate promotional or commercial postings to discussion groups or bulletin boards" [3]

In today's business world, product advertising is of major importance. Businesses use a vast array of advertising methods, such as television, radio, newspaper, Internet, emails. One of the features of the electronic mail, the possibility to send one message to large amount of people, facilitates spammers. They used this free service to overflow our mailbox with different kinds of advertisement. Table 1 presents a list of types of spam distribution [4]:

Money making opportunities	35%
Adult entertainment	11%
Direct marketing product and services	10%
Informational and how-to guides	9%
Internet services, computer hardware and software	7%
Other products and services	25%
Non-commercial	3%

Table 1. Distribution of spam senders

Email is a free service for the sender, but the recipients pay the bill. Spammers use automatic software, with prices varying from $100 to $600 [5]. Since there is no "free lunch" someone has to pay the cost. Internet services providers implement spam filtering programs, which cost money, bandwidth to deal with high volume Internet traffic and staff to cope with problems caused by spam. Regular users spend time to set filters, download, read, and delete messages and submit complaints. Their privacy is also endangered.

There are a lot of studies dealing with the actual price paid by the user and ISPs. Computer Mail Services has created a calculator that projects the cost of spam:

> "It shows that a company with 500 employees, each of whom receives five junk e-mails per day and spends about 10 seconds deleting each one, can expect to lose close to $40,000 per year in wasted salaries and 105 days in lost productivity" [6]

A study conducted in 2001 from the European Union calculated spam costs to be about $8.6 billion a year worldwide. Nowadays, 50 to 80 % of the emails are spam and this cost is even higher. Therefore it is very important to find ways to limit spam. In order to fight against spammers, we first have to identify its sources. Spamhouse.org published a list of 200 known spam operations responsible for 80% of the spam received in Europe and North America [7]. Since the United

States ratified CAN-SPAM [8] act on January 1, 2004 it is interesting that about 75% of these website are operating in the United States. The battle with spammers will be really tough, but there is a real chance that email will be overwhelmed and subsequently discredited as a method of communication. Some studies claim that because of the spam, users decrease their email usage. That is why it is so important to limit the amount of unsolicited junk emails. The next source of spam is hackers who break into innocent people's computers using high bandwidth connections and hi-jack processing power and resources to send spam.

Literature Review

The name of the first article in my literature review is "Spam!"[4]. This is a study which was performed by Lorrie Faith Cranor (AT&T Labs) and Brian A. LaMacchia (Microsoft Corp). It was published in August 1998 in Communications of the ACM, Volume 41, Issue 8 Pages: 74 – 83. The authors of the study provide a brief background of the growth of unwanted/unsolicited emails, list factors contributing to it and discuss various techniques available for reducing such traffic. According to them these mechanisms include filtering, counterattack, opt-out lists, channels, fees and legislation.

The article is related to my experiment, because it presents interesting history information about the proliferation of the spam. It is necessary to explore the beginning of this problem and the idea behind sending spam. The understanding of the principles which spammers use to send emails would help in creating more reliable filters. The article also provides a very interesting data analysis about the content of the spam emails in the period of March – May in 1997. Although it may seems that this analysis is a little bit obsolete, in such fast developing sector as Internet, it carries valuable data for the content of the spam. Probably a contemporary study will give similar results about the distribution of the spam categories. The article describes a study about the number of spam messages received regarding the day of the week. It concludes that there was a decrease in the number of emails received during the weekends and major holidays. The authors of the study describe also how the automated spam filters works. They mention

- filtering solutions,
- counterattack solutions
- opt-out lists
- channels
- payments

- referral networks
- fee restructuring.

The next item is a research article titled named “Why Am I Getting All This Spam?” [9]. This article was published in March, 2003. The initiator of the study is the Center for Democracy and Technology. They wanted to determine what the sources of spam are. To do so, they set up hundreds of different e-mail addresses and then waited six months to see what kind of mail those addresses were receiving. They found out that the vast majority of the spam received, over 97% of it, was delivered to addresses that had been posted on the public Web and public postings to Usenet newsgroups and forums. This statement was proved by an experiment, in which they removed some of the email addresses “in order to determine how long an e-mail address, once placed on the public Web, would continue to receive spam after its removal”. The outcome is a significant decrease in the number of spam received. The instigators also proposed some valuable tips in order to avoid the receiving of spam. The authors of the study registered their emails in over 30 well – known websites, such as amazon.com, ebay.com, cnn.com, monster.com, expedia.com etc. Probably not all of them are so dangerous, but some of them may sell our email addresses

to third party companies. This has to be a red light for the people and force them to have multiple addresses for different web services.

Since the spam became one of the biggest computer annoyances, the United States Senate issued CAN-SPAM Act of 2003 (Controlling the Assault of Non-Solicited Pornography and Marketing Act). The law requires, among other things, that recipients be allowed to opt out of being included in a marketing mailing list, simply by clicking a link on an e-mail notice from the business companies. In addition, the CAN-SPAM Act establishes harsh penalties for senders whose e-mail messages fail to meet its requirements. PC World has tested this act by conducting an experiment [10]. They signed up in 100 heavily trafficked US websites (11 categories, such as sports, travel, retail, gambling, and entertainment) and after that tried to unsubscribe from receiving marketing emails. According to the CAN-SPAM act 10 days after unsubscription the user should not receive any emails. PC World states that 15 of these websites did not stop sending spam and surprisingly one of these websites is amazon.com. After contacting Amazon by phone, the messages from them have stopped. The CAN-SPAM act shows the desire of the US government to deal with this annoyance. Obviously there is a lot to do in order junk emails to be eliminated, but this is only the first step.

Next article is “Should spam be on the menu”, published in *Communications of the ACM* by J. Sipior et al in 2004 [11]. The motto of the article is “Email marketing is an excellent, low-cost way to reach consumers, but spammers are currently looting much of its potential”. The authors describe the history of the spam and point out that the first spam was sent in 1975. 1994 is the year of the start of the real proliferation when the law firm of Cantor and Siegel spammed more that 6000 Usenet newsgroups. The authors also affirm that about 60% of all Internet email is spam, which leads to about 25% decrease in email usage. They make several recommendations for dealing with spam: using opt-in lists and labeling emails with ADV and ADV-ADULT, which comes from advertisement and adult advertisement. This would facilitate spam filtering programs, but the question is “Are spammers willing to include these amendments?”

Methodology

The research question, which I would like to explore, is “What will be the amount of unsolicited message a user will receive if registered in a particular web site?” Some websites require registration in order to use their “free” services. I decided to conduct an experiment to investigate, which websites are willing to misuse their client email

addresses. Though spam could be received not only from websites but also from individuals in my study, I concentrated in websites as sources of spam. I created 30 emails using Yahoo mail service and registered each of them in 30 different websites. I wanted the name of the accounts to be in unison with my project, therefore I named them spammasterpaperXX@yahoo.com, where XX is from 01 to 30. Yahoo offers a pretty big amount of space (250 MB), so there is no danger to run out of space. In addition, it is expected to increase this space to 1 GB in April. I choose a wide variety of web places. Table 2 shows the full list of websites and their category distribution:

http://riseofchaos.com/index.php	Gaming
https://www.pokerroom.com	Gaming
www.playapal.com	Gaming
http://freedom.qfetbays.biz/SendForm/sendform.aspx	Mortgage
http://freedom.poepneat.biz	Mortgage
www.dating.com	Dating
www.match.com	Dating
http://www.beted.com/default.aspx	Betting
http://www.freelotto.com	Gambling
www.gambling.com	Gambling
http://www.grandonline.com/welcome.php	Gambling
http://www.easywarez.com/	Warez
http://www.orgsites.com/fl/hacker/index.html	Warez

www.hackthissite.cim	Warez
http://www.free-software-free-software.com	Free software
http://giftfox.com	Free stuff
http://thefreeproject.com	Advertising
https://www.comcast.com	Advertising
https://www.compusa.com	Computer Store
www.cdw.com	Computer Store
www.bestbuy.com	Computer Store
www.expedia.com	Travel agency
http://www.travelocity.com/	Travel agency
slashdot.org	News
http://groups-beta.google.com	Usenet
www.imdb.com	Movie database
www.moneybookers.com	Work online
www.weather.com	Weather Channel
www.aol.com	ISP
www.real.com	Real player

Table 2. List of the websites and their category

In order to be representative I used one account in only one website. I tried to choose a wide variety of websites to determine which category is most "dangerous". My suggestion is that probably gambling, mortgage and advertising web sites will send the biggest part of spam. I left one email address in the Google's newsgroup. Usenet are

one of the places, from where spammers get our email addresses. They use software robots to scan Internet web pages and collect email addresses. I selected several web sites of big companies that have good reputation, but could be possible spam-senders: Weather.com, Real.com, Imdb.com and Aol.com.

Results and analysis

The purpose of the experiment I made was to check what will be the amount of spam received when registered in different websites. There is a contradiction what exactly spam is. Some people claim that spam is every unwanted message they received. Others prefer to receive emails with promotions and special offers from the business companies. They do not qualify this as spam and often use these messages to buy products. Therefore, it is quite difficult to distinguish spam from emails that could be in benefit for the users.

To facilitate my result analysis, I made "screenshots" of the number of emails I received every month. I created my email accounts in the second week of January and registered them in the list of target websites. Then I counted the number of emails in each account on the 1st day of February, March and April. Probably a period of about three

months is not enough to create general conclusions, but at least will provide patterns of what can be expected further.

I created a chart for each of these months and the February's data is presented in Figure 1.

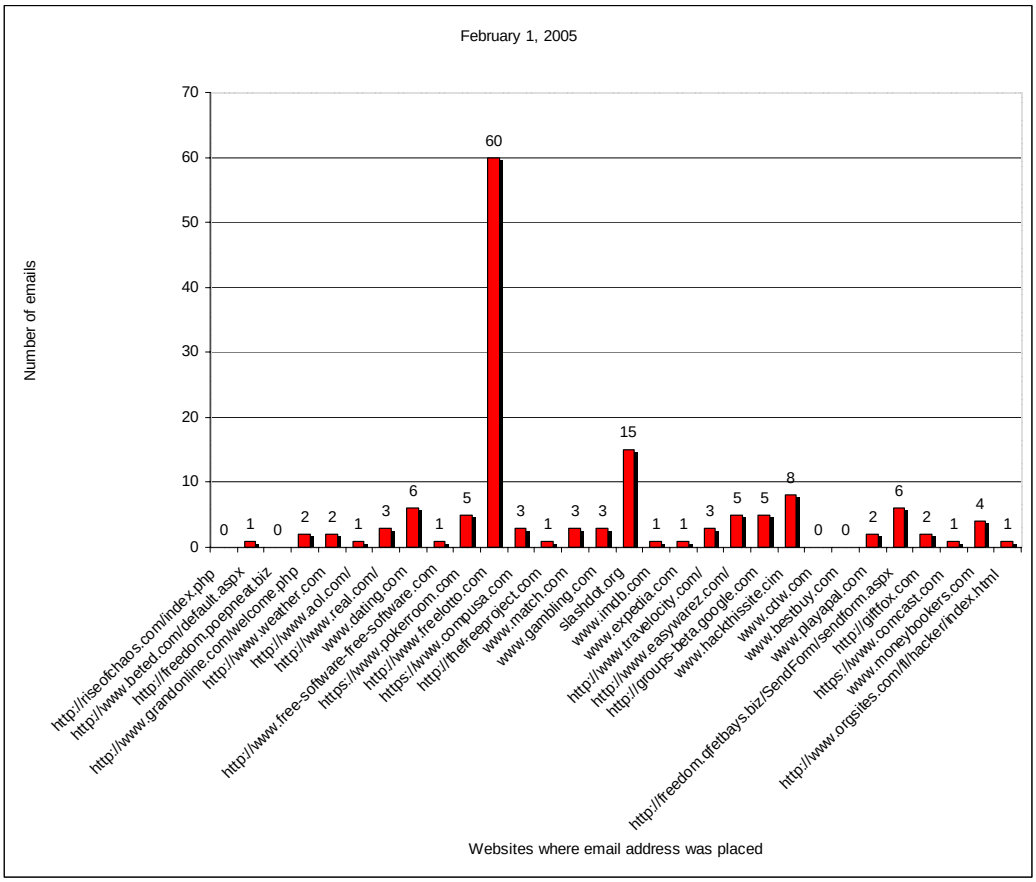

Figure 1. Number of emails received in the period January 15 - January 31

My suggestion before performing the experiment was that websites offering free products (iPods, hardware, software applications), gambling and gaming portals, will be most intensive

sources of spam. In this two-week period I received total of 145 emails. The diagram shows an uneven distribution. There is an obvious leader – Freelotto.com. They sent 60 emails, which are 41% of the total number of spam during this period. The second place in the chart is for Slashdot.org. However, the emails sent by them, were not exactly spam but electronic newsletters. Everyday, they send a message to the people registered in their website, containing a newsletter and headline mailer with the news related to the Information Technology area. These emails comprise useful information especially for people interested in the Internet, information security and politics. There are no significant trends in the amount of emails received from other websites, but I have to mention that there were four "good" websites, that did not send a single email (Table 3):

riseofchaos.com	Gaming
freedom.poepneat.biz	Mortgage
www.cdw.com	Online store
www.bestbuy.com	Store

Table 3

Next counting I made was on March 1, which includes the period from January 15 to February 28. The graphical representation of the data is in Figure 2.

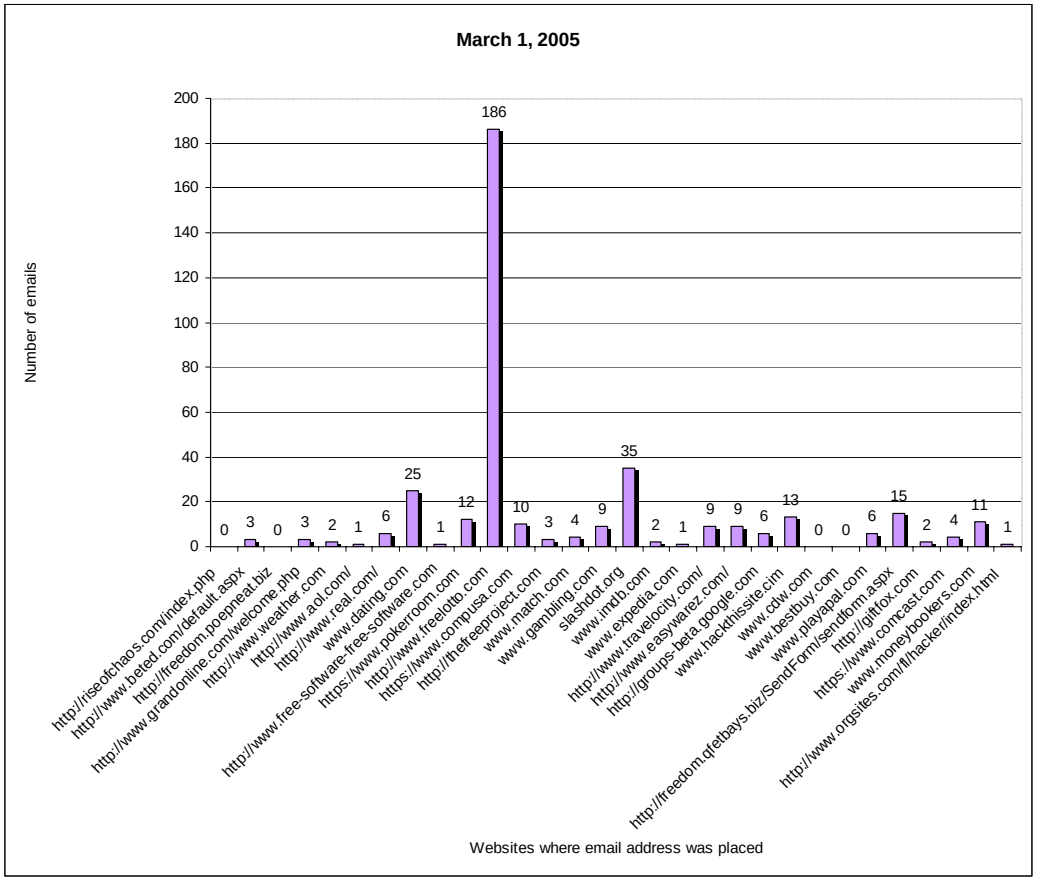

Figure 2. Number of emails received in the period January 15 - February 28

Similarly to the first period, Freelotto.com leads in the number of sent messages. They needed only one month to increase three times the amount of spam to 186. In the presented period I received 379 emails. Freelotto.com enlarges their portion, which totals 49% of all emails.

Slashdot.org holds the second place with 9% and the third is for Dating.com. The four sites that did not send spam in the first period, kept their behavior and I did not received emails from them. There are ten websites with no change in the number of sent emails, including the four "good" mentioned earlier.

The next period, for which I collected data, is from January 15 and April 1 – Figure 3. There were not any big or surprising changes. The top three senders kept their positions. Freelotto.com doubled the amount of emails, totaling 339 for the whole period. During the experiment; which continues about two and half months, I received 739

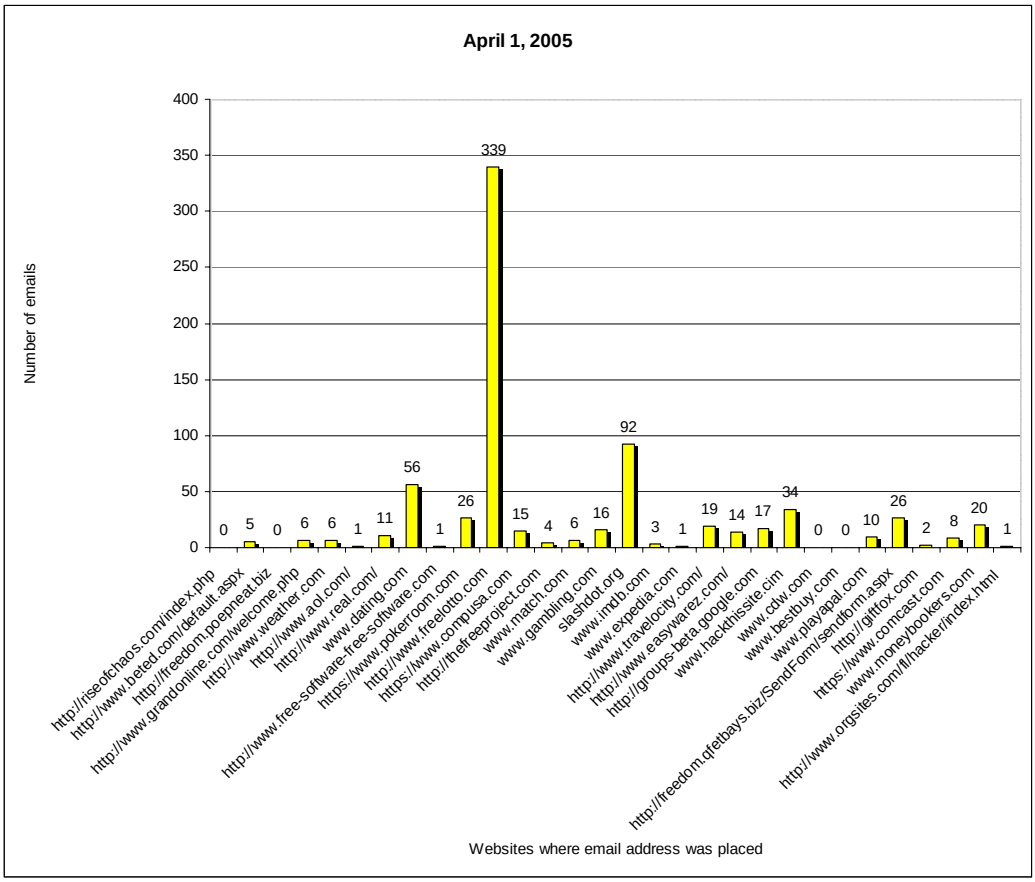

Figure 3. Number of emails received in the period January 15 - March 31

messages. The chart showed that Freelotto.com has sent 45% of all emails. Slashdot.org continued to send me their newsletters – 92 for the whole period. Dating.com is again on the third place with 56 emails. I registered one of the accounts in Google's newsgroup, which is a Windows XP forum. The actual number of emails was 221, but I classified only 17 of them as spam. The others are questions and answers from the members of that forum. According to the data collected, there are ten websites that sent between 0 and 3 emails for the entire experimental period. Some of them sent only one welcome email

or request for email verification. Therefore these ten websites could be qualified as safe places.

The websites I choose can be distributed into ten categories. It was interesting to find out, which of then is the biggest spam - generator. Figure 4 represents this distribution.

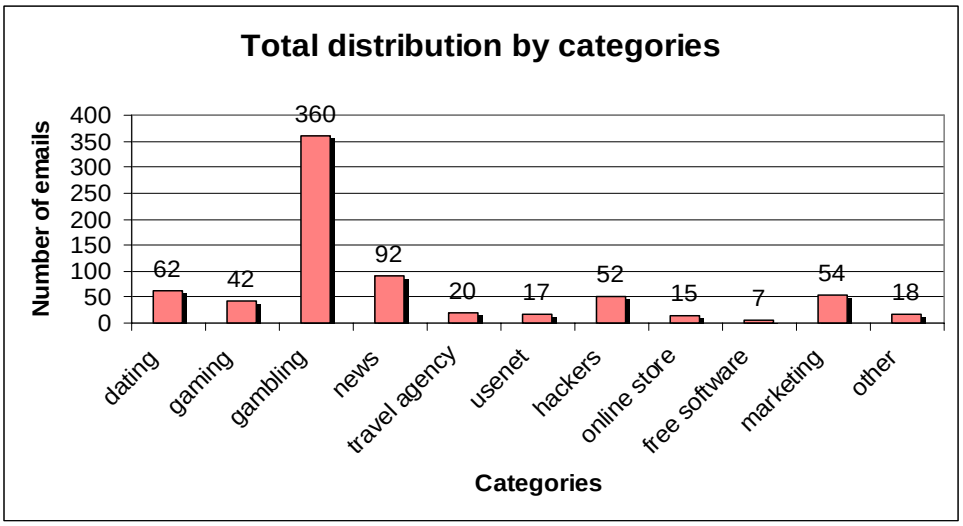

Figure 4. Total distribution by categories

With 360 emails, gambling is on the top, which is about 49% of all spam. 339 of them came from Freelotto.com. Next category is the news. The major part of these 92 emails is the newsletters sent by Slashdot.org. As mentioned earlier, the messages received from this websites are not exactly spam. Dating, hackers and marketing have slight difference in their spam -score, respectively 62, 52 and 54 emails. In the beginning of the experiment I expected to receive lots of spam from the hackers' websites I registered. However, the result showed that these 52 messages are only 7% of all spam that I got. I could

conclude from this number that we can get viruses, Trojans and other malicious programs from hackers, but not so much spam.

Conclusion

Email marketing became a cheap and effective way for advertising services and products. It can efficiently reach a target market with personalized communication, at a cost lower than any other alternative media. This paper describes an experiment in which I registered 30 email accounts in 30 websites. My idea was to find out which of these websites are most dangerous in terms of sending spam. The results showed that gambling sites sent about 50% of all spam that I received. In general, these emails contained "free offers", financial services and money - making opportunities.

The conducting of this experiment is important, because it tried to distinguish good from bad websites. The users will know which websites are safe to register and which ones to avoid. As the results demonstrated, there are about ten websites that can be classified as good and other nine, with more than 20 emails, as spam-senders.

Email is a low cost, convenient, fast and relatively reliable service that facilitates our work and life. It is an irreplaceable way of communication especially for long distances. However, the

proliferating of the spam jeopardizes it as service. The users spend money ant time to get rid of this annoyance. The business world lose big amount of money dealing with spam, money that could be used for innovations, investment or development. Thus, must be found effective way to restrict spam propagation.

Further Research

This experiment was limited in time. It lasted only two and half months. The data collected, probably is not enough for making general conclusion and establishing patterns. That is why this experiment could be expanded for a longer period of time – one or two years. The increasing the amount of observed websites would help for collecting more exhaustive data as well. The CAN-SPAM act requires from websites to include unsubscription or opt-out option. Other direction, in which this experiment could continue, is attempting to unsubscribe from these websites and investigating the effect of this action. I will continue to watch the behavior of these sample websites further. It will be interesting to find out what will be the ratio between the big spammers and other websites.

Recommendations

There is a growing concern that the volume of the spam sent each day may increase significantly, even "kill" email as service. This is probably the worst scenario about electronic services. Understanding how spammers get our email addresses and what can be done for our protection, definitely will make our life easier.

The modern Internet society started to battle against spam-senders, by using spam-filtering program, government acts and teaching users how to avoid spammers. Recently in Leesburg, VA a man was convicted in the nation's first felony case against illegal spamming and was sentenced to nine years in prison for "bombarding Internet users with millions of junk e-mails" [12]. This case classifies spamming as a crime which becomes a precedent that would hopefully mark the beginning of a decrease in the number of spam in our mailboxes. But the most helpful action is self-protection. There are several precautions that the user can follow in order to defend from aggressive advertisements. The most important are to

- Be careful about disclosing email address and keep their primary addresses only for people they know
- Review the privacy policies of websites to make sure that they do not inadvertently agree to release their email addresses to third

parties. Probably this is the last thing that a user will read, but it is important to check for hidden clauses

- Block images which spammers can use to confirm their email address
- Be very careful when publishing email addresses in news groups, chats and forums
- Disguise email addresses posted in a public place. For example using name at domain dot com instead of name@domain.com
- Use multiple email addresses for different purposes – emails only for well-known people, such as friends, relatives and colleagues, and other for registering in websites, web postings and electronic commerce
- Use longer email addresses – spammers exploit software robots for sending spam and guess our emails. They use a method called dictionary attack to find valid addresses. The longer and complicated names we create, the less spam we will get
- Install anti-spam filters. Although they are not perfect and have some weaknesses, they could decrease significantly the amount of spam in our mailboxes.

Users' behavior is very important when they receive spam. A recent study, conducted by Mirapoint and Radicati Group, found that

"nearly a third of e-mail users have clicked on links in spam messages" and "one in ten users has bought products advertised in junk mail" [13]. These numbers are really disturbing and could explain the proliferation of the spam. The people should be educated about the troubles related to email as service. They could get not only spam messages, but also could be target of phishing or scamming.

Limitations

During the experiment, I had several problems such as defining what exactly spam is. Some people could find a message that advertise a product, service or promotion as helpful, but other as a big annoyance. I tried to classify a message as spam if it contains one or more of the well-known spam words or the advertisement was too aggressive [14].

The time was also a limitation for the experiment. I spent only two and half months for collecting data. A longer period of observation would give more information for analysis, which would increase the experiment's representativness. Other improvement for the study is the including of more websites from more categories.

References

1. Berghel, H. (1997). Email – the good. The bad, and the ugly. *Communications of the ACM*, 40(4), 11-15

2. http://www.m-w.com/cgi-bin/dictionary?book=Dictionary &va =spam

3. http://www.getnetwise.org/glossary.php#S

4. Cranor, L.F., & LaMacchia, B.A. (1998). Spam!. *Communications of the ACM*, 41(8), 74-83

5. dir.yahoo.com/business_and_economy/business_to_business/ marketing_and_advertising/direct_marketing/direct_email/software/

6. ww.businessweek.com/technology/content/mar2002/tc2002031_8613.htm

7. 200 Known Spam Operations responsible for 80% of your spam. http://www.spamhaus.org/rokso/index.lasso

8. http://www.ftc.gov/bcp/conline/pubs/buspubs/canspam.htm

9. Center for Democracy & Technology (2003) Why Am I Getting All This Spam. http://www.cdt.org/speech/spam/030319spamreport.shtml

10**.** Spring, T. (2005) Spam Law Test. http://www.pcworld.com/news/article /0,aid,118702,pg,4,00.asp

11. Sipior, J.C., Ward, B.T., & Bonner, P.G. (2004). Should spam be on the menu? *Wireless sensor networks*, 47(6), 59-63

12. Spammer sentenced to 9 years. http://msnbc.msn.com/ID/7432555

13. http://news.bbc.co.uk/2/hi/technology/4375601.stm

14. http://wiki.wordpress.org/SpamWords