Global Information Assurance Certification Paper

Case Study - Assessing the Impact of Unsolicited Commercial E-mail in a Large Corporation

GIAC Security Essentials Certification (GSEC)
Practical Assignment – Version 1.4b, option 2.

Joseph McComb
January 26, 2004

Table of Contents

Figures and Tables

Abstract

Unsolicited commercial e-mail has become an increasing issue in corporate environments. This case study examines the impact of unsolicited commercial e-mail (also known as spam) on the productivity of employees in the research division of a large global corporation. Questionnaires were utilized to assess the volume of spam that employees received in order to conduct an impact study. The results suggest that 68.8% of the internet messages received by the research division are spam and that employees receive an average of 4.46 spam messages per day. An employee was timed while sorting 315 e-mail messages that were a mixture of legitimate e-mail and spam. The employee averaged 3 seconds per message to determine if the message was legitimate.

A gateway filtering service was implemented that inserts a tag in the subject line if the filter determines that the message is spam based upon the content. Questionnaires indicate that the filter correctly identifies 91.8% of the spam e-mail messages. Using client-side filtering, employees were able to separate out the tagged messages. An employee was timed while sorting the filtered messages and averaged 0.67 seconds per message. This suggests that the research division employees will regain 2.33 seconds per spam message saving 10,500 hours per year in lost productivity.

Before – Origin of the Impact Study

Introduction

Unsolicited commercial e-mail has become an increasing problem in large corporations. Estimates place the volume of unsolicited e-mail advertising at 20%-70% of the total volume of internet e-mail (Lagesse, 2003). Other estimates suggest that unsolicited commercial e-mail costs $874.00 per employee per year in lost productivity (Nucleus Research, 2003). Samples of employee unsolicited commercial e-mail indicate that often the e-mail originates from forged return

addresses and contains characteristics to evade gateway filters (presented in this paper). This fraudulent unsolicited commercial e-mail is known as spam to the majority of e-mail users and ranges from an annoyance to a moral insult, as in the case of pornographic spam.

In Spam Filtering in a Small Business Environment (2003), Richard Snow presents an excellent analysis of implementing gateway filtering in a small business environment. The present case study differs in that it focuses on determining the impact of spam in a large enterprise and focuses on assessing the impact of the spam reducing measures via employee questionnaires. In particular, the large enterprise presented issues that were not observed in the Small Business case study. For example, one driver for this case study was that much of the upper management did not receive unsolicited commercial e-mail and had not experienced the issue first hand. Because the upper management had only anecdotal evidence reported by employees on the impact of spam, a six month impact analysis was requested. This case study documents a methodology that was used to collect data on the impact of spam so that upper management could approve a solution that was acceptable to the business and based on a cost/benefit analysis. This case study also examines the impact of the gateway filtering solution upon the user community.

The Company

The statistics were collected in a global Fortune 500 company of approximately 60,000 employees that is broken into a divisional hierarchy such as manufacturing and research. In this paper, one division is responsible for handling infrastructure aspects for the entire company. There are two key infrastructure teams: (1) the e-mail server team and (2) the e-mail client team (in this case a single person). Another key player is the information security unit for the research division, which is the unit I work in. The overall breakdown of the teams with their responsibilities is:

1. E-mail Server Team: This team is responsible for maintaining the internet mail servers.
2. E-mail Client Owner: This person is responsible for maintaining or enhancing the e-mail client.
3. Research Division Information Security: This unit is responsible for the information assets of the research division.

The Infrastructure Division and the Research Division each have their own separately funded information technology (IT) units and each are approximately 700 employees in size. This case study focuses on the research division that represents 10,000 employees and amounts to 1/6 of the company.

Chronology of the Problem

In July of 2002, a director in the research division sent a message to my director asking if there was anything that could be done about the spam he was receiving in his business e-mail. By August of 2002, I was asked to prepare information on what other large companies were doing to control spam. This information would be presented to the IT management committee of the research division in October along with recommendations. At that time, I contacted the e-mail server team and found that they were also preparing a presentation for the infrastructure leadership committee, with recommendations on how to control spam. The presentation by the e-mail server team was scheduled a couple of weeks before my presentation and they felt that their proposal to filter spam at the internet gateway would be accepted.

In October of 2002, the e-mail server team presented the infrastructure leadership committee with a proposal to reduce unsolicited e-mails by filtering messages that were passing through the internet mail servers. The leadership committee rejected the proposal and issued a formal statement that the company would not implement spam filtering because the company could potentially lose business or personal e-mail to the filters. Ultimately, this meant that they would not supply the e-mail server team with funding for gateway filtering. Instead, the leadership committee recommended that individual employees should utilize the junk mail filters included with Outlook, the primary e-mail client.

A couple of weeks after the presentation by the e-mail server team I presented information to the IT management committee of the research division. In the presentation, I noted that the company had a policy that the employee could utilize internet e-mail for personal use provided that it did not adversely affect productivity (or break other rules in the policy such as unauthorized distribution of proprietary information). This meant that the employee could subscribe to mailing lists or apply for online accounts with mail order companies using the company e-mail address. Some companies, such as Trans Pacific, (DiSabatino, 2002) limit the amount of spam by enforcing a business use only policy for their e-mail accounts. The current policy had been implemented recently and could not be changed easily without more management review.

The presentation also focused on current industry direction. Research by Osterman Research reveals that 54% of companies have implemented anti-spam technology (Osterman Research, 2002). However, other reports have shown that there is no single strategy that companies have employed to reduce unsolicited e-mail (Gaspar, 2002). Many corporations employ anti-spam filters that block known spammer IP addresses or block specific content. Some companies also train users in the use of client-side filters or establish policy to prevent employees from disclosing their e-mail address to spammers (Gaspar, 2002).

Near the beginning of the presentation, a question was asked which illuminated an underlying issue as to why the company was reluctant to filter at the gateway.

One of the committee asked, “Can’t you just set a rule in Outlook to block the e-mail from the company sending the spam?” It turned out that some of the committee did not receive spam and did not realize that many spammers change their address so that the messages are difficult to filter.

A week after my presentation, I received a forwarded voice message from the Vice President of the research division to work with the e-mail server team to rigorously quantify the amount of spam that the research division was receiving over a 3 to 6 month period. Based on the results I was to work with the e-mail server team to implement a solution to reduce spam for the company.

The before situation can be summarized in four key points. First, the company had committed to a policy that allowed employees to use the company e-mail address for personal activities. Second, the infrastructure leadership did not wish to commit to filtering at the gateway because potentially business e-mail would be lost. Third, some of the upper management had not experienced spam first hand and did not know the extent of the problem. Fourth, there had not been a formal spam impact study on the company’s employees to determine if it was affecting productivity. In summary, the management did not know if spam was actually a problem and needed a formal impact study to assess if a solution would be needed.

During – Assessing the Impact of Spam

Planning the Impact Study

The impact study to quantify the amount of spam received by the research division presented a couple of initial issues. First, all of the company e-mail is handled on an enterprise scale through a set of mail servers that makes it difficult to separate the divisional accounts out. Second, since the e-mail server team had recently had their proposal to filter at the gateway rejected, their upper management would be unwilling to commit resources to the project. In talking to the head of the e-mail server team, he expressed that he “just wanted this problem to end.” As it turned out, he often received e-mail from employees in the company who wanted something done about the spam problem and he was tired of the problem. In the end, he agreed to the project as long as it did not heavily impact his staff.

In December 2002, I met with the e-mail server team to discuss project planning and implementation. The crucial part of the meeting was planning what statistics would be needed to assess the impact of spam on the productivity of employees. In addition, we examined the following solutions to reduce spam which had been implemented by other companies:

1. Disposable E-mail Aliases: Lucent developed a system of revokable e-mail addresses in the 1990s (Lucent Technologies, 1998). We considered allowing each e-mail user a second “disposable” account.
2. Filtering with tagging message: Instead of deleting the message, a gateway filter tags the message and passes it on to the user, so the user must still look at it. The user can sort the spam away from untagged messages.
3. Filtering with deletion of the message: A spam filter examines messages at the gateway and deletes a message if it is suspected to be spam.

The solution of disposable e-mail addresses was interesting, but would introduce more maintenance for the additional e-mail accounts. In particular, the solution of tagging the message was appealing because it would give the business an opportunity to see how well the filter was working without losing legitimate e-mail.

Early in the meeting, one of the team asked the question “how do we define spam?” The point was that one user could consider certain e-mail as spam while another user might consider the same e-mail as legitimate. Actually, this situation does occur on a limited scale, as was shown through some of the surveys: some users consider e-mail from mailing lists of legitimate companies as spam, and other users see this as legitimate. The survey results described later in this paper suggest that this legitimate junk e-mail is around 3-5 % of the e-mail that users consider as spam. However, since we did not know this in advance, we decided to attack the project on two fronts, surveys from e-mail users and spam statistics from the internet gateway.

The e-mail server team would use a spam content filter which was supplied with the internet gateway to count the amount of spam passing through the gateway every week for a period of six months. By default, the filter contained 30 words commonly found in spam (such as sex or mortgage) and used wild cards to catch variations of the key words. The filter was not used to alter or delete the message, but would register an event that a message containing the word had passed through the filter. Because the filter had a limited word list, we expected that this estimate of the amount of spam would be low.

The second set of statistics was a series of user surveys designed and conducted by me. This set of surveys would answer the questions about the actual impact of spam on the individual user. Although it is easy to measure how many messages a filter may catch at the e-mail gateway, it is impossible to know if the filter is doing a decent job without user surveys that assess what the employee experiences. In addition, the surveys would allow me to separate out my division from the rest of the company, and focus on the impact of spam in the research division. Last, the surveys allowed me to collect information on how the users defined spam and could provide information on the spam definition question raised earlier in the meeting.

Defining Spam

The humorous thing about being selected to work on this project is that I get very little spam myself. In fact, when I started the project, I did not get any spam at all! Thus, when I began my research on spam I had asked everyone I knew if they got spam. Even better, I asked if they could send me a days worth! I plowed though spam messages, reading the headers, pinging servers and verifying e-mail addresses. Finally, I wrote up a short description of the common characteristics I found in the messages:

1. Spam messages often have forged from addresses so that the spammer's identity cannot be traced. Sometimes, the spammer will forge the recipients own e-mail address as the from address. Even if the return message is not forged, most spammers change their e-mail addresses frequently, so that they cannot be blocked by sorting on return e-mail address.

2. Spam messages often contain HTML formatting which is used to deliver images in the message from an outside server. These images are used to detect that the message was opened by the intended recipient, confirming that this is a live e-mail address. The HTML image tag often contains a unique identifier so that the loading that particular URL can be correlated with a particular e-mail address.

3. The hypertext image link is often obscured using hexadecimal. For a tutorial on this, pc-help.org has done a fantastic page at http://www.pc-help.org/obscure.htm (PC-Help, 2002)

4. Spammers often disguise key words in the subject and the message body to avoid content filtering by the e-mail client. For example, S*X may be used in place of SEX in the subject line.

5. Spammers often put garbage words (such as dshfoiewqhr) in the subject of body so the message passes through some filters.

6. Spammers will hide words in the body of the message that are not commonly found in spam, such as "geostationary."

7. Spammers will forge the host name of where the message originated to obscure their identity. This forgery is revealed when the ip addresses in the header are checked against dns entries of host names.

Working my way through each user's spam, gave me an appreciation for the depth of the problem. Some users received several hundred spam per day and often mixed in with their submitted spam were legitimate mailing lists (~3-5 % of the spam). Because my company has disk space quotas on the amount of e-mail

a user can keep in their inbox, I had expected that the size of the spam message might be an issue and that spam might push a user over the quota. However, the average message size of my spam samples was only 3.6 kb (often the spam message downloads an image from a remote server, keeping the message compact). The major problem appeared to be the large volume of spam, rather than the size of the messages.

While I had good technical data on spam, I needed a simple definition for the surveys. Based on this, I focused on the aspects of the message that the user would normally see in an unsolicited mail. First, spam originates from outside the company. Second, spam is unsolicited and typically from a sender unknown to the user. Third, spam is not directly related to company business.

The Questionnaire Text

I wanted to maximize the number of individuals responding so that I could obtain the most accurate impact numbers possible. The questionnaire construction went through a number of drafts, and I made an effort to make the text so that it would take the individual as little time as possible to answer.

A number of people suggested that I use an automated method for gathering survey results. The company uses a survey tool in which a link is embedded in the e-mail and when the user clicks on the link, their browser opens to a web page with the survey. Because of the way the survey tool is implemented, if I wanted to follow-up with the user, I would have to have them type in their name. Also, I felt that users would be more likely to respond to the e-mail if they could hit reply, rather than follow a link to a web page[*].

The actual questionnaire text read:

> Dear <actual first name>,
>
> In collaboration with the E-Mail Server team, Research Division Information Security is researching methods for reducing the business impact of unsolicited e-mail, which is also known as SPAM. For accuracy and completeness, it is extremely important that you respond, even if you do not receive SPAM. *If you do not receive SPAM, please reply to this message by typing NO underneath question 1 and return the e-mail.* However, if you do receive SPAM, please also answer questions 2 and 3.
> Please return this e-mail by XX/XX/2003.
>
> For this survey, SPAM may be defined as e-mail that (1) originates outside of the Company, (2) was unsolicited, typically from a sender unknown to the recipient and (3) is not directly related to Company business. If you have comments on this definition, please write them in the comments section.
>
> 1. Do you receive SPAM (unsolicited e-mail at your Company e-mail address - please answer YES or NO) ?

[*] Other web based studies within the company had response rates close to half of the response rates observed in this study.

2. On average, how much SPAM do you receive in a single day ?
____ messages / day

3. Do you use Outlook rules or filters to block SPAM (please answer YES or NO) ?

If you have comments about SPAM or Outlook filters, please write them here:

Thank you in advance for your time !
Sincerely,
Joe McComb
Research Division Information Security

The final draft of the questionnaire was a simple text message that was designed to be rapidly answered by the user whether they received spam or not. Test users were sent the questions and timed for how long it took them to answer. The users answered the questions in an average of 30 seconds, but few of the test users supplied comments. Because the company had asked e-mail users to filter spam with Outlook filters, the third question was posed to find out how many users were actually filtering on the client end.

The questionnaire was piloted within my department (approximately 50 users) to test usability. In particular, I was interested if the second question had the proper scale of messages/day, or if a scale of messages/week should be used. The pilot revealed that messages/day would be adequate. Some employees who did not receive spam noted the italicized line "*If you do not receive SPAM, please reply to this message by typing NO underneath question 1 and return the e-mail.*" made the survey "easy" to fill out.

Employee Sampling Methodology

The sample consisted of 500 randomly selected research division employees and represented 5% of the total of division (10,000 total employees). Questionnaires were distributed in batches of 100 e-mails for the months of January – May, 2003. The surveys were typically targeted for release during the first week of the month and were typically sent on a Monday or Tuesday (the one exception was survey #4 which was sent on a Friday). Recipients were given two weeks to respond to the survey. The survey was typically released between the hours of 1:00 PM to 4:00 PM so that employees had had a chance to sort out their spam and other morning e-mail.

Results of the Surveys

Of the 500 e-mail questionnaires distributed, 393 people responded (79% response rate). The overall response rate was skewed by survey #4, which was sent on a Friday and had a 65% response rate. It was noted that the recipients who responded to survey #4 were more likely to receive spam than in surveys #1-3. In surveys #1-3, an average of 45 % of the sample reported receiving spam

but in survey #4, 54% of the sample reported receiving spam. This suggests that the employees who had not responded to the questionnaire may be individuals who do not receive spam and thus ignored the questionnaire.

Based on the assumption that the 21% of people who did not respond to the questionnaire, did not received spam, the results are presented in two forms. First, the raw results that are based upon the sample of the 393 people who responded to the survey is designated as Unweighted. Second, Weighted results are based upon a pool of 500 persons with an assumption that 107 recipients (21%) did not respond because they did not receive spam. The Weighted results have 107 people who receive no spam added to sample of 393 people who responded to the survey. The Weighted results provide a more conservative estimate of the volume of spam, and based on the results of survey #4 are probably more accurate.

Estimates of the Amount of Spam

Table 1. Percent of Yes and No responses to the question "Do you receive spam?" (Weighted assumes that people who did not respond to the questionnaire do not receive spam and these non-responders have been added to make a larger sample).

Do you receive spam ?	Unweighted	Weighted
Yes	48%	38%
No	52%	62%
Average Spam per Individual	5.67 spam per day	4.46 spam per day

The average amount of spam per day per employee is shown in Table 1 for both weighted and unweighted samples. At the very least, employees receive an average of 4.46 spam messages per day. This figure is misleading because spam is not distributed evenly between employees. The majority of users (85%) fall into a level where they receive no more than five spam per day and is shown in Figure 1.

Figure 1. Unweighted Distribution of Amount of Spam received among employees.

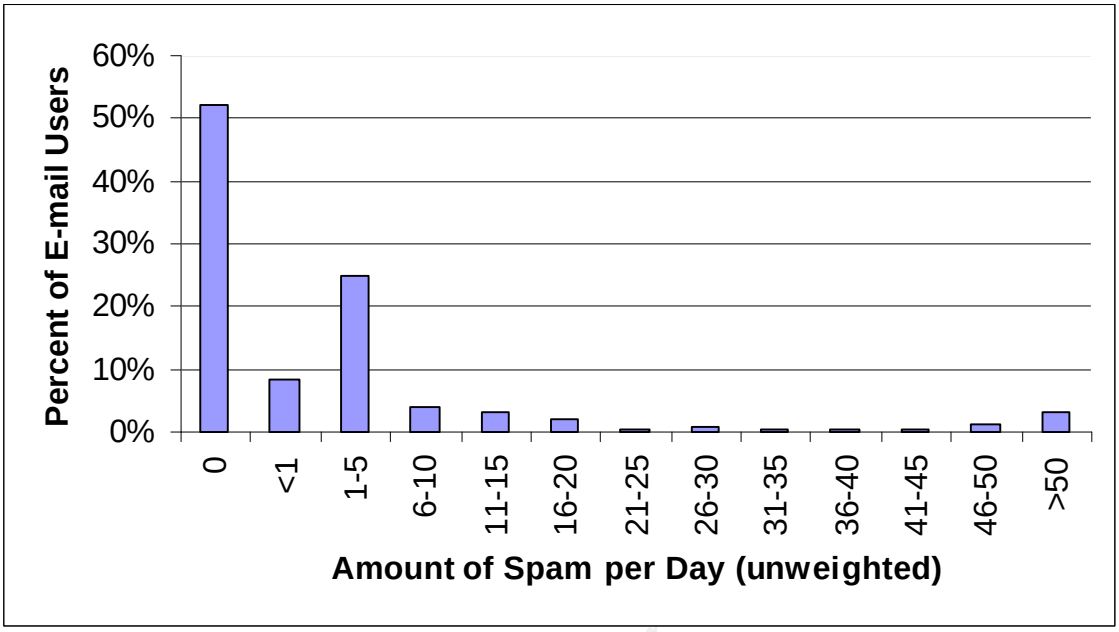
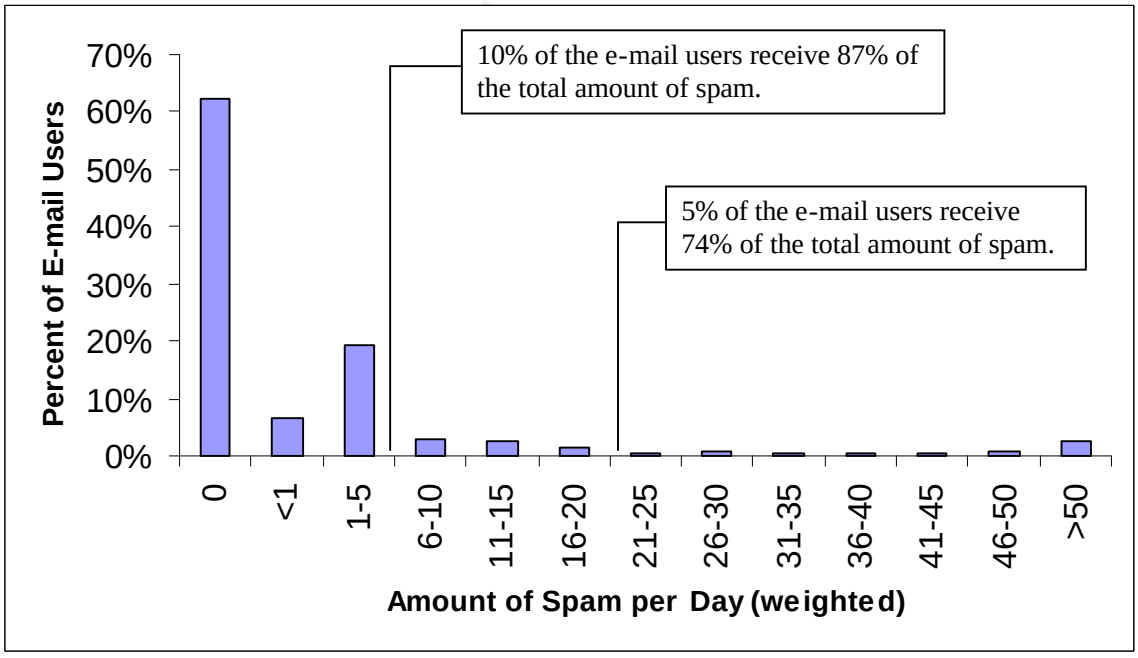


Figure 2. Weighted Distribution of Amount of Spam received among employees. The first marker indicates the 10% of employees who receive more than 5 spam per day account for 87% of the total volume of spam. The second marker indicates the 5% of employees who receive more than 20 spam per day account for 74% of the total volume of spam.



The graph of weighted data (Figure 2.) assumes that the individuals who did not respond to the questionnaire do not receive spam and thus skews the data towards 0 spam per day. The shift due to the weighting places 90% of the users

into the category of receiving 0 to 5 spam per day. As noted on Figure 2, this suggests that the majority of users are not drowning in spam, but rather a small portion of the total employees account for the majority of the spam. For example, 87% of the spam is received by 10% of the users (greater than 5 spam/day). In fact, 74% of the volume of spam is received by 5% of the users (greater than 20 spam/day).

Much of the variability in the amount of spam an employee receives appears to be related to how they utilize their business e-mail account. In following up with some of the users, I found that the users who got little to no spam usually used their e-mail address very conservatively on the internet. They typically did not sign up for mailing lists and did not post to newsgroups. In particular, this trend was present in the upper management, who typically used their e-mail accounts for mainly business purposes and in the questionnaires they reported little to no spam.

One question that was posed by a few users is “can my e-mail address be stolen by surfing the web?” Using Uri Raz’s site on how e-mail addresses are harvested (Raz, 2002), I tested the three primary methods of harvesting from a web browser and found that the Company’s standard browser did not give out an e-mail address to anonymous FTP site or in the HTTP_FROM header. I found that it was possible to automatically instruct the e-mail client to create a new e-mail message from javascript embedded in the web page, but that the e-mail user would need to click the send button to send the message. Overall, these harvesting methods did not pose a threat to the companies e-mail accounts.

Estimating the Percentage of Spam in Internet Mail

During the January – May study period, the e-mail server team set a default filter in the internet e-mail servers to log when an inbound message passed though and carried one of 30 key words which are commonly found in spam. The general expectation was that this list would only catch a portion of the spam. Readings of the total spam detected were taken on a weekly basis. During the study period, 43,514,654 external e-mail passed through the e-mail servers and 5,742,110 (13.22%) were detected as spam by the default key word filter.

Assuming 60,000 e-mail users, the volume of e-mail during the study period indicates that each patron receives an average of 6.48 external e-mail per day. If the amount of spam from the surveys is reflective of overall company, then 68.8% (4.46/6.48) of the inbound e-mail is spam. This figure is at the high end of the industry estimates for the amount of spam in inbound e-mail, which range from 20% – 70% (Lagesse, 2003). It is possible that the users overestimated the amount of spam they were receiving on the questionnaire. However, follow-up with some of the questionnaire recipients has shown that the estimates are accurate. In general, it is agreed within the spam reduction team that between 50 – 70% of the internet e-mail entering the company is spam. In summation, the

statistics showed that the default filter would only capture a small portion of the spam.

Results on Use of Filters

Overall, the majority of employees who are receiving spam were unaware that Outlook provides junk mail filters (the filters are under the actions menu) and a common comment to the question regarding filtering was:

“I don't know what filters are.“

Only 14% of the users who receive spam reported using filters, but many commented that they were dissatisfied with the filters because spammers easily evade the integral e-mail client filters. One user wrote:

“Blocking SPAM at the desktop is really unproductive. The tools don't work well because they are slow and too easily "spoofed" It would be much better to block it at the server, either by using a SPAM recognition service, or by limiting the number of identical messages that can be sent to company employees from outside. Also, by blocking SPAM at the server, you would cut the total e-mail load on company servers at least in half (my guess is even more)”

Estimating the Cost of Spam to the Research Division

As noted earlier, the major cost of spam is productivity loss due to the e-mail user having to sort and delete spam messages from their inbox. While spam does absorb disk space, this is minimized because most users delete the spam out of their inbox. A sample of 30 spam was analyzed and the typical message size is 3.6 kb. With 10,000 users in the research division receiving 4.46 spam per day, this results in:

10,000 users x 4.46 spam per day per user x 3.6 kb /spam = 156.8 MB per day

A major cost to the research division is the productivity loss due to spam. This was estimated by timing an individual as he sorted through 315 messages which he had received that day which were a mixture of spam and legitimate e-mail. To sort and delete took and average of 3 seconds per spam. For the user group studied in the research division, most individuals rapidly distinguished spam and deleted it. This lead to a general productivity loss of:

(4.46 spam / user / day) x (10,000 users) = 44,600 spam per day.
(44,600 spam / day) x (3 seconds / spam) = 133,800 seconds lost per day.
(133,800 seconds / day) / (3600 seconds /hour) = 37.2 hours lost to spam per day.

The company has a general estimate of the productivity dollars for the average worker at $75.00 per hour (including benefits and other expenses).

(37.2 hours lost to spam / day) x ($ 75.00 / hour) = $2,790.00 per day lost to spam.

Spam is sent 365 days a year. Discussion with survey participants revealed that they spent more time after a weekend sorting spam than during the rest of the week. In addition, comments in the questionnaires indicate that the spam / day estimate is based on a full week rather than a work week. Consequently, the loss estimate is based off of a 365 day year.

($2,790.00 per day) x (365 days /year) = $ 1,018,350 productivity lost per year.

The Gateway Filtering Solution

Shortly before the final survey in May, a formal spam reduction task force was formed and consisted of the e-mail server team, the e-mail client owner and research division information security. The statistics obtained from the questionnaires were presented to the task force and to the infrastructure management. The surveys were repeated and statistic data was gathered at the company level. From this data, the infrastructure leadership committee decided that the company had a problem with unsolicited commercial e-mail.

The task force chose to attack the problem in two areas. First, web sites were created to educate the users on how to avoid receiving spam (see Raz, 2002 for tips). Second, the team began to examine products that would filter unsolicited e-mail at the internet gateway. The earlier presentation by the e-mail server team that had been rejected by upper management showed that the company was risk adverse to losing legitimate business e-mail. The team recommended the solution of tagging the message with the filter and forwarding the message to the client. This solution would allow the business to see the efficiency of the filter without losing messages. The solution of providing disposable e-mail accounts was seen as difficult to maintain in the environment and too costly to implement.

Ultimately, a filtering service was selected from the manufacturer who had supplied the software for the internet e-mail gateway server[*]. This solution was seen to be as effective as the other providers examined (Brightmail and Ironmail) but was cheaper because it was an add-on to the existing e-mail server software.

Using the filter, messages were tagged according to the probability that they were spam. A tag was inserted into the message header and also in the subject line which read [Moderate Spam] or [High Spam] according to the probability that

[*] There was discussion among company reviewers if the actual filter that was purchased should be named in this section. One area of the concern was around endorsement of the product and another area was that it reveals something about the infrastructure. Based on this feedback, I removed the reference.

the message was spam. Pornographic spam was further segregated by tagging the spam as either [Moderate Adult Spam] or [High Adult Spam]. Users were instructed to use their e-mail client filters to direct messages with these tags to a junk e-mail folder.

The filter detected spam according to content of the e-mail message and would assign a probability based on the key words within the message body. The filter was updated twice daily as a service provided by the filter supplier. They created the updates by catching spam in honeypot e-mail accounts on the internet. In addition, the filter could be adapted by user input. Users could submit spam that had been missed by the filter to a mailbox so that the filter could be adapted to catch those messages in later rounds. In addition, users could submit messages that had been tagged but were not spam (called false positives) so that the filter could be adapted to let these messages through.

After – Assessing the Impact of Gateway Filtering

Piloting the Filter

The new filter was piloted for a month-long period during July 2003 on a test base of 40 users who were chosen from the upper management IT leadership committees. Some of the individuals received spam and others received little to none. The upper management was specifically chosen because it was felt that before the filter could be established in the company, upper IT management should directly experience how well the product worked.

During the pilot, the amount of spam correctly tagged was measured via surveys. A user was timed to determine how long it took to sort the spam from legitimate e-mail that had been caught in the junk mail folder. The pilot indicated that just over 90% of the spam was being tagged correctly and the users were highly satisfied with the filter. The user sorted through 446 caught messages in five minutes (~1.5 spam/second) which is a 450% improvement over the 1 spam/3 seconds that had been observed earlier. These results suggest that 2.33 seconds would be reclaimed for each spam. For the research division this amounts to:

(44,600 spam / day) x (2.33 seconds / spam) = ~ 103,900 seconds saved per day
(103,900 seconds / day) x (1 hour / 3600 seconds) = 28.86 hours saved per day.
(28.86 hours / day) x (365 days / year) = ~10, 500 hours saved per year.

The Divisional Survey

Based on the pilot, the filter was put into company-wide production in October 2003. Prior to implementation, the task force released a message to all e-mail users, explaining the details of the filter. One month after implementation, I

followed up with e-mail users who had responded to the first questionnaire who had reported that they received spam. The sample consisted of 75 users from the group who received 90% of the spam.

Similar to the first questionnaire, the follow-up went though a series of drafts and was designed to be answered by the employee in as short a time as possible. The questionnaire was distributed via e-mail just after 1:00 PM and the survey text read:

Dear <actual user name>,
I am following up on a series of surveys that I sent earlier this year to determine the impact of unsolicited e-mail (SPAM) on the Research Division. Over this period, a SPAM reduction team was created and in October, a Company-wide filter was put in place to identify and tag unsolicited e-mail. I am writing to Research Division employees who were included in the first impact study to gather feedback on the effectiveness of the current filter.
Can you please return this e-mail to me by XX/XX/2003?

1. Of the body of unsolicited e-mail messages you receive, approximately what percentage is correctly tagged as SPAM by the filter?
_____ %

2. How many messages has the filter tagged as SPAM, but were legitimate business e-mail?
_____ messages

If you have comments about SPAM or filtering, please write them here:

Thank you in advance for your time !
Sincerely,
Joe McComb
Research Division Information Security

Results of the Follow-up Questionnaire

The questionnaire had a 72% response rate and general comments indicated that the employees were “very happy that this new system is in place.” One interesting result was that people who reported that they received 5 messages or less per day reported that the filter was only 70 % effective. Conversely, people who received more that 5 messages per day reported a 96% tagging accuracy. Over the entire pool of individuals, 91.8% of the total amount of spam was correctly tagged as spam.

The difference between individuals who received 5 unsolicited messages or less versus more than 5 per day is suspected to be a difference in how individuals define spam, but may also be attributed to that some employees did not even notice the tags in the subject line. Both of these explanations are supported by follow-up conversations with employees and questionnaire comments. Some employees in the 5 or less per day category called and asked about the definition of spam e-mail, which may indicate that the individuals in the 5 or less per day category may consider legitimate junk e-mail (such as from legitimate outside

companies) as spam. On the other hand, several individuals commented that they either had not noticed the tags or had thought that new legislation had been enacted to force spammers to provide tags in the subject.

The false positive rate (legitimate e-mail tagged as spam) was found to be 0.06%, which is a remarkably low value. Based on this value the spam reduction task force is examining deleting the e-mail that has been tagged as a high probability of being spam at the internet gateway. Ultimately, this would prevent the high probability spam from reaching the employee.

Estimating Productivity Gain

From the earlier estimates of productivity losses to the research division, a cost model can be constructed for the company:

Before the filter

(4.46 spam per user per day) x (60,000 users) = 267,600 spam per day.
(267,600 spam per day) x (3 seconds per spam) = 802,800 seconds lost per day.
(802,800 seconds per day) / (3600 seconds per hour) = 223 hours lost to spam per day.
(223 hours lost to spam / day) x ($ 75.00 / hour) = $16,725.00 per day lost to spam.
($16,725.00 per day) x (365 days /year) = ~$6,104,000 productivity lost per year.

After the filter

Based on the employee testing values, a user clears 1.5 spam/second.

(4.46 spam per user per day) x (60,000 users) = 267,600 spam per day.
(267,600 spam per day) x (0.67 seconds per spam) = ~180,000 seconds lost per day.
(180,000 seconds per day) / (3600 seconds per hour) = 50 hours lost to spam per day.
(50 hours lost to spam / day) x ($ 75.00 / hour) = $3,750.00 per day lost to spam.
($3,750.00 per day) x (365 days /year) = ~$1,369,000 productivity lost per year.

Gain in Productivity:

```
  $6,104,000 productivity lost per year.  (before)
- $1,369,000 productivity lost per year.  (now)
---------------------------------------------
  $4,735,000 return on investment per year due to productivity gains.
```

Conclusion

After the filtering service was put into place, the task force has begun to examine the false positive rates and re-examine if the company will accept having the high probability spam e-mail messages deleted at the internet gateway. The work is ongoing, but the employees have provided positive feedback on questionnaires that they are willing to lose some legitimate e-mail if it means that they receive less spam.

A recent article in InfoWorld (Roberts, 2003) presents research by Nucleus Research with comments from other researchers in the spam reduction area. InfoWorld reports that the Nucleus Research found that the average employee receives 13.3 spam per day and spends on average 6.5 minutes per day deleting spam (Nucleus Research, 2003). Nucleus Research(2003) reports that on average companies lose $874.00 per employee per year to spam in lost productivity. Within the present case study, this would translate to 60,000 employees x $874.00 =$ 52,440,000.00 in lost productivity per year. Also in the InfoWorld article, Joe Fisher, a researcher from Tumbleweed, comments that the costs and time in the Nucleus Research report may be too low.

The results found in this case study indicate that the average employee received 4.46 spam per day and spends approximately 15 seconds each day sorting and deleting the spam. However, the three second per spam value rests on the direct observation and timing of a single employee. For the analysis requested by the upper management, these figures were considered valid enough to make a business decision. Clearly, however, more direct observation of a larger set of employees would be needed to accurately estimate the productivity loss due to spam.

In addition, this case study suggests that the definition of what constitutes spam differs from individuals who receive little spam versus those who receive a great deal. Employees who report that they receive 5 or less spam per day appear to consider legitimate mailing lists as part of the spam, but this result would need to be confirmed with follow-up studies of the samples of e-mail which the employees consider to be spam.

In summation, this case study provided an analysis that showed that the enterprise was potentially losing several million dollars in productivity to unsolicited commercial e-mail. The results of the study were used in the business decision to purchase a commercial filtering service. As shown through employee questionnaires and direct user observation, the filtering service does appear to decrease the productivity loss due to spam. However, because of the small sample of observed users, this sample would need to be increased to confirm the productivity loss values calculated in this case study.

Bibliography

DiSabatino, Jennifer. “Spam taking a toll on business systems.” ComputerWorld. 2002.
URL:http://www.computerworld.com/networkingtopics/networking/story/0,10801,68439,00.html

Gaspar, Suzanne. “Fighting back against spam.” NetworkWorldFusion. 2002.
URL:http://www.nwfusion.com/research/2002/0513spam.html

Lagesse, David. “That Damn Spam.” USNews.com. 2003.
URL: http://www.usnews.com/usnews/biztech/articles/030428/28spam.htm

Lucent Technologies.“The Lucent Personalized Web Assistant.” 1998
URL: http://www.bell-labs.com/project/lpwa/

Osterman Research. “Osterman Research Survey on Email Content Filtering Issues.” 2002.
URL:http://www.ostermanresearch.com/results/surveyresults_cf0802.htm

PC-Help. “How to Obscure Any URL.” 2002.
URL: http://www.pc-help.org/obscure.htm

Raz, Uri. “How do spammers harvest email addresses ?” 2002.
URL: http://www.private.org.il/harvest.html

Roberts, Paul. “Report: Spam costs $874 per employee per year.” InfoWorld. 2003. URL: http://www.infoworld.com/article/03/07/01/HNspamcost_1.html

Nucleus Research. “Spam: The Silent ROI Killer.” Research Note D59. 2003.
URL: http://www.nucleusresearch.com/prspam.html

Snow, Richard. “Spam Filtering in a Small Business Environment, a Case Study.” SANS Reading Room. 2003. URL:http://www.sans.org/rr/catindex.php?cat_id=19

Acknowledgements

The author would like to acknowledge the work of Carl Weilandics for providing internet gateway spam filtering statistics from a default filter during the December – May, 2003 study period.

Last Updated: June 24th, 2017

Upcoming Training

SANS Paris 2017	Paris, France	Jun 26, 2017 - Jul 01, 2017	Live Event
SANS Cyber Defence Canberra 2017	Canberra, Australia	Jun 26, 2017 - Jul 08, 2017	Live Event
SANS Columbia, MD 2017	Columbia, MD	Jun 26, 2017 - Jul 01, 2017	Live Event
SANS London July 2017	London, United Kingdom	Jul 03, 2017 - Jul 08, 2017	Live Event
Cyber Defence Japan 2017	Tokyo, Japan	Jul 05, 2017 - Jul 15, 2017	Live Event
SANS Cyber Defence Singapore 2017	Singapore, Singapore	Jul 10, 2017 - Jul 15, 2017	Live Event
Community SANS Phoenix SEC401	Phoenix, AZ	Jul 10, 2017 - Jul 15, 2017	Community SANS
SANS Los Angeles - Long Beach 2017	Long Beach, CA	Jul 10, 2017 - Jul 15, 2017	Live Event
SANS Munich Summer 2017	Munich, Germany	Jul 10, 2017 - Jul 15, 2017	Live Event
Mentor Session - SEC401	Macon, GA	Jul 12, 2017 - Aug 23, 2017	Mentor
Mentor Session - SEC401	Ventura, CA	Jul 12, 2017 - Sep 13, 2017	Mentor
Community SANS Atlanta SEC401	Atlanta, GA	Jul 17, 2017 - Jul 22, 2017	Community SANS
SANSFIRE 2017	Washington, DC	Jul 22, 2017 - Jul 29, 2017	Live Event
SANSFIRE 2017 - SEC401: Security Essentials Bootcamp Style	Washington, DC	Jul 24, 2017 - Jul 29, 2017	vLive
Community SANS Fort Lauderdale SEC401	Fort Lauderdale, FL	Jul 31, 2017 - Aug 05, 2017	Community SANS
SANS San Antonio 2017	San Antonio, TX	Aug 06, 2017 - Aug 11, 2017	Live Event
SANS Boston 2017	Boston, MA	Aug 07, 2017 - Aug 12, 2017	Live Event
SANS Prague 2017	Prague, Czech Republic	Aug 07, 2017 - Aug 12, 2017	Live Event
Community SANS Omaha SEC401*	Omaha, NE	Aug 14, 2017 - Aug 19, 2017	Community SANS
SANS Salt Lake City 2017	Salt Lake City, UT	Aug 14, 2017 - Aug 19, 2017	Live Event
SANS New York City 2017	New York City, NY	Aug 14, 2017 - Aug 19, 2017	Live Event
SANS Chicago 2017	Chicago, IL	Aug 21, 2017 - Aug 26, 2017	Live Event
Community SANS Trenton SEC401	Trenton, NJ	Aug 21, 2017 - Aug 26, 2017	Community SANS
SANS Virginia Beach 2017	Virginia Beach, VA	Aug 21, 2017 - Sep 01, 2017	Live Event
Community SANS San Diego SEC401	San Diego, CA	Aug 21, 2017 - Aug 26, 2017	Community SANS
SANS Adelaide 2017	Adelaide, Australia	Aug 21, 2017 - Aug 26, 2017	Live Event
Virginia Beach 2017 - SEC401: Security Essentials Bootcamp Style	Virginia Beach, VA	Aug 21, 2017 - Aug 26, 2017	vLive
Mentor Session - SEC401	Minneapolis, MN	Aug 29, 2017 - Oct 10, 2017	Mentor
SANS Tampa - Clearwater 2017	Clearwater, FL	Sep 05, 2017 - Sep 10, 2017	Live Event
SANS San Francisco Fall 2017	San Francisco, CA	Sep 05, 2017 - Sep 10, 2017	Live Event
Mentor Session - SEC401	Edmonton, AB	Sep 06, 2017 - Oct 18, 2017	Mentor