Factors involved in estimating cost of Email spam

Farida Ridzuan, Vidyasagar Potdar, Alex Talevski

Anti Spam Research Lab, Digital Ecosystems and Business Intelligence Institute,
Curtin University of Technology.
farida.mohdridzuan@postgrad.curtin.edu.au, {v.potdar, a.talevski}@curtin.edu.au

Abstract. This paper analyses existing research work to identify all possible factors involved in estimating cost of spam. Main motivation of this paper is to provide unbiased spam costs estimation. For that, we first study the email spam lifecycle and identify all possible stakeholders. We then categorise cost and study the impact on each stakeholder. This initial study will form the backbone of the real time spam cost calculating engine that we are developing for Australia.

Keywords: spam cost, email spam, spam lifecycle

1 Introduction

Spamming in email refers to sending unwanted, irrelevant, inappropriate and unsolicited email messages to a large number of recipients. Sending email is fast, convenient and cheap; making it as an important means of communication in business and personal. This is supported by the report from Radicati Group saying that there is a growth of email users from time to time [1]. Dependencies on email usage throughout the whole world provide a huge opportunity to the spammers for spamming.

Spamming activities starts from spammers (who create and send spam), but its impacts goes far beyond them, involving Internet Service Provider (ISP), company, and users (spam email recipients) since they represent the key stakeholders. It is undeniable that each stakeholders involved in this activity has to bear some costs associated with spam.

Throughout our study, there are a few papers discussing on the costs of email spam, but most of them focuses only on one stakeholder, which is the user. Not only that, most of the results are from commercial anti spam vendor. So, it is unclear on how unbiased these reports are. For instance, [2] estimated that company with an average number of employees of 12,000 has to bear the cost of $2.4 million but by deploying the anti spam solution, they would be able to save $1.2 million. Nucleus Research estimated that the loss of productivity for spam management in US is more than $71 billion annually [3]. It is also estimated in [4] that deploying Spamhaus in large corporation and mid-sized corporation could save $400,000 and $27,000 respectively.

Therefore, the main aims of this paper are to 1) identify spam stakeholders, 2) understand email spam lifecycle, 3) identify cost categories and parameters for each

cost categories, and 4) derive the cost impacts based on the identified cost categories and related parameters towards each stakeholder.

This paper has been organized in the following way. Section 2.0 will enlist three main stakeholders in email spam lifecycle i.e. spammers, ISP and users. Section 3.0 will give a brief overview of the lifecycle of email spam. Details on which party involve in every stage, tools used by the spammers are also included in this section. Section 3.0 continues by introducing 5 cost categories and its related 17 parameters that can be used to estimate the cost of email spam. Section 4.0 begins by laying out the email spam costs in detail towards three different stakeholders: spammers, users and ISPs. The last section provides the discussion and conclusion.

2 Spam Stakeholders

In this section, we list all the key stakeholders in the email spam lifecycle. These include

=> Spammer
=> Internet Service Provider (ISP), and
=> User

We now explain the details of each stakeholder in the following sections. We specifically outline where each stakeholder plays a key role.

2.1 Spammer

Spammer starts the lifecycle of email spam by creating spam messages and sends them to the users. Spammer uses various techniques and tries to bypass filters deployed by other stakeholders. Spamming gives a few benefits to the spammers such as to generate revenue, get higher search rank, promote products and services and others [5], which motivates them to continue spamming even with the existence of spam laws such as the CAN-SPAM Act [6].

2.2 Internet Service Provider (ISP)

Internet Service Provider (ISP) provides internet access both to users and spammers. They are involved in the lifecycle of email spam because without their service, spammers would not be able to send emails in bulk. On the other hand, emails users would not be able to read their legitimate email without ISP. Due to the spamming activities by spammers, ISP would have to prepare large bandwidth for their users, which indirectly increases ISP's operational costs.

2.3 User

The third stakeholder in the lifecycle of email spam is user. User is the actual recipient of an email spam. Spammer's goal is that the email should reach the end

user and ensure that user is interested in opening, reading and responding to the email. Apart from the filtering system, the end user is the key to decide whether the spammer's campaign would be successful or not. This is because, in the end, if there is any spam email that gets to the inbox; the user can choose either to respond or ignore that email. The lifecycle ends when the user ignores the spam email but it continues if the user replies or takes action based on that email. Users depend on efficiency of the anti spam software to avoid getting spam emails. Most of the problems faced by the users are quite similar. They are afraid of losing legitimate emails filtered by anti spam filter but at the same time, they do not like to spend too much time checking spam emails. The user may even waste more time if s/he gets interested in a spam email spam because they will spend more time browsing unnecessary websites.

3 Lifecycle of Email Spam

The three main stakeholders in email spam lifecycle are users, spammers and ISPs. Based on these stakeholders we have classified email lifecycle into seven main stages. Email spam lifecycle starts from spammer's end and then continue to traverse to the recipient's end. We categorise email lifecycle into the following seven stages as follow:

=> Get email addresses
=> Create spam messages
=> Send spam
=> Filter spam by the ISP
=> Filter spam on server side
=> Filter spam on client side
=> Spam that bypasses all filters [7, 8].

As mentioned earlier, this section will focus more on the lifecycle of email spam, tools used by spammers, problems encountered by spammers in sending the spam and what has been done by anti spammers as countermeasure in each stage. The first three stages involve spammers while the next four stages involves with the recipients of spam email. These seven stages are described in the figure below with three stakeholders involved: users, spammers and ISPs.

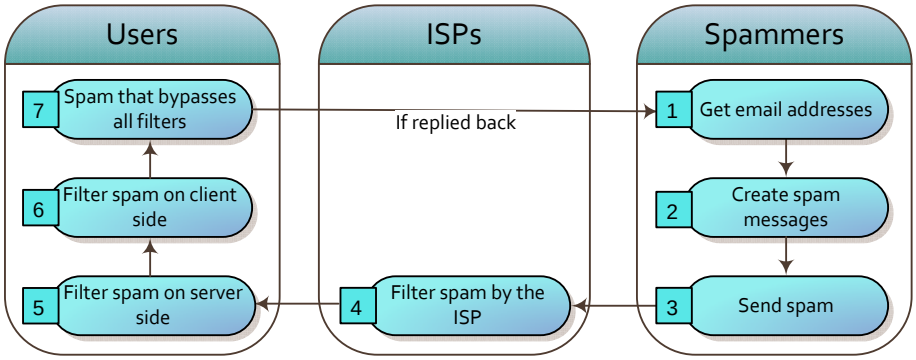


Fig. 1. Lifecycle of Email Spam.

3.1 Get Email Addresses

The lifecycle of an email spam starts with the spammers gathering a list of email address. These email addresses can be gathered through several well known ways such as obtaining email list & newsletter subscriber list through hacking or buying it, using spam bots to crawl and collect email addresses from websites, randomly generating name combinations for certain domain and others. Several other ways of spammers getting email addresses are explained in [9]. In fact, in a recent case, spammers were successfully in getting a huge list of email addresses with its passwords, that significantly increased the rate of spam [10] since they are able to manipulate the account itself and use it in order to get other active email addresses.

One of the easiest ways for spammers to gather email addresses is to use automated tools e.g. Speed Email Extractor, Power Email Harvester, Email Grabber, Teleport Pro, Email Spider or Email Extractor from EmailSmartz and others [11-15]. Trial and limited version are downloadable for free. Some of this software such as Power Email Harvester [12] and Email Extractor [15] could even be used to send bulk mail.

There are plenty of ways for spammers to gather a list of email addresses. Nevertheless, it would be wasteful if the list that they have gathered contains fake email addresses. Spammers need to test the list of email addresses by sending an email to detect whether it's fake or not. If it is, spammer usually gets an auto reply message saying that the address is not valid. Still, this does not cost much for spammers to stop spamming. In order to maintain a genuine email database, there is no other way than to keep updating the database and getting a new list.

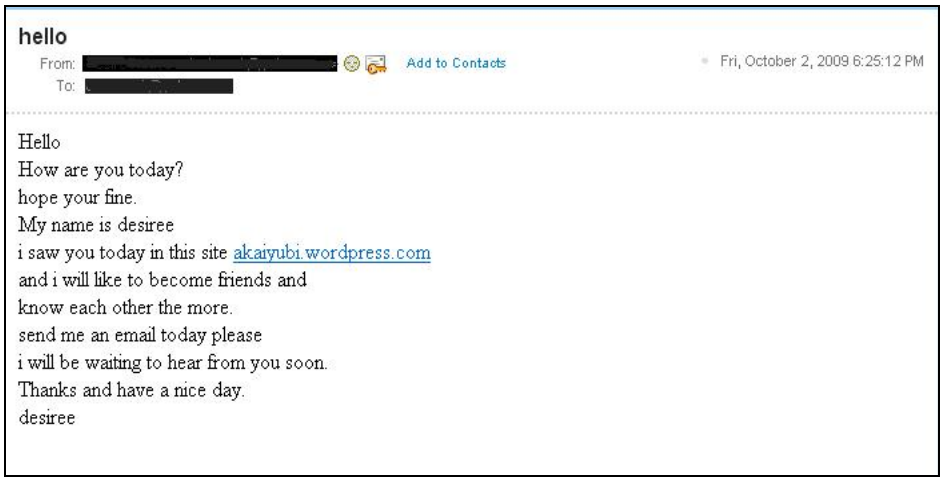
Preventing spammers to successfully gather email addresses, companies need to deploy security measures on network server to prevent spammers from hacking the server. In case where spammers use crawlers to obtain email addresses from websites, web administrator could use tools like Spam Preventer 1.0 [16] to avoid websites from being harvested by spam bots. Nowadays, users also have been educated not to

simply put their email address on the websites and use address munging technique such as by replacing “[dot]” instead of “[.]”.

3.2 Create Spam Messages

Stage 2 shows that spammers need to create messages before sending the email. It is possible to just send a simple message to thousands of users but the rate of success for such emails being read is low. Research is needed to provide users to read what they need [17] so that spammer could sends a specific email to their target group. We believe that spammers have their own database of words and phrases to create messages and this system would allow them to create messages and send huge amount of spam messages faster. This system is supposed to be robust against simple keyword filter.

Common users would have to follow certain tips to avoid from being mistakenly seen as spammers, such as given in [18-20]. Similarly, spammers could also follow these tips in order from being detected. Hence, spammers will try to imitate real user’s behavior. Their target is to get users interested in reading the messages and make an active action either by ensuring that the user clicks on the link provided or reply to the email [17]. Nowadays, spammers are smart enough to avoid using common keywords since they are easily detectable. Example below shows that spammers are now trying to act as if the spammer knows the user personally.


hello

From: [redacted] Add to Contacts Fri, October 2, 2009 6:25:12 PM

To: [redacted]

Hello
How are you today?
hope your fine.
My name is desiree
i saw you today in this site akaiyubi.wordpress.com
and i will like to become friends and
know each other the more.
send me an email today please
i will be waiting to hear from you soon.
Thanks and have a nice day.
desiree

Fig. 2. An example of smart spam message.

Spammers would try to avoid using sensitive blacklisted keyword and exploit the spelling of those words to bypass filters. Spammer also uses personal words such as “Hey” or “Hello” such as in Figure 2 as their message title to encourage users to open the email. This makes it difficult for anti spam filters to differentiate between spam and ham (i.e. legitimate) messages. Nevertheless, there is no other way for anti spam filters that uses content based method to keep updating their keyword databases.

Hence, using behavior analysis along with content based analysis might provide a better solution [21].

3.3 Send Spam

Spammers then will send the created messages to thousands of recipients. There are several ways for spammers to do this. The most important thing for spammers is to ensure that their identity remain unknown when sending spam. Spammers might use various different free email services provided by Hotmail, Yahoo! or Google so that they cannot be detected by the recipient. The spammers might also use unsuspected third party mail servers and spam can be relayed through them indirectly hiding the spammer's identity. Spammer could also scan for open insecure proxy servers or botnets, which can be used to collect email addresses and then send an enormous amount of email simultaneously.

Spammers then create or use program or software that could be used to send spam to the list gathered earlier in a faster way. Some of the tools that can be used to send email in a huge volume or mail bombers are Bulk Mailer, Avalanche, Dark Mailer, PostCast Server [22-25]. Meanwhile, in order to detect insecure proxy, spammers could use YAPH (Yet Another Proxy Hunter) [26] or Proxy Hunter [27]. An easier way for spammers to send email is to use email sending company such as Aweber Communications, Constant Contact and JangoMail [28-30].

On the other side, if the spammers are using open proxy server, there might be other spammers that are spamming from similar proxy as well. This could make it obvious for others that this particular server is used to send spam hence shortening the life span of using that proxy from being banned. In this way, spammers then need to find a new open proxy. Spammers could also be sending spam email through email advertising companies, which are sending a numerous amount of email legally. The price that spammers have to pay for an email is also comparatively very low [31]. Nowadays, using botnets is one of the easiest and effective ways to send spam. Nevertheless, there are some botnets that have been detected such as Srizbi, Bobax and Rustock [32]. Still, spammers could just find other botnets that have not been detected and continue spamming.

In this case, the anti spam group just have to keep detecting active botnets that are sending huge amount of spam email and blacklist them. Companies should also keep updated list of banned hosts and flag suspicious IP address. To avoid company's server from being used by spammers, company needs to deploy anti relay functionalities on their servers. It is without a doubt that more active countermeasure needs to be taken in order to avoid spammers from keep spamming.

3.4 Filter Spam by the ISP

Once spammers send spam email, those spam messages will first be filtered by the ISPs. These anti spam filter tools usually implement Sender Reputation, Sender Authenticity, Content Analysis (though some of the service provide claimed that they are implementing a better solution than this method, but it is unclear of what method

they are using), Network Analysis and others. The pricing of these tools' depends on the number of mailboxes that the ISP wishes to protect, numbers of messages to be handled daily and type of institution. Some of the tools cost depends on the number of domains, or number of servers.

Tools that can be deployed by the ISP are MailCleaner, SpamFilter ISP from Logsat Software and SpamTitan [33-35]. Several other choices for the ISP are the Anti Spam for ISP by Kaspersky and MXForce [36, 37]. As of November 2009, ovh.net, telefonica.es and tiscali.it are some of the top three spam service ISP reported by Spamhaus [38]. Some other services to detect and blacklist ISP used for spamming are RBL(Real Time Blacklist) such as provided by [39] and [40].

ISP on the other hand is spending a huge amount of money for the anti spam filters [41, 42]. ISP also needs to control and manage this tool. At the same time, they do not want to lose their customers, which may also include spammers. It is also believed that there are ISPs that sell services to spammers in order to gain profit [38]. This is one of the reasons why spammers are unstoppable.

3.5 Filter Spam on Server Side

Company usually deploy anti spam filter on their server. Therefore, users would not see all the emails that were actually sent to them because most of these email messages have been filtered by the server. The cost of implementing anti spam filter on the server depends on several factors such as how many email can the tools cater at a time, methods and additional services provided by the anti spam filter, number of users, duration of license and others. These filters usually provide solution by using Bayesian filtering, keyword checking, email header analysis and DNS blacklist [17].

Some of the commercial tools that could be implemented on the server side include EMP 7 Enterprise Anti Spam, SpamFighter Exchange Module, Spamfighter Mail Gateway and Web & Mail Security GFIMailEssentials Anti-spam Solution for Exchange/SMTP/Lotus [43-46]. In addition, companies could also opt for open source tools such as SpamAssassin and Anti-Spam SMTP Proxy Server [47, 48].

Main problem for applying filter on server side are the costs of renewing the license every year and managing the tools itself especially for a small company. Company usually does not favor to commit for a longer duration license because the service provided might not satisfy the company and there might be better services provided by other anti spam companies in the future.

3.6 Filter Spam on Client Side

As an additional precaution, users themselves can install desktop application to filter spam. Tools that can be used by the users are BullGuard Spamfilter, Cloudmark Desktop, ClearMyMail, MailFrontier Desktop, Spam Filter Express and others [49-53]. These tools come with various features and different advantages and disadvantages. Some tools need to be updated frequently; some are licensed, which has to be paid every year and some need more time to train.

With this filter, it is much harder for the spam messages to get through to the users. But still, not everybody would want to spend money on additional filter. Some users are just satisfied with what is provided by the application. The logic comes when users use free email services, therefore, they would not want to spend more on additional filtering and just hope the free services that they chose provide the best anti spam technology. These are the groups that spammers usually choose to spam. Even for the users that implement this filter, spammers could just hope that they will still check the spam folder and response to the spam message.

As for users, the problem comes when receiving smart spam messages. Users without knowledge could easily fall for the trap especially with spammers imitating real friendly behavior. Still, the tool used to filter spam on client side usually comes with a high costs and need to be maintained personally by user.

3.7 Spam That Bypasses All Filters

Spam that is not caught by both filters then can be read by the users. If the users fall for the trap and reply that email, user's email address is confirmed to be active and will then go into spammers' email database. As a result, user then will receives more email spam. The only problem for spammers is that most of the spam messages sent by them are already filtered. Hence, they are targeting specific topic or product to specific person in order to get them to read the messages. As a countermeasure, knowledge of what spam is should be given to educate email users from being easily manipulated.

4 Cost Categories of Email Spam

We have identified 5 cost categories with 17 parameters to estimate the cost of email spam. This section will further provide description of each cost category with its associated parameters. Parameters in each cost category will be defined considering that we are collecting a huge collection reference of web spam and trying to estimate those cost based on the collection.

4.1 Storage Cost

Storage cost refers to the cost spent for server storage used to store any information such as list of email addresses, spam for spammers and blacklisted IP addresses for companies and ISPs. In the effort to avoid losing legitimate messages, it is easier to flag an email or spam content so that it could be checked by the user itself. Once the user checks it, the user either would read it and clear the messages as non-spam or delete it if it is a spam. This process requires additional storage and includes the cost of filtering because the efficiency of this method depends on the filter itself. Suppose the email or content itself contains big attachment files or large sized images, this will increase the storage requirement and its cost. Storage cost for email spam can then be defined as follow:

$$C_s = f(a, b, c, d). \quad (1)$$

where

a = monthly fee/ GB for storage,
b = spam message received/day,
c = message size,
d = duration of storage.

Monthly fee per GB for storage is actual server cost paid for each GB. Measurement of this cost will consider the current general cost of storage. Spam message received per day is defined as spam messages received by the user per day. Measurement of this unit will consider all spam emails received by all the users on a certain period.

Message size is then defined as size of spam email. Measurement of this unit will consider the actual size used to keep the spam email message in storage. Duration of storage is the parameter defined to calculate how long (days) a message is stored before it is checked and deleted. In order to measure this, there is a need of a close observation towards the spam email to measure when a spam email is checked and deleted since it was first received.

4.2 Bandwidth Cost

Bandwidth cost is the cost used for connectivity. In this case, all parties are going to bear the cost of i.e. spammers for spamming and users for checking emails. Bandwidth cost function for email spam can then be defined as follows:

$$C_b = f(e, f, g). \quad (2)$$

where

e = annual fee for connectivity,
f = email percentage representing bandwidth,
g = spam percentage of all email.

Annual fee for connectivity is the actual cost users have to pay for Internet connection. Measurement of this cost will consider the current cost that users have to pay for the connectivity. On the other hand, email percentage representing bandwidth is considered as the proportion of bandwidth used just for email purposes. Measurement of this cost would need to consider previous research done by [3, 8]. Spam percentage of all email is defined as the proportion of spam messages from all received emails. Based on data collection, the proportion of spam email messages can be measured based on the storage that it uses.

4.3 Human Resource Cost

Human resource cost for spam filter is the cost used by the associated party for filtering spam. Considering that not everyone has knowledge of spam, companies usually hire a professional team to handle any issues arising from spam. This could include help-desk support or network team specially hired for fighting spam. Spam sometimes cause serious problem if they are embedding virus or worms with the attachment [7]**.** Ignoring the cost associated with virus and worms, the cost for human resource can be defined as follow:

$$C_{hr} = f(h). \quad \textbf{(3)}$$

where

h = salary for human resource incharge to support spam.

Salary for human resource incharge to support spam is the amount of salary paid for person to manage spam-related problem. Measurement of this cost depends on the current salary usually paid to the network administrator, support team members or help-desk officer, which requires further survey on current situation in order to determine its precise and accurate value.

On the other hand, spammers are also using their time to create smart spam, find a list of genuine email addresses and send spam etc. This cost is associated from stage 1 to stage 3. In this case, we define human resource cost for spammers as follow:

$$C_{hr} = f(i). \quad \textbf{(4)}$$

where

i = time used for spammers to spam.

4.4 Annual Productivity Cost

Three different substance to take into account when calculating productivity cost are the 1) process of inspection and deletion of spam that gets through to the inbox, 2) process of identifying legitimate email from spam folder and 3) helpdesk support [54]. In our case, annual productivity cost is measured for the time that is spent on each spam message and the cost of helpdesk support is already calculated in human resource cost. This cost may vary depending on user's knowledge. Even if a spam message is flagged, a user might actually reply the email. This is because a message might be spam to one, but it might not be for others. Nevertheless, this cost can be defined using several parameters as follow:

$$C_{ap} = f(j, k, l, m, n, o, p). \quad \textbf{(5)}$$

where

j = time to clear out spam/each check,
k = time to look for false positive in spam folder/each check,
l = time to focus back on work after each check,
ml = employee salary,
n = how many times users check email/day
o = how many working days
p = number of employees in one organization

Time to clear out spam/each check is considered as the amount of time needed to delete any spam email. Measurement for this cost depends on how fast a user can interact with system which also depends on how familiar users are with the application.

Time to look for false positive in spam folder/each check is defined as the amount of time needed to check and determine if there is any a legitimate content that was mistakenly flagged as spam. Measurement for this cost may vary depends on how knowledgeable users are about spam. It is also possible to measure this based on author's experience.

Time used to focus back on work after each check is the parameter used to define the amount of time needed to an employee to focus back to work after each check. This cost is usually measured by taking an average value of user opinion. Measurement for this cost has not been decided yet but it is also possible to measure this based on author's experience.

Employee salary is the parameter that would consider the salary of an employee which varies depends on position hence having a different effect on the total amount of this cost. Measurement for this parameter need further survey on current situation.

How many times users check email/day is considered as the frequency of a user checking the application. It is also possible to use a predetermined default value for this parameter.

For parameter how many working days, it is possible to just use a predetermined value that is 22 days permonth considering that there are 30 days in every month.

Number of employees in one organization is the parameter could be measured through the number of account holders for email application. These parameters would play an important role in calculating the cost of software because some software are licensed and buying them depends on the number of employees in an organization.

4.5 Software Cost

There are a lot of software tools available for both email spammers and users. Considering that each party deploys these tools for their email application, it is important to measure this cost based on current survey of the cheapest and most effective spam tools for spammers and anti spam solutions for users. This cost can be defined as follow:

$$C_{sw} = f(q). \tag{6}$$

where

q= software costs.

Table 1 shows that parameters used in cost calculation for each party i.e. spammer, company and ISP. Further explanation on each cost calculation for all the parties will be provided in the next section.

Table 1. Parameter used for spammer, company and ISP.

	Parameter	Abb.	Spammer	Company	ISP
	Storage Cost				
1	Monthly fee/GB for storage	*a*	✓	✓	✓
2	Spam message received/day	*b*		✓	✓
3	Message size	*c*	✓	✓	✓
4	Duration of storage	*d*	✓	✓	✓
	Bandwidth Cost				
5	Annual fee for connectivity	*e*	✓	✓	✓
6	Email percentage representing bandwidth	*f*		✓	✓
7	Spam percentage of all email	*g*		✓	✓
	Human Resource Cost				
8	Salary for human resource in charge to support spam	*h*		✓	✓
9	Time used for spammers to spam	*i*	✓		
	Annual Productivity Cost				
10	Time to clear out spam/each check	*j*		✓	
11	Time to look for false positive in spam folder/each check	*k*		✓	
12	Time to focus back on work after each check	*l*		✓	
13	Employee salary	*m*		✓	
14	How many times users check email/day	*n*		✓	
15	How many working days	*o*		✓	
16	Number of employees in one organization	*p*		✓	
	Software Cost				
17	Software costs	*q*	✓	✓	✓

5 Spam Cost Impact Towards Stakeholders

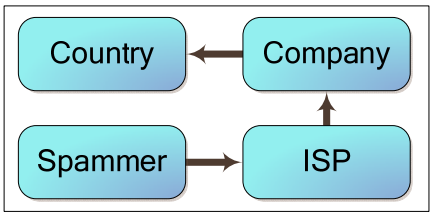


Fig. 3. Email Spam Cost Impacts.

Figure above shows the cost impact of email spam towards four parties: spammer, company, ISP and country. Based on our understanding, spammers relatively bear the lowest cost followed by company, ISP and country. However, in this section, we are only going to focus the cost impact of spam towards three stakeholders that we have defined in Section 2. For this section, we are going to address company as a stakeholder since company is considered as a group of users. Based on the generic parameters set in previous section, each cost associated for spammer, company and ISP are going to be identified.

5.1 Cost of Email Spam for Spammer

Based on the lifecycle that we have mentioned earlier, storage cost is associated with spammers in stage 1 and 2 and with users in between stage 4 and 5. Hence, parameter set for storage cost and bandwidth cost used by spammer are as follow:

$$C_s = f(a, c, d). \tag{7}$$

$$C_b = f(e). \tag{8}$$

With free web crawler, spammers can use UbiCrawler, WebSphinx, Wget, Polybot and Teleport to gather a list of email addresses [14, 55, 56]. In this stage, spammer could use email grabber such as Speed Email Extractor, Power Email Harvester, Email Grabber, Teleport Pro, Email Spider or Email Extractor from EmailSmartz and others for free [11-15]. Usually, the latest version of the software which provides more functionality would cost the spammers in the range of AUD19($16.95) to AUD165($149.95). Nevertheless, storage cost in stage 1 for spammers is relatively low as they can just use normal capacity to store a list of billions email addresses. Thus, the only cost that spammers have to spend is their time as the software that they bought is a one-time cost. Time uses for spammers to spam is defined as in Equation 4.

Some of the tools that can be used to send email in a huge volume or mail bombers are Bulk Mailer, Avalanche, Dark Mailer, Direct Mailer which costs spammer in the

range of AUD45($40) to AUD550($499). YAPH(Yet Another Proxy Hunter) which is an open source or Proxy Hunter for AUD33 ($29.95) are tools that can be used to detect insecure proxy servers. Spammers that generate high revenue could also opt by sending email using companies such as Aweber Communications, Constant Contact and JangoMail [28-30]. These companies provide services which would cost spammers as low as AUD0.0044 per email recipient

5.2 Cost of Email Spam for Company

Regardless of how users or companies deal with spam, there are costs that they have to bear and these costs are far beyond than financial costs. For example, according to Nucleus Research, there are three ways used by a company to deal with email spam which are: confirmation process, quarantine strategy and delete strategy [3]. Each strategy possesses different risk thus causes different costs such as loss of productivity, loss of time and storage cost.

Storage cost is associated between stage 5 and 6 for company. Storage cost parameters for company can be defined similarly as in previous section. Using the quarantine strategy, all email marked as spam needs to be kept in storage until the users check and delete it. This process is implemented in order to avoid losing important business emails.

Bandwidth is a valuable resource for company. Now that spammers are getting smarter and more creative by embedding various attachment types, downloading all these unnecessary spam messages consumes large bandwidth as well. Bandwidth cost is associated during the transmission of email spam from spammer's side to user's side which is between stage 3 to 5 and this cost can be defined as in previous section.

Email provides a faster and smoother communication approach, but when used by spammers, it could turn its advantages towards being one of the reason contributing to a big cost for a company. Company need to spend their money on spam filtering tools, hire related personnel to deal with this problem and even provide training for their employees to improve their understanding of spam. Support cost for spam filter is associated with stage 5 and 6 in the case of lifecycle of email spam and can be defined as in Equation 3.

Employees also need to spend time in checking spam folder to avoid losing legitimate messages. They then need to read and delete or mark as non spam for each email messages for future safety. This situation worsen if the employee decide to reply or get interested in the product or services by the spam messages which lead them to spend more time browsing unnecessary websites. They also usually take time before getting back to do their work. This affects the employees' productivity [57, 58]. This cost is associated with stage 5 and it is calculated as in Equation 4.

As far as software costs goes, open source tools that can be deployed in company's server are SpamAssassin and Anti-Spam SMTP Proxy Server. Conversely, there are commercial tools which are EMP 7 Enterprise Anti Spam, SpamFighter Exchange Module which could costs AUD40($36) per user, Spamfighter Mail Gateway with AUD22.50($20.40) per user and Web & Mail Security GFIMail Essentials Anti-spam Solution for Exchange/SMTP/Lotus with 10 to 24 Mailboxes at a price of AUD43.05 per mailbox.

Based on our observation, companies would have to bear the cost of buying one license for every single user and this cost usually gets cheaper if they buy more licenses. In this case, a small company would have to pay a higher price for anti spam tools. For example, the cost of using EMP 7 Enterprise Anti Spam is AUD28.60 ($25.99) for every user. This cost gets cheaper if the company buy the license for more users that is AUD10.60 ($9.60) for each email recipient for 500 to 999 users. For additional filtering, users themselves could equip their desktop with anti spam filter tools which could cost them below AUD55 (US$50).

5.3 Cost of Email Spam for ISP

ISPs need to spend additional costs for storage and bandwidth to provide services to their client who can be spammers or normal users. Spam which is transmitted at the same time with legitimate content causes increase usage of network bandwidth and storage capacity. This cost parameters are similarly defined as in Equation 1 and Equation 2. ISP could also deploy anti spam tools such as MailCleaner, SpamFilter ISP from Logsat Software which cost AUD660 ($600) per server. Another tool called SpamTitan with single appliance license for 50 users covered for 1 year could be bought at a price of AUD435 ($395).

6 Discussion and Conclusion

Symantec reported that the average of spam volume is 87% of all email messages[59]. Another firm, Ferris Research in their recent report estimated the cost of email spam for the whole worldwide is $130 billion[8]. It shows that spam impacts are not limited to an individual but to a certain degree it could affects a country. Several figures have been produced by Japanese researchers showing that a huge amount of money was spent or wasted due to spam mail. It is reported that 960 billion yen was calculated for GDP loss from several big industries in Japan [57].It is a great loss suffered just to pay for time spent in handling email spam and the labour needed in order to process spam mails. This labour loss is associated with time spent for email spam such as what we have defined in Equation 4. It is proved by [57, 58, 60] that email spam harm the economy of a country which without proper effort could further reduces the economic growth globally.

This paper first identifies the lifecycle of email spam and the cost of spam associated with each lifecycle stages. Considering that we are going to measure the amount of spam accurately based on a huge reference of spam collection, there is a need to formulate all associated costs accordingly which was done in Section 5 where cost categories have been defined with 16 related parameters.

Regardless of facing all these costs, it is important to take note that there are several key issues in calculating the cost of spam. A considerable amount of report has been published on the cost of email spam. Nevertheless, there is no guarantee that surveys or report done are unbiased as most of the report will finally try to show that the cost of spam can be reduced by using their product, hence it is also possible that in earlier stage, they would try to maximize the cost of spam.

This paper provide an overview of three different stakeholders bearing spam costs including spammer's cost to spam, ISP's cost and company's cost in combating spam. We are trying to define a much more general way in calculating spam costs in a case where a huge reference real data collection is done. As a conclusion, it is important to continue on researching on real time spam cost calculator in order to ensure that company are spending a considerable amount of cost in combating spam. By studying cost of spam, it is hoped that spammer's in future would have to spend more than the amount that they gain so that they would lose their interest/benefit for spamming.

References

1. *Number of e-mail users worldwide to reach 1.6 billion in 2011, says Radicati Group* [cited 2009 15 November]; Available from: http://software.tekrati.com/research/9512/.
2. Windows & .NET Magazine, *The Secret Cost of Spam*.
3. Nucleus Research, *Spam : The Repeat Offender*, in *Research Note*. 2007.
4. Osterman Research Inc., *How Spamhaus Cost-Effectively Eliminates Spam*, in *An Osterman Research White Paper*. March 2008.
5. Hayati, P. and V. Potdar, *Evaluation of spam detection and prevention frameworks for email and image spam: a state of art*, in *Proceedings of the 10th International Conference on Information Integration and Web-based Applications \& Services*. 2008, ACM: Linz, Austria.
6. *PUBLIC LAW 108–187—DEC. 16, 2003*. 2003 15 November 2009 [cited; Available from: http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=108_cong_public_laws&docid=f:publ187.108.pdf
7. Nagamalai, D., B.C. Dhinakaran, and J.-K. Lee, *An in-depth analysis of spam and spammers*. 2009.
8. Ferris Research, *Spam, Spammers and Spam Control*. March 2009, Ferris Research: San Francisco, Calif, USA.
9. Raz, U. *How do spammers harvest email addresses?* 2009 [cited 28 October 2009]; Available from: http://www.private.org.il/harvest.html.
10. Lanxon, N. *More email passwords posted to the Internet: Experts detect spam increase*. 2009 20 October 2009 [cited; Available from: http://crave.cnet.co.uk/software/0,39029471,49303853,00.htm.
11. Download3000. *Speed Email Extractor 6.0 Free Download*. [cited 2009 13 November]; Available from: http://www.download3000.com/download-speed-email-extractor-count-reg-44069.html.
12. Brothersoft. *Power Email Harvester 1.45 Download*. [cited 2009 13 November]; Available from: http://www.brothersoft.com/power-email-harvester-86025.html.
13. Emailgrabber.net. *Email Grabber*. [cited 2009 13 November]; Available from: http://www.emailgrabber.net/.
14. Tenmax.com. *Teleport Pro - Offline Browsing Webspider*. [cited 2009 13 November]; Available from: http://www.tenmax.com/teleport/pro/home.htm.
15. E-mailSmartz. *Email Marketing Software | Email Generator*. [cited 2009 13 November]; Available from: http://www.emailsmartz.com/.
16. *Spam Preventer 1.0*. [cited 2009 13 November]; Available from: http://wareseeker.com/Security-Privacy/spam-preventer-1.0.zip/346607.

17. Spammer-X, J. Posluns, and S. Sjouwerman, *Inside the SPAM Cartel: By Spammer-X*. 2004: Syngress.
18. San, G. *Creating Email Copy - Spam Words to Avoid* [cited 2009 24 November]; Available from: http://ezinearticles.com/?Creating-Email-Copy---Spam-Words-to-Avoid&id=2510485.
19. *How to stop yourself from being unfairly labelled a spammer*. [cited 2009 24 November]; Available from: http://www.emailaddresses.com/email_spam_lists.htm.
20. Connick, E. *How to stop your e-mail from being seen as spam*. [cited 2009 24 November]; Available from: http://www.helium.com/items/1063324-how-to-stop-your-e-mail-from-being-seen-as-spam.
21. Hayati, P., et al., *HoneySpam 2.0: Profiling Web Spambot Behaviour*. *Accepted*, in *PRIMA 2009*. 2009: Nagoya, Japan.
22. Brothersoft. *BulkMailer 2.3 Download*. [cited 2009 13 November]; Available from: http://www.brothersoft.com/bulk-mailer-16936.html.
23. Wareseeker. *Avalanche 98.8.20*. [cited 2009 13 November]; Available from: http://wareseeker.com/Email-Tools/avalanche-98.8.20.zip/4692
24. Brothersoft. *DarkMailer 1.13 Download*. [cited 2009 13 November]; Available from: http://www.brothersoft.com/dark-mailer-90307.html.
25. Postcast Server. *PostCast Server - Free SMTP Server*. [cited 2009 13 November]; Available from: http://www.postcastserver.com/download/.
26. Sourceforge. *YAPH - Yet Another Proxy Hunter for HTTP Connect, Socks4 and Socks5 Servers* [cited 2009 13 November]; Available from: http://yaph.sourceforge.net/.
27. *Proxy Hunter*. [cited 2009 13 November]; Available from: www.proxyblind.org/proxy_hunter.shtml.
28. AWeber Communications. *Email Marketing Software, Email Newletters and Autoresponders by AWeber*. [cited 2009 14 November]; Available from: http://www.aweber.com/.
29. Constant Contact. *Email Marketing Solutions from Constant Contact*. [cited 2009 14 November]; Available from: http://www.constantcontact.com/index.jsp.
30. JangoMail. *Email Marketing & Personalized Business Email Delivery Service*. [cited 2009 14 November]; Available from: http://www.jangomail.com/.
31. Judge, P. (2003) *The state of the spam problem*. **Volume**,
32. Stewart, J. *Top Spam Botnets Exposed*. April 8, 2008 [cited 2009 29 October]; Available from: http://www.secureworks.com/research/threats/topbotnets/.
33. *Mailcleaner Anti spam solution for enterprise or ISP*. [cited 2009 14 November]; Available from: http://www.mailcleaner.net/.
34. LogSat Software. *Spam Filter ISP - spam filter server for Windows | Spam Filter ISP*. [cited 2009 14 November]; Available from: http://www.logsat.com/.
35. *SpamTitan | Leading way to better e-mail security*. [cited 2009 14 November]; Available from: http://www.spamtitan.com/.
36. Kaspersky Lab. *Kaspersky Anti-Spam ISP Edition*. [cited 2009 14 November]; Available from: http://www.kaspersky.com/corporatesolutions?chapter=4157640.
37. *MX Force:: Managed Email Security Services* [cited 2009 14 November]; Available from: http://www.mxforce.com/.
38. Spamhaus. *Spamhaus Statistics : The Top 10*. The 10 Worst Spam Service ISPs 2009 28 October 2009 [cited 28 October 2009]; Available from: http://www.spamhaus.org/statistics/networks.lasso.
39. *Spam Blacklist Checker, RBL Black listed IP address, blacklist, check blacklists*. [cited 2009 14 November]; Available from: http://www.spamblacklist.com.au/.
40. *RBL.JP*. [cited 2009 14 November]; Available from: http://www.rbl.jp/.

41. Sipior, J.C., B.T. Ward, and P.G. Bonner, *Should spam be on the menu?* Communications of the ACM, 2004. **47**(6): p. 59-63.
42. Atkins, S. *Size and Cost of the Problem*. in *56th IETF Meeting*. March 2003. San Francisco, CA.
43. Korsmeyer. *Enterprise Email Security for Exchange Server, Domino and GroupWise*. [cited 2009 14 November]; Available from: http://www.jak.com/.
44. Wareseeker. *SPAMfighter Exchange Module 3.5.0.0*. [cited 2009 14 November]; Available from: http://wareseeker.com/Email-Tools/spamfighter-exchange-module-3.5.0.0.zip/3667af4a9.
45. SPAMfighter. *Anti Spam Gateway for Mail Serves*. [cited 2009 14 November]; Available from: http://www.spamfighter.com/Product_SMTP.asp.
46. GFI. *Anti-spam filter for Exchange Server and Lotus Notes*. [cited 2009 14 November]; Available from: http://www.gfi.com/mes/.
47. *SpamAssassin : Welcome to SpamAssassin*. [cited 2009 14 November]; Available from: http://spamassassin.apache.org/.
48. *Stop spam with the Anti-Spam-SMTP-Proxy(ASSP)*. [cited 2009 14 November]; Available from: http://assp.sourceforge.net/.
49. BullGuard. *BullGuard Antivirus, Antispyware, Firewall, Spamfilter, Backup and Support*. [cited 2009 14 November]; Available from: http://www.bullguard.com/main.aspx.
50. Cloudmark. *Cloudmark Messaging Security - Block Spam, Fraud, Phishing & Viruses*. [cited 2009 14 November]; Available from: http://www.cloudmark.com/en/home.html.
51. *Spam Blocker from Clear My Mail, Spam Filter and Anti-Spam Solution*. [cited 2009 14 November]; Available from: http://www.clearmymail.com/default.aspx.
52. *Spam Blocker for Microsoft Outlook*. [cited 2009 14 November]; Available from: http://www.mailfrontier.com/products_matador.html.
53. *Spam Filter Express is anti-spam software and spam blocker to stop spam email.* [cited 2009 14 November]; Available from: http://www.spam-filter-express.com/.
54. Ferris Research. *Industry Statistics.* 2009 [cited 2009 5 November]; Available from: http://www.ferris.com/?page_id=1078.
55. *Laboratory for Web Algorithmics.* [cited 2009 25 November]; Available from: http://law.dsi.unimi.it/index.php?option=com_frontpage&Itemid=1.
56. Girardi, C., F. Ricca, and P. Tonella (2006) *Web crawlers compared*. International Journal of Web Information Systems **Volume**, 85-94
57. Takemura, T. and H. Ebara, *Economic loss caused by spam mail in japanese industries.* RCSS Discussion Paper Series, 2008.
58. Takemura, T. and H. Ebara. *Spam mail reduces economic effects*. in *Digital Society, 2008 Second International Conference on the*. 2008.
59. Morss, D., D. Harnett, and C. Edwards, *State of Spam : A Monthly Report*. September 2009, Symantec.
60. Ukai, Y. and T. Takemura, *Spam mails impede economic growth.* The Review of Socionetwork Strategies, 2007. **1**(1).