Summer 2005

E-Mail Spam Filtering Solution For The Western Interstate Commission For Higher Education (Wiche)

Jerry Worley
Regis University

Regis University
School for Professional Studies Graduate Programs
Final Project/Thesis

Disclaimer

Abstract

E-mail Spam Filtering at
The Western Interstate Commission for Higher Education (WICHE)

By Jerry Worley

WICHE staff, consultants and constituents that connect to WICHE's network are provided with internet access and e-mail accounts. In the past, email was not monitored or controlled, other than scanning e-mail attachments for viruses with Norton Antivirus for Exchange Server. This has become a problem for several reasons. Just a few of the reasons are viruses, wasted staff time, inappropriate material stored on the network and inappropriate use of network resources. The WICHE Spam filtering solution includes monitoring and restriction of incoming and outgoing email, email attachment and storage limitations and filtering incoming email for Spam. All internal, incoming and outgoing e-mail is scanned on the Exchange server for inappropriate key words. Limits are placed on email accounts, such as mailbox size, message size, attachment size and maximum number of recipients. All incoming e-mail is scanned for several indicators of Spam. These indicators include originating mail server, message subject, sender address, number of recipients and message content, based on key words. The messages are scanned and stopped at a front end SMTP server before they enter WICHE's internal network.

Table of Contents

List of Illustrations/Figures/Tables

Chapter One

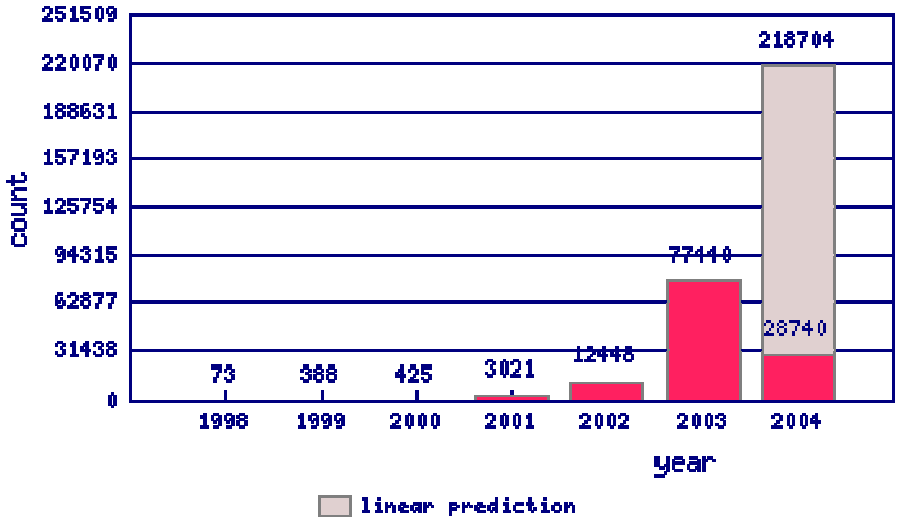
1.0 Introduction

This project addresses the problem of Spam email at the Western Interstate Commission for Higher Education, also referred to as WICHE. Over the past few years, the organization has grown from using computers for a select few specialized tasks to relying on them to complete the majority of the day to day workload. The central tool that users have adopted for communicating and processing information is email. Along with the legitimate email traffic comes the opportunity for businesses to use this new medium as a cheap and effective method for advertising. All of this new, unrequested and sometimes offensive email traffic has been given the label of "Spam". Spam is a new and very quickly growing problem. Implementing a new email monitoring and filtering solution will allow the I.T. department to limit the unproductive and potentially damaging files that are being sent through the internet via email.

1.1 Problem statement

Staff at WICHE all have desktop or laptop computers and access to the Internet and e-mail. Everyone uses email for communication and transferring files on a daily basis. There are also many very undesirable and sometimes offensive messages that are transferred via e-mail. When these activities are not restricted on the network there is the potential for many problems. Viruses can be distributed as E-mail attachments, spread across WICHE's internal network through file shares, and spread from WICHE's network to other company's networks, which has a negative impact on WICHE's reputation. Lost productivity is also an important consideration. Staff time lost sorting

through, reading and removing undesirable messages on company time is a major consideration. Spam email is increasing at a very rapid rate, as the following graphic from Bloodgate.com (2004) shows.



SPAM statistics (actual Spam shown in red)

Another consideration is improper resource usage. Bandwidth is consumed by inappropriate traffic. Server disk space and processor usage by inappropriate traffic can be considerable. I.T. staff time fixing problems caused by undesirable email traffic is another side effect.

1.2 Review of existing situation

WICHE currently uses SurfControl email filter, version 4.7 for email filtering. SurfControl runs on a front end mail server in the DMZ zone, on port 25. The same physical server also functions as a front end Microsoft Exchange server, running on port 26. Incoming mail is filtered by a set of rules, which scan for included/excluded

addresses, domains and content dictionaries. Any mail that is flagged has the subject line appended to include a possible Spam warning. All outgoing mail also goes through SurfControl, but is allowed through by the first rule, which checks for a source address from WICHE's internal network. To control outgoing spam, the firewall is blocking port 25 for all computers except the mail servers. This prevents all computers other than mail servers from relaying email. To control relaying through mail servers, they will only accept connections from authenticated clients or other internal mail servers. WICHE does do some mass mailing, but all of it is setup and monitored by the I.T. department and mailing lists are generally by subscription only and must have an operational unsubscribe link attached to them. There is not currently any client software for Spam control, but Outlook clients are configured with a Spam folder and a rule to route messages with possible Spam in the subject to that folder. The existing system does filter out most of the Spam, but also has some drawbacks.

1.3 Goals of project

The main goal of this project is to filter out virtually all of the incoming Spam without stopping any legitimate traffic. The system must be flexible enough to allow each client to receive mail that they want to let through and to block the mail that they do not want to let through. It must also be easy enough to configure and use that all staff can use it without extensive training on the software. The system cannot take up very much I.T. staff time, as support time is already at a premium. The system must be flexible enough to allow traffic that needs to get through, regardless of whether the senders address is on

any particular Black List, so that WICHE staff interaction with customers is not interrupted.

1.4 Barriers and/or issues

WICHE is a customer focused organization, and must post several email addresses on its websites. WICHE staff must remain available to all of their customers at all times. We cannot block email that is from WICHE customers and if we do, it is the responsibility of WICHE's I.T. department to ensure that those email messages go through in the future. If we subscribe to a Relay Black List service and their servers are attacked by a DOS attack, we must allow incoming messages that go through that filter. However, if we let all of the incoming email through it becomes unmanageable. We have staff mailboxes that have five hundred messages per day coming in, and about 90% of them are considered Spam.

1.5 Scope of project

This project will address the growing numbers of Spam email messages and will attempt to create a solution that will block more Spam from the users' mailboxes. Users spend far too much time sorting through Spam email. The project will analyze available methods of Spam reduction and will select the best method for WICHE staff. The solution that is selected must also allow legitimate email to get through. Packaged products for turnkey services, server side solutions and client side solutions will be analyzed. The creation of a custom software program will also be considered to compliment a packaged solution that may address some, but not all of the requirements.

The solution that is chosen will save the organization money by reducing the amount of staff time and network resources spent processing Spam. It will also increase the level of customer service by allowing staff to respond to legitimate email faster, respond to legitimate email that is not blocked and avoiding confusion by separating legitimate email from Spam in the user mailboxes.

1.6 Definition of Terms

Actual Spam –

Email messages that have been marked by a mail filtering system as being spam, and are positively determined to be spam messages.

Anti-spam compliance –

All email that originates on the WICHE network must either be for personal correspondence or must be sent to a subscription based list where the recipients request membership. Each message sent to the list must have a method for recipients to opt out if they no longer wish to receive the subscription based messages.

DOS attack –

An attack by a hacker that is designed to overwhelm and shut down a server or network by sending a large number of bogus requests for service.

False positives –

Email messages that have been marked by a mail filtering system as being spam, but are actually not spam messages.

Flagged as Spam –

Email messages that have been marked by a mail filtering system as being spam. These messages may or may not actually be spam messages.

Filtering –

Monitoring email messages and flagging them based on source, subject, content or other monitoring criteria.

RBL –

Relay Black List. A list of mail servers, developed by a third party monitoring service, that indicates that they may be performing relay activities that could result in the propagation of Spam email.

Relaying –

Any computer that sends and/or receives email through a SMTP connection is considered to be relaying email. The computer does not need to be a mail server to be relaying mail. To help prevent spam, all computers that are not required to relay mail should be configured to prevent relaying. All computers that are required to relay mail should be configured to prevent unauthorized relaying.

SMTP –

Simple Mail Transfer Protocol. The protocol that mail servers commonly use to communicate with each other. Generally resides on port 25.

Spam –

Unsolicited e-mail, often of a commercial nature, sent indiscriminately to multiple mailing lists, individuals, or newsgroups; junk e-mail.

Virus –

A self-replicating computer program, which spreads by inserting copies of its code into each computer that it infects. A Virus also commonly does damage to the computer it resides on, such as deleting files or sending out Denial of Service attacks.

WICHE -

The Western Interstate Commission for Higher Education. A non-profit organization, focused on higher education issues in the western United States.

Chapter Two

2.1 Research methods used

Several different methods were used to perform research on the spam problem at WICHE. The most important factors in doing research were to use information that was provided by others who are currently facing problems with Spam on their networks, to address the needs of the staff at WICHE and to objectively compare available products using accurate information. In order to meet each of these criteria, the research methods chosen were the Internet, WICHE staff interviews, and product literature for selected Spam filtering products.

2.1.1 Internet

The main reason for using the Internet for doing research was that a wide variety of applicable experience and opinions could be reviewed, using the most current information possible. When network administrators have a problem such as Spam that is propagated through the Internet, they generally use the internet to share information on how to most effectively combat that problem. All of the product literature for each of the package solutions available was also accessed over the Internet.

2.1.2 Staff Interviews

Email is used at WICHE for a wide variety of purposes, and staff requests vary greatly. Some users only use email for internal correspondence, while other users correspond with the public through the email system. Some users receive five to ten messages per day, while others receive up to five hundred per day. Some mailboxes are

over Two GB in size, while others are relatively small. The following figure shows the large size of many of the mailboxes on the WICHE network.

Mailboxes

Mailbox	Size (KB)	Total Items
Jer...	2,398,871	37,509
Pat ...	2,242,817	38,430
Che...	1,677,075	34,490
De...	1,275,409	52,831
She...	1,128,214	48,025
Ann...	986,282	22,796
Rac...	982,131	13,539
Mar...	818,274	13,674
Mic...	804,986	17,818
Crai...	569,791	27,552
Mar...	566,384	12,199
Dav...	481,233	7,917
Mar...	452,092	5,517
Sall...	448,250	9,535
Jen...	413,827	6,296
Rus...	409,005	5,568
Chu...	383,396	5,559
Sco...	350,357	2,977
Sha...	315,537	13,672
Deb...	250,296	5,989
San...	193,378	8,882
Den...	164,630	15,838
Bria...	159,910	4,070
Jerr...	144,238	4,401

WICHE Mailbox sizes

For some staff, it is critical that absolutely no legitimate email be restricted from coming in to them. For other staff, they would rather see more questionable messages being blocked, even if it means a small percentage of legitimate email being blocked as well. There is no distinct way to separate the users into different categories of email usage, so it was important to poll all staff on how the Spam problem relates to their unique circumstances and to come up with a solution that meets the needs of as many staff as

possible. A survey was distributed to all staff to solicit input from everyone, and six staff members were chosen at random for face to face interviews.

2.1.3 Product Literature

For each of the packaged products available, they have a product overview, a product datasheet, a white paper, a demo version of the product, sample data and reports from the product, administrator guides and user guides. The product overviews were used to compare the basic features of each product. The product datasheets were then compared to look at more details about the products, such as the hardware requirements and the specific methods used by the product to block Spam. The white papers were reviewed to gain more insight into the details of the spam blocking methods and the general philosophy of the product's design team. The demo versions were downloaded and installed on a test server and were used for experimenting with the product on a trial basis. The sample data and reports were reviewed to determine user friendliness and reporting features of each product. The administrator and user guides also show a great deal about how user friendly and easy to install and configure each product is. Each of these pieces of product literature helped to determine what the exact capabilities of each packaged product were.

2.2 Methods currently used to block Spam

There are many methods that are actually used to block Spam. SurfControl Plc (2003) addresses many of these methods. Content filters search through the content of the message for keywords and score the message based on those keywords. If the message

scores over a certain score, it is flagged as Spam. The advantage of content filters is that they look at more than just the address the message is coming from, and can positively block all occurrences of a specific word. They are effective in blocking offensive and derogatory language. They also have many drawbacks. They can block legitimate mail, such as blocking a message with the word “breast” when the sender was referring to the chicken breast he had for lunch. They can also be circumvented by inserting graphics of a word instead of text or inserting symbols into a word, such as “bre@st”. White lists contain a list of email addresses that are always allowed and never marked as Spam. They are effective in allowing all email from a legitimate address, but are time consuming to configure and do not address the problem of spammers being able to spoof the sending address. Black Lists contain a list of email addresses that are always marked as Spam. They are effective in blocking known spammers, but also do not address the spoofed sending address limitation of email or the problem of spammers sending mail through open relays in legitimate mail servers. If the legitimate mail servers are added to the blacklists, then any legitimate email coming from them will also be blocked. Bayesian filters start with no rules and learn from the user, based on what the user wants to block, and what the user wants to allow. This is effective because it blocks only what the user decides to block, but also requires the user to enter data before the filter is effective for that user. Challenge-response sends a message back to the sender, and makes them manually verify that they want to contact the user before the mail is allowed through. Reverse DNS lookups verify that the message is actually coming from the IP address listed as the mail server’s address in the sender address. Challenge-response and reverse DNS are often used in conjunction with Black Lists and white lists can be very effective

in blocking Spam, but can also block a large amount of legitimate traffic, depending on the specifics of the implementation. Spam filtering packages may use one or several of these methods to block Spam. The product literature for the packages usually indicates which methods it uses to block Spam, but seldom goes into any detail about how those methods are specifically used.

2.3 Packaged products available

There are currently many packaged products available for filtering Spam from email. They can be divided into three main categories, turnkey services, server side and client side. The server side products run on a mail server and the client side programs are installed on each email client computer. Turnkey solutions are also available, which are basically server side products that an external provider is running for you. The price range and functionality varies greatly from product to product and some are more customizable than others. Turnkey services will not be analyzed further because of cost, control and customization limitations. The following table compares the advantages and disadvantage of each type of solution.

Feature	**Turnkey Service**	**Server Side**	**Client Side**
Blocks Spam	Yes	Yes	Yes
Customizable by administrator	No	Yes	No
Customizable by users	Partial	No	Yes
Black Lists	All	All	Custom

White lists	All	All	Custom
RBL's	Yes	Yes	No
Spam dictionaries	Yes	Custom	No
subject to DOS attacks	Yes	No	No
Requires user training	No	No	Yes
Price	Highest	Average	Lowest
Effectiveness in blocking Spam	Good at server level	Good at server level	Depends on user input
Blocks legitimate email	Yes	Depends on Admin options	Depends on user input

Feature Comparison

2.3.1- Server Side

The server side products currently available are generally used by network administrators to control Spam coming into, internally and leaving their networks. Server side solutions are effective because they are controlled by the administrator and require no input from the users to be effective. They also allow the product vendor to periodically update the software to combat emerging threats as they are created. This is an important feature in the battle against Spam, because the methods that spammers use change on a daily basis. Some of the more popular ones are SurfControl, SpamAssassin and SpamFilter ISP. They use a wide variety of methods to identify Spam and can be configured to block, quarantine or just flag the Spam and generate many different log file entries and alarms. (SpamAssassin Project, 2005) The main disadvantage of server side

programs is that they block the Spam that the network administrator wants to block and not what the users want to block, and are generally not configurable from user to user.

2.3.2 - Client Side

There are many products available to block Spam at the mail client. A few of the packages available are Mailwasher, McAffee Spamkiller, SpamNet, and AntiSpamWare. The advantage of client side products is that each user can generally configure their own Spam filtering software to filter out what they consider Spam and allow the mail that they don't consider Spam. The disadvantages are that they create more work for the network administrators, they do not filter any mail at the server and they must be installed and configured separately for each client, each time that client logs into a new computer.

2.4 Discussion of filtering internal and outgoing messages

WICHE has made the decision to not apply Spam filtering to internal and outgoing messages. The organization is small, and internal Spam is not a problem. The I.T. department is also structured as a service provider to the other departments and is paid to provide services to them. Each user pays a set fee to the I.T. Department each month for I.T. Services. In a traditional environment, the I.T. department would be protecting the interest of the organization as a whole by monitoring internal and outgoing email. In the WICHE environment, the fact that each department pays for its own usage is enough of a deterrent to internally generated Spam. The firewall does block SMTP for all addresses except the mail servers, which results in all email going out through mail servers. All users are authenticated on those servers to prevent relaying. Any email that is

generated from a bulk mailer on a server is configured by the I.T. department and is checked for anti-Spam compliance before being put online.

2.5 I.T. resources

Resources are somewhat limited at WICHE. The level of desktop support is not high enough to provide individualized assistance with Spam filtering. Any client level application must be extremely easy to use and have a default configuration that will work well for the majority of staff. The organization also rotates computers from user to user as the computers get older, which means that if client configurations are stored on the desktop, they must be easily moved from computer to computer. The budget for new hardware and software is also somewhat limited. Funds for I.T. services are generated by user fees, which would need to increase for any significant purchases related to Spam filtering.

2.6 Summary

There are many packaged solutions for filtering Spam. Most of them are either server based and designed for the systems administrator to design the rules, or client based and designed for the user to configure his/her own rules. The packaged solutions use several different methods to block Spam. Each method has its advantages, as well as its disadvantages. It is also possible to design a custom solution that will work together with a packaged system. This would allow the users to have additional functionality that the packaged system may not have. In order to select the best possible solution for the WICHE network, extensive research needs to be done on the internet, the users must be

interviewed, the product literature must be examined, the packaged products must be tested and the features of the products must be compared with the requirements of the organization.

Chapter Three: Project Methodology

3.1 - Research & Analysis Phase

The research and analysis phase of this project consists of analyzing the effectiveness of the existing solution, soliciting input from staff, analyzing the options that are currently available and selecting a new solution based on the business and technical requirements.

3.1.1 - Problem Analysis

The existing solution provides some of the functionality for the server side component, but not all of it. There is still a high percentage of the Spam email that is getting into staff mailboxes, and an unacceptable amount of legitimate email that is getting blocked. The existing solution also does not have any client side functionality at all. This is something that must be addressed in the solution.

3.1.1.1- Interviews and information-gathering

A Spam survey was distributed to all staff. The results of the survey are shown in the following table.

SPAM Survey

	Please rate the following aspects of WICHE's current SPAM filtering system on a scale of 1-5:
Best--------Worst	
1 2 3 4 5	

4	**5**	6	0	0	Makes managing my email easier
1	**6**	5	3	0	Removes spam from my inbox
1	**9**	2	3	0	Does not remove legitimate email from my inbox
8	**5**	2	0	0	Allows me to recover legitimate email that has been flagged as spam

Please rate the importance of the following features in a SPAM filtering system on a scale of 1-5:

Most.-------Least					
1	**2**	**3**	**4**	**5**	
10	5	0	0	0	Removes spam from my inbox
9	3	2	1	0	Does not remove legitimate email from my inbox
13	2	0	0	0	Allows me to recover legitimate email that has been flagged as spam
6	2	**0**	4	3	Notifies sender if message is flagged as spam
6	**2**	3	1	2	Notifies me if message is flagged as spam
3	3	**5**	3	0	Spam email address list can be customized by me (the user)
3	1	3	**5**	3	Relies on the sender to remove his/her own name from our spam list
0	1	6	**4**	3	Relies on SPAM blacklist service, and requires the sender to remove his/her own name from blacklist
4	**7**	2	1	0	Relies on SPAM blacklist service, and allows me (the user) to remove legitimate names from blacklist
5	**3**	4	1	1	List of people (email addresses) that should never be flagged as

spam, even if on blacklist or other spam list

Method to not generate auto replies to spammers when “out of

4 **6** 2 1 2 office”

0 1 **1** 0 0 Other client flags subjects, ease of use_____

In a perfect world, we would filter out all spam and allow all legitimate email through. Assuming we do not live in a perfect world, please answer the following questions:

Would you rather:

5 A. filter out all the spam and occasionally block a legitimate email

2 B. Filter out some of the spam and never block a legitimate email

C. Somewhere in between A and B, with some spam getting through and

6 some legitimate email getting blocked

If legitimate email is flagged as spam, would you rather:

1 A. leave it up to the sender to remove his/her name from the spam list

6 B. leave it up to each user to maintain his/her own list of allowed addresses

C. flag the message, but leave it in a place where the user can get to it, much

6 as we are now doing

Comments

current system is working well

in addition to flagging the message, it would be nice if we could teach the system to

allow mail from legitimate users, much like the training system on spambayes

no complaints

System does not flag as much spam as it should

3.1.1.2 - Requirements Analysis

Requirements analysis consists of the following categories:

- Business Requirements
- Technical Requirements
- Software Requirements
- Hardware Requirements
- Training Requirements

The business requirements for this project are to create a Spam filtering system for WICHE that filters out incoming Spam email but does not block legitimate email. It is very important to not place restrictions on incoming email because the organization is one that provides services to the general public, and encourages communication from the general public. This system must contain a component that blocks Spam without any input from the user. This is important because some users do not have the time or knowledge level to configure a Spam filter themselves. The system must also contain a component that each user can configure independently of the server configuration. Each user may require different levels of customization, and the system must allow them to customize their own filter if they need to do that.

The technical requirements are for a server based system that is administered by the I.T. department and a client tool, which allows each user to customize their filter if they need to. Both components must be used, in order to meet the business requirements.

The Spam filter currently in place for the server side is SurfControl SMTP filter, version 4.7. It is meeting some of the requirements for blocking Spam, but not all of them. This software will either need to be upgraded to a more functional version or replaced with a different server side component. There currently is not a client side software package in place. A client tool must be developed to add the functionality that the users require.

The mail servers currently in use are the back end Exchange 2003 server for staff mailboxes, the exchange front end mail relay in the DMZ zone of the network, and the SurfControl SMTP filter, which runs on the same machine as the exchange front end. All servers have Intel processors and are running Windows 2000 or Windows 2003 Server. All client machines have Intel processors and are running Windows XP. The budget does not include any funds for purchasing new hardware, so the solution must run on the existing servers and client computer.

Staff training resources are very limited at WICHE. Many staff travel the majority of the time and have a difficult time scheduling time for training. For these users, the server component must run without any user intervention. For other users that require extensive client side customization, training sessions will be provided on a periodic basis.

Staff turnover can be high for some positions, so training must be scheduled regularly to accommodate new staff.

3.1.2 - Decision Analysis

The final decision was made to continue to use SurfControl SMTP filter for the server side software, but to upgrade to version 5.0 for several new features that it adds. The new version allows the RBL functionality to be used effectively without blocking legitimate email, and if it does block legitimate email, the sender will be notified. It also allows customized White Lists and Black Lists to be added for each user, which must be available in order to add the client side software. The final decision for the client side software was made to create a simple Java application that will run in the background on each client computer. The java application will strip the subject or email address from email messages and store them in white list and Black List files for each user. The white lists and Black Lists will be applied to the server filter for each user. This client tool meets the requirement that each user needs to have the option to customize their own list of allowed and denied email addresses if they require more or less filtering than the standard server configuration.

3.2 - Design Phase

The design phase of the project is very important, and had to be completed before the construction phase was started. All staff were updated with the plan details, and construction was coordinated around any major organizational events that could have

been affected by possible problems with the implementation. The project plan, support plan and training plan were each developed in this phase.

3.2.1 - Create the Project Plan

The project plan was to upgrade the server side filtering software package, create and install a client side Java application for customized client filtering and to install each one of these components separately on the SMTP mail relay and all client machines on the network.

3.2.2 - Design the Support Plan

Support for the server side software was purchased with the software package. The decision has been made to purchase a three-year support and maintenance contract from SurfControl. The support will include daily updates to the Spam dictionary files, product patches, service packs and version upgrades and technical support for installation and configuration problems. Since the client software was developed in house, the support for it will be done by WICHE I.T. services as required. The goal was to create a simple client solution that will not require frequent upgrades.

3.2.3 - Design the Training Plan

Training was completed after the upgrade. All staff had the option of attending training, but were not required to attend. The server side software is transparent to the users and the client is not required for basic operation. If the users desired to learn how to increase the level of functionality of their filtering solution, they had the option of

attending the training sessions. Training was held on a recurring basis for new staff and any applicable updates to the system.

3.3 - Construction Phase

The construction phase was broken down into three steps. It consisted of purchasing server software, creating the client side application and testing, in that order. This is the part of the project where the tangible product was created and readied for installation on the network.

3.3.1 – Purchase any required software packages

The decision has been made to upgrade the server side software from SurfControl SMTP filter version 4.7 to SurfControl SMTP filter version 5.0, sp1. This version adds the server side functionality that is required to meet the business and technical requirements. The client side solution will be a custom built Java application that allows the users to create customized white lists and black lists on a user by user basis. It will also include user specific data files that can easily be moved from computer to computer as the users require.

3.3.2 - Create any customized software that may be required

The custom Java solution for the client side is an application that runs on each client computer. It allows each user to submit addresses and subjects from email messages to white lists or black lists on the SMTP server. The client white lists only affect the user who submits them, and are applied before the server side rules. This is

important because if the server has Black Listed an address and the user has white listed it, the message must get through the filter to that user. The Client side Black List is applied after the server side rules. The Java application also creates a text file, in comma separated file format and stores it on the file server, with the rules for that client. If the client moves to a new computer, the file will ensure that their previous saved settings will still be applied. If the client software is not installed on a client computer, the server side filter will still operate normally.

3.3.3 – Install and test solution on test platform

The server side solution was installed on a test computer first. It was configured with SurfControl 4.7 in an identical configuration to the existing SMTP relay. It was then upgraded to version 5.0, sp1. There were numerous problems with the upgrade. The server was unable to connect to the database, and had to be reconfigured by SurfControl tech support. These changes were documented for the actual server upgrade. Several rules also had to be removed before the upgrade, and then added back in after the upgrade. This was due to a new method that was used in version 5.0 to handle these rules. Again, notes were taken for the actual upgrade. Once the test server was running, the client solution was installed and tested extensively on a desktop computer before being rolled out to the staff members. There were some logic problems in the Java coding, but they were easily corrected. The solution ran for seven days on the test server and client test machine without further incidents.

3.4 - Implementation Phase

Implementation consisted of two phases, the server upgrade and the client rollout. It was important to deliver the upgrade without interrupting service to the users. It was also important to make the experience a positive one and not a negative one. Just like with people, the first impression of a new software package is the lasting one.

3.4.1 - Deliver the product

The server upgrade was implemented first. This was completed at night, in order to have uninterrupted service for the users. The upgrade was completed without incident, due in large part to the testing server catastrophe and the lessons learned from it. The users left on Thursday night with the old version and arrived Friday morning with the new version. The server upgrade was transparent to the users, but generated a large number of help desk calls due to the increased efficiency of the filter. Users thought the email server was down, because they were not receiving the large amount of Spam that they were used to getting.

The client install was implemented next. The Java program was installed on a test group of five users first. They were consulted daily for one week to determine that the software was meeting their needs. In general, they seemed pleased with the product. It was performing as expected. The next step was to install it on the remaining client computers. This was done over the next week. For the most part, this portion of the process was accomplished without incident.

3.4.2 - Deliver the end-user training

User training was completed after the upgrade. This is different than the ordinary process, which would consist of at least some training before the user is upgraded to a new product. This decision was made because of the unique nature of this project, in that the server side operation is transparent to the user, and the client side does not need to be operational, unless the user wants the specific functionality that it adds. The only problem that was discovered with not pre-training the users was the unexpected drop in Spam in their inboxes that was caused by the increased efficiency of the new server side solution. This generated several calls for support, but was quickly dissipated with an explanation memo, distributed to all staff. After the users all had the client software installed, three instances of the one hour training sessions were conducted, in order to provide traveling staff the opportunity to attend. One of the requirements was that the client software was easy to learn to use, so it did not take a long time to train the users. They also had a chance to see and experiment with the client software before being trained on it.

3.5 - Maintenance Phase

Maintenance is an important feature for a Spam filter, because spammers are continually seeking ways to get around Spam filters. In this way, Spam filtering is very similar to virus protection. It requires constant and continuous updates.

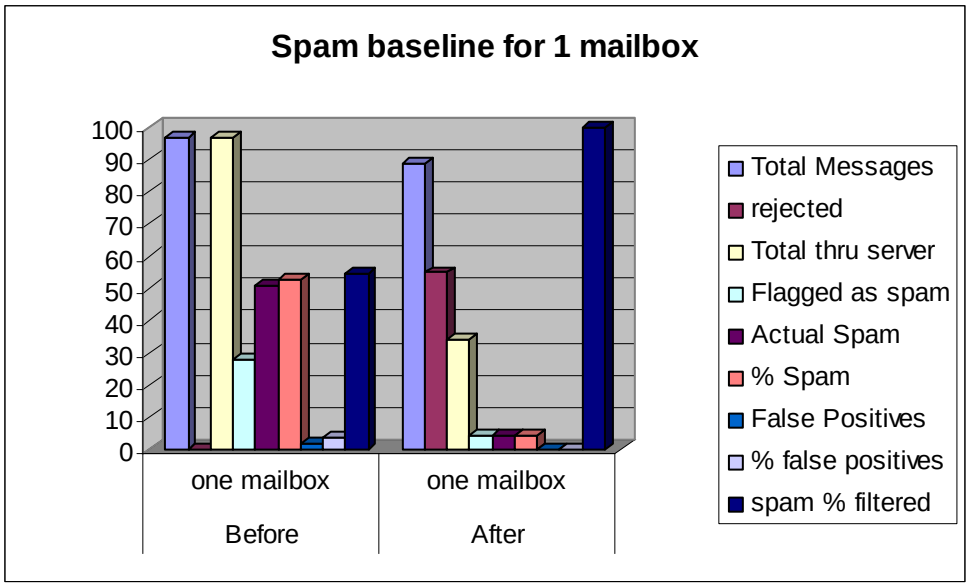
3.5.1 – Create plan for periodic updates

A three year maintenance contract was signed with SurfControl, and live updates were scheduled to occur for the server software, on a daily basis. The client software will

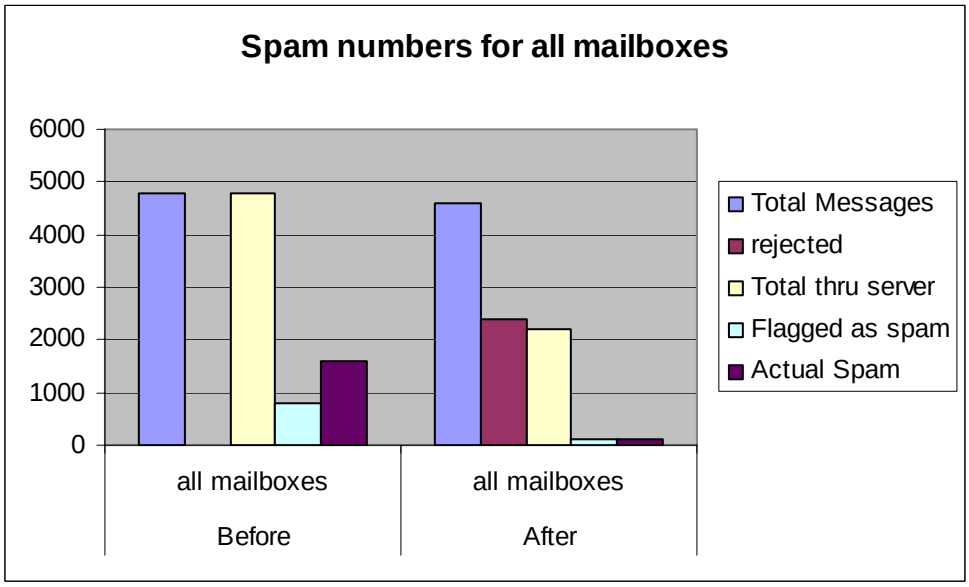
not be updated unless problems with its operation are discovered, but part of the functionality is that the clients will update their own white lists and black lists on an ongoing basis. Each client will contribute to the effectiveness of their own software package.

3.5.2 – Create plan for periodic effectiveness testing

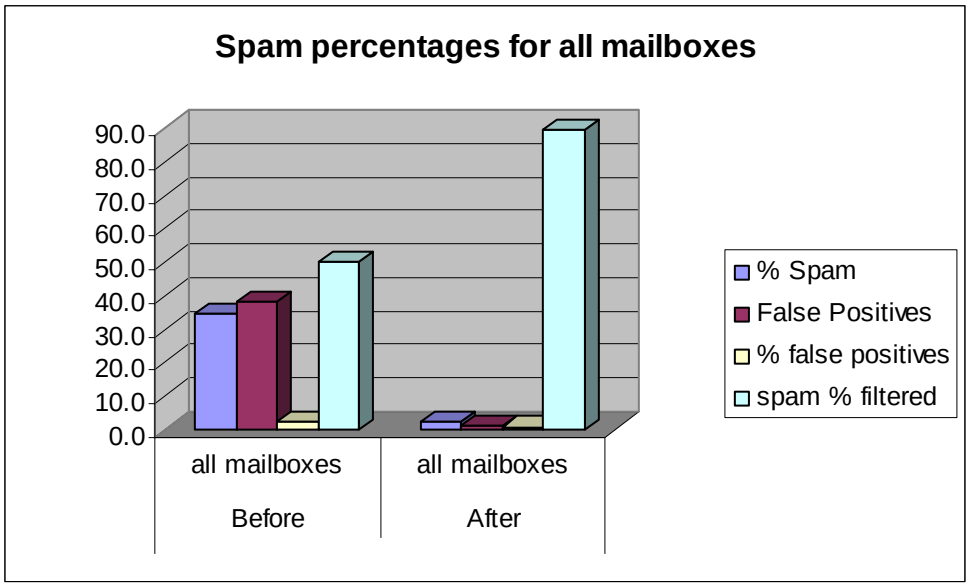
Effectiveness testing will be accomplished by periodically monitoring the incoming email for a twenty-four hour period of time. The following figures show the effectiveness of the previous version just prior to the server upgrade, compared to the effectiveness just after the server upgrade. The first figure shows indicators for one mailbox. This mailbox is used as a Spam attractor mailbox, and the address is posted on the website, as well as being distributed to known spammers. This mailbox should receive more Spam message than any other mailbox on the system. Before the upgrade, messages were being flagged as Spam, but not rejected at the server. The subject line was modified to indicate that they may be “possible Spam”. One key difference is that while no messages were rejected before the implementation, many are being rejected after the implementation. As a result of that, many less messages are being flagged as Spam, but the total percentage of Spam that is being filtered has increased.



The next figure shows total traffic over a twenty four hour period, with the same notes as the first figure. The amount of flagged Spam has decreased, due to the large number of messages that were rejected at the server.



The following figure has percentages for all traffic. The percentage of false positives has decreased, while the percentage of Spam filtered has increased.



The goal for continued effectiveness monitoring will be to have less than ten percent of incoming mail that is Spam, while still filtering over ninety percent of the Spam that is not rejected. If the percent Spam rises over ten percent, then the rejection process is not achieving its goal. If the Spam percent filtered is less than ninety percent, then the filtering rules are not achieving their goal. The percent false positives is also important, and should remain under three percent. The client software will only be as effective as each user makes it. Effectiveness testing will be done quarterly, consisting of a twenty-four hour server sample of the server effectiveness, and a user survey of the client software effectiveness.

3.5.3 - Wrap up the project

The project has been viewed as a success, due to the increased effectiveness of the filtering, the low numbers of problems that occurred during the upgrade and the satisfaction of the staff with the new product. The server upgrade is operating more effectively, and is also operating more efficiently due to the large amounts of Spam that it

is rejecting completely. Over half of the incoming mail traffic is being rejected, which increases efficiency of the back end mail server as well. The client tool is simple, yet effective and allows the users to feel like they have some control over the way their email is filtered.

Chapter Four: Project History

4.1 - How the project began

Five years ago Spam email was a minor inconvenience, at the most. In the last five years, it has grown into an epidemic. Many of WICHE's staff members receive several hundred legitimate email messages each day. When they are combined with several hundred Spam email messages, the workload becomes unbearable. A study of spammer techniques of harvesting email addresses revealed that they get them from many of the same places that the organization needs to make them available to the general public in order to accomplish WICHE's mission. It was determined that WICHE could not prevent spammers from harvesting WICHE's email addresses and needed to filter the incoming Spam instead. WICHE's first Spam filtering solution was very limited, and did not filter it to the level that staff expected it to. After dealing with a growing problem for several years, the decision was made to revisit the problem and come up with a solution that would be more effective and more flexible for different users.

4.2 - How the project was managed

WICHE is a small organization, with about sixty users. The project was managed by the I.T. Manager, with all changes being approved by the department managers who are effectively the customers of the I.T. department. The PC tech assisted with the implementation of the client software, and SurfControl tech support also played a key role, with the I.T. Manager being their point of contact in the organization.

4.3 - Was the project considered a success

Overall, the project was considered a success. It reduced the amount of spam coming into users' inboxes and gave them the functionality to manage their own filters in addition to the filtering on the server side. One unique aspect of a Spam filter is that it can be extremely effective one day, and the next day is extremely ineffective. Only time will tell if the solution continues to be effective. The staff at WICHE is pleased with the performance of the software. There has not been any negative feedback submitted to the I.T. Department, about the changes in the system since the upgrade. The costs were also kept to a minimum, which was one of the key criteria of the project.

4.4 - What changes occurred to the plan

The client side Java package was originally planned to be a plug-in to Microsoft Outlook. This became more difficult than was originally expected, and was therefore changed to a separate program that was placed in the startup folder and runs in the background on all client computers. Another change that was made after the start of the project was to use Relay Black Lists to reject messages from known spammers. Originally, WICHE managers did not want to completely reject any mail from anyone, due to the fear that legitimate users would end up on these lists, as well as the possibility of message failures due to problems connecting with RBL servers. The key factors that changed our minds were the incredibly large amount of Spam that was coming from these RBL listed servers, the increased accuracy of the RBL lists and the ability of the software to allow messages by default, if the RBL server could not be contacted.

4.5 - How did the project end

The last stages of the project included training the users and monitoring the effectiveness of the new software. The monitoring will continue over time to ensure that the solution remains effective. The training will also continue as staff turnover occurs and new updates are implemented.

4.6 – Benefits

There were many benefits to this project. Increased staff productivity and lowered staff frustration with high levels of Spam were the most important ones. Increased network and server speed and efficiency were also important benefits of the project.

4.6.1 – Productivity

Productivity has been increased by the new software. Staff time spent reading and replying to Spam has been minimized. Spam samples taken after the updates were in place indicate that the amount of time spent dealing with spam each day has gone from fifteen minutes per day, per user to under three minutes per day, per user. I.T. staff time dealing with problems related to inappropriate material and non-work related problems has also been cut down to less than half of the previous level. Since the implementation was completed, there has not been any down time related to viruses brought in through links in Spam email messages.

4.6.2 Network resources

A large portion of WICHE's internet bandwidth can be attributed to Spam email. E-mail storage space and mail server resources are also affected by this traffic. The average mailbox size is also increased. As monitoring and filtering restrictions are put in place, these resources will be utilized to their maximum potential. Expensive internet T-1 lines will be utilized more effectively, resulting in an initial drop in usage and slower growth time. Mail servers will have a longer life cycle due to the decreased demand on their resources. Mailbox disk space will be reduced. LAN bandwidth will be decreased.

4.6.3 Inappropriate material on the network

There are also other problems related to inappropriate material on the network. Users can be offended or embarrassed by unintentionally opening offensive or inappropriate material. The organization can be held legally accountable for this material if it does offend someone. Graphics files can be very large and consume network resources. Graphics also have the potential to contain material that may be very offensive to some people. Blocking these materials at the server is a very important aspect of a Spam filtering system.

4.7 - Project Summary

Before the project was started, the WICHE Spam filtering system consisted of a server side filtering system, on the front end Microsoft Exchange mail server. The software was SurfControl SMTP filter, version 4.7. The system did not reject or block

any email at all. If it flagged email as Spam, it modified the subject to include a ~ possible Spam ~ tag. The client Outlook software was configured with a rule to move messages with this tag into a separate folder. There were many problems with this system. At times, only about one half of the actual Spam was getting flagged as Spam. Another problem was that there was such a large volume of possible Spam being flagged that it made it very difficult to find legitimate messages that were falsely identified as possible Spam. Another major problem was that the client did not have any ability to customize his own filtering. Several products were reviewed and the decision was made to upgrade the SurfControl server side software to version 5.0 and to install a custom made client software to give the users some added control to the filtering system. The upgraded server software added the functionality of rejecting Black Listed email and also flagged possible Spam more efficiently. The client software also added another layer of efficiency for those users that wish to use it. The system is in place, fully operational and blocking Spam very effectively for the time being

Chapter Five: Lessons Learned

5.1 - What was learned from the project experience

Researching different solutions for this project, it was discovered that there are many different methods used for controlling Spam. There are server side, client side and external service provider solutions available. The server side software packages include content filtering by heuristics, Spam dictionaries, html parsers, image scanners and many other methods, as well as white lists and Black Lists. The client side software packages generally allow the user to create white lists and Black Lists, as well as filter based on words in the subject and message body. External service providers will also filter your email for you, charging on a per-user or per-messages basis. By learning more about each of these filtering methods, the decision that was most effective for the organization was chosen.

5.2 - What would we have done differently

Since Spam email is a relatively new phenomenon, there is not very much that we could have done about the problem before we did. As it becomes more and more of an issue for businesses, software providers will build more and more anti-Spam functionality into their software. If there were an email server and client package that had integrated Spam protection, it would be the best option and the easiest to administer. As new products become available, we may discover that one of them is a better solution that the system we have in place. We may have made a mistake by purchasing three years of maintenance and support without knowing this. Only time will tell.

5.3 Did the project meet initial expectations

One of the benefits of monitoring email traffic is that we will be able to analyze exactly how many Spam messages are entering the network each day. It will be important to purchase and develop software that has the capability to monitor and record this. We will also need to create reports to display the information in a useful manner. Recording this will give us a good idea of how successful the new filtering tools will be. As the new software is implemented, we should also be able to see an increase in available network resources. This will be due to the decrease in unauthorized traffic and storage of Spam email messages. Before the implementation is started, we must record a baseline of several resources, such as storage space, internet bandwidth utilization, LAN bandwidth utilization and various server performance counters. This information will be compared to measurements taken after the Spam filtering implementation is completed. Factors such as other ongoing projects and regular network growth should also be taken into account.

5.4 Effects on staff productivity

Computers are predictable and easy to monitor and record statistics on. People, on the other hand are much more unpredictable. It is easy to say people will be more productive because the resources for them to be unproductive are removed, but the reality is usually not that simple. It will be very difficult to measure exactly how productive staff is with the Spam filtering software in place. There may even be a problem with morale that would decrease productivity, if staff is unhappy with new restrictions. It will be very

important to work closely with the users on the network to ensure they are not disgruntled with any changes or new restrictions.

5.5 Best practices to filter without impacting workflow

Any changes to the network that impose restrictions will have an impact on workflow.. There will be tasks that cannot be done the same way they were done before. The important thing is how we deal with those obstacles. It will be important to keep a record of any workflow that was impacted by the changes. It will also be important to implement the changes slowly and run a wide variety of test between each change. The software will need to be restrictive, but must also be flexible enough to allow for exceptions created by unique circumstances. The most important lesson to be learned will be how to work with the end users to make the changes a positive experience instead of a negative one

5.6 - The next stage of evolution for the project if continued

The next stage of evolution would be to further integrate the client software with outlook. The administration of multiple client software packages is more difficult and the operation of two different software packages is more difficult and time consuming for the users. The integration of the client white lists and Black Lists with the server software is also somewhat inefficient. SurfControl is promising a client software package with version 6 of their SMTP filter. This may solve many of these problems, but ideally Spam filtering would be very tightly integrated with the email system on both the server side and the client side.

5.7 - Conclusions/recommendations

At the completion of this project, I would recommend a server side and a client side solution for organizations that have similar business requirements. If the business did not need to make its email addresses publicly available on websites, the system would not need to be as extensive as the system we implemented. I would also recommend that anyone looking for Spam filtering should try to find a packaged system that meets their needs, and is sold with regular updates. A system may work well for a period of time, but spammers are very dedicated and will find ways around systems that work. I also feel strongly that RBL services alone are not sufficient. They do part of the job, but not all of it. Many solutions require that the sender take actions to get their message to pass through the Spam filter. These are convenient, because they take the responsibility away from the receiving party, but are not customer focused. They work well for people who want to be difficult to contact, but may frustrate and discourage potential customers or constituents in customer focused organizations.

5.8 – Summary

Spam is a growing problem. The knowledge I have obtained from this project has been very extensive. It will help me to battle the problem, as well as to share my knowledge with others who are facing similar problems. Each organization has different business requirements and the best solution will be different for each organization. The knowledge I have gained from this project and from the MSCIT program at Regis has helped me to realize this and to be able to ask the right questions when asked for my opinion on Spam filtering. I have met many people who claim to have the “best answer”

for Spam, but I now realize that the best answer for one organization may not be the best answer for another organization.

References

Julian Haight (1998-2005). *Spam statistics* – Retrieved April 10, 2005, Website: http://spamcop.net/spamstats.shtml

The Apache SpamAssassin Project (2005). *SpamAssassin Frequently Asked Questions* – Retrieved April 10, 2005, Website: http://wiki.apache.org/spamassassin/FrequentlyAskedQuestions

SurfControl Plc (2005). *SurfControl Total Filtering Solution* – Retrieved April 10, 2005, Website: http://www.surfcontrol.com/products/total_filtering.aspx

SurfControl Plc (January 6-14, 2003). *Surf Control Web filter sample Reports* – Retrieved April 10, 2005, Website: http://www.surfcontrol.com/general/guides/superscout_web/web_reports.pdf

David E. Sorkin (1999-2004). *United States Federal Spam Laws* – Retrieved April 10, 2005, Website: http://www.spamlaws.com/federal/index.html

David E. Sorkin *Colorado State Spam Laws* – Retrieved April 10, 2005, Website: http://www.spamlaws.com/state/co.html

The office of the Federal Privacy Commissioner (March 30, 2000). *Guidelines on Workplace E-mail, Web Browsing and Privacy* – Retrieved April 10, 2005, Website: http://www.privacy.gov.au/internet/email/

Bloodgate.com (February 17, 2004). *Yearly spam statistics* – Retrieved April 10, 2005, Website: http://bloodgate.com/spams/stats.html#yearly