

TECHNOLOGY WHITE PAPER

Eliminating Spam

*A Long-Term Solution using **Exclaimer Anti-Spam technology***

Author: Gary Levell, Exclaimer Ltd
Ref: 131106



Exclaimer
UK +44 (0) 845 050 2300
USA 1-888-450-9631
info@exclaimer.com

Table of contents

Abstract	2
Introduction	2
What is 'spam'?	3
How is spam typically blocked?	3
What is Bayesian analysis?	4
Dealing with false positives	6
Exclaimer's anti-spam solution	7
Identifying the Client IP Address	8
Spam classification at protocol time or transport time?	12
Installing Exclaimer on a boundary server	12
Installing Exclaimer on an internal server	13
Scanning the Internet Header for the Client IP	13
Exclaimer's anti-spam individual tests	15
1. Auto Whitelist	15
2. Trusted/Blocked IPs	15
3. Spoofed domains/IPs	15
4. SPF tests	16
5. DNS black & white lists	17
6. DNS RHS black & white lists	17
7. Local black & white lists	17
8. Detection centre fingerprint test	18
Exclaimer's anti-spam deployment strategy	20
Evaluation Deployment	20
Final Deployment	22
Conclusion	23
Glossary	24
References	26

▶ **Abstract**

This paper attempts to highlight some of the problems facing Exchange server administrators in dealing with spam and spam attacks. It introduces how Exclaimer Anti-Spam can help eliminate spam by dynamically filtering incoming mail.

▶ **Introduction**

As an Exchange Server Administrator managing multiple mail accounts, dealing with spam can be a life altering experience. Now, as much as I appreciate these spammers concerns over my health, size (ahem!) and my soon to be deactivated bank account, I have to say the frequency at which I (and many other people) receive these types of email message can be a little intense to put it mildly.

As I'm sure you know, the problem is that spam comes in many different shapes, sizes and disguises, and what is spam to one person may not be to another. This combined with ability that spammers have to constantly find new ways of bypassing spam filters, poses a whole series of problems for the developers of Anti-Spam applications.

This document looks to address the core issues surrounding spam and anti-spam, and how best to manage them within an Exchange server environment.

▶ **What is 'spam'?**

In simple terms 'spam' is unwanted email. This may be a crude definition but it accurately reflects what many people mean by 'spam'.

The fact of the matter is, is that the term 'spam' is used to refer to a host of different types of unwanted email. This can include email that people could find personally offensive, email that is unexpected or contains strange characters, and email that appears to have come from someone they know but didn't.

▶ **How is spam typically blocked?**

Many anti-spam solutions use a technique called content analysis to block unwanted emails.

This technique attempts to analyze the content or characteristics of the message that make it look like spam.

For instance, some spam emails use all uppercase letters (which to me just looks like shouting), or large bold red letters, or (in the case of HTML email) include various benign HTML tags in words (which are designed to confuse the content analysis). Some spam comes with the message in a picture and a short extract of Shakespeare to enlighten my day.

Other tests look at the formation of the message or the format in which the message is sent e.g. a message received only in HTML is often, but not always, spam. At Exclaimer we believe that content analysis is a flawed approach to identifying spam. It contributes to the cyclical nature of the prevention of spam, and also leads to false positives (legitimate email being wrongly classified as spam).

Each time a new trend in spam is spotted by the anti-spam community, a software product or patch is released to deal with it. The content analysis tests are modified to spot the new trend. However, as soon as the spammers find their email being blocked they simply adapt their approach, and so the cycle continues.

This is not all bad. It has, after all, helped to make it harder for the spammers to deliver email. Still, since it costs them very little to continue trying, things have now reached fever pitch. The anti-spam community now talks in terms of 100's of millions of messages in a run. This is often achieved with an army of compromised Trojan machines that unleash a torrent of emails in the hope that, by the time someone complains about their host or web

site, they will have delivered the message to enough people to make their campaign effective.

This is despite the fact that we have never met the proud owner of a “genuine” Rolex watch purchased online; nor anyone who has bought shares in the next best hot stock, or purchased dubious pharmaceuticals, has increased the size of parts of their anatomy, or has actually contacted the foreign national in whose sticky mitts resides millions of dollars that they are prepared to share with you just because you are... well, you!

According to Paul Graham¹, a spam campaign only requires a response rate of 15 in a million to be judged successful. Yes, you read that correctly, only 15 idiots in a million are needed to sustain the business of a spammer and ensure the rest of us suffer from inbox graffiti, reduced internet performance, Trojan viruses etc.

¹ Graham, Paul – A plan for spam
(<http://www.paulgraham.com/spam.html>)

What is Bayesian analysis?

Bayesian analysis is a more complex form of content analysis that automatically assigns a score to each word or phrase in a message based on previous scrutiny.

How does it work?

The theory goes that you take a representative sample of, for example, 1000 email messages and classify them into spam and not-spam (sometimes called ham!). A training engine then analyses this block of classified messages (known as the corpus). It breaks each message apart and scores each word or phrase based on its presence in a spam or ham message.

So you would expect the word “Viagra” to have a high spam score and a low ham (not-spam) score. With a word like “and” you would expect it to have roughly equal spam and ham scores, or the word “brickwork” to have a low spam score and a high ham score.

Once you’ve trained your system in this way, each incoming message is classified as either spam or ham by applying a complex statistical calculation based on the results of the Bayesian analysis.

It’s important to note that the cornerstone of a Bayesian system is its training! The good news is that this method is very effective if your database is well trained with both recent spam and ham messages. In fact it can be so good that this can tend to cloud peoples’ judgment on the disadvantages of this technique.

As a classification tool, Bayesian analysis is very effective. However, the spammer is also aware of the Bayesian analysis techniques and, quite simply, aims to get his

message classified as ham. (He is not at all happy with it being classified as spam!)

He fills his message with a variety of distracting information, perhaps snippets of Shakespeare or an excerpt from a recent news feed. This distorts the Bayesian analysis result. Although the message clearly contains the words "Viagra" and "cheap" which increase the spam score, it also includes the words "theretofore" and "enunciate" which the analyzer has probably not been trained to recognize. The **unrecognized words "water down" the effect of the condemned words in the Bayesian score and the message is not clearly identified as spam.** This may be enough to allow the message to be delivered. Success as far as the spammer is concerned.

You need to keep on top of your Bayesian analysis training to maintain its effectiveness – something that most busy IT people don't have time to do.

This problem is exacerbated in a corporate environment where the analysis is already diluted by the presence of a huge number of different words and phrases within ham messages. This doesn't help with scalability and low maintenance, especially if your company makes Viagra!

Analyzing the content of emails to identify and block spam messages simply doesn't work in the long term. **Exclaimer Anti-Spam** applies a series of different tests to identify spam without needing to scan the content. This **provides a more effective, language independent, long-term and low maintenance solution.** We'll look in detail at these tests later in this paper.

▶ Dealing with false positives

There is no sense in denying it. If you block email, you will at some point come across a false positive i.e. a legitimate email that has been incorrectly classified as spam and is rejected by your system. This is the subject of much anecdotal information on the Internet and in various other anti-spam products' documentation.

Claims to have zero false positives abound – Rubbish!

At Exclaimer we accept that false positives are a reality. Although we try extraordinarily hard to reduce their numbers to as close to zero as possible (can you sense our obsession with this?), there is, and always will be, the chance that they will occur. Once you admit this, you can deal with it professionally, which is the approach we take.

By ensuring that the proper Internet protocols are adhered to, we can rely on an already impressive infrastructure to politely inform the sender that an email was not delivered correctly. This allows the sender to take the appropriate action. They can call you, send a fax, or perhaps refer to your website to investigate why the mail was blocked and try to correct the problem.

We appreciate that the sender may find this irritating. But, on reflection, how much more irritating (and perhaps damaging) would it be if they were left with the impression that their email had mysteriously disappeared, been delayed, ignored, or deleted. Denying the reality that it was blocked is not a professional way to deal with the problem.

Our approach of informing the sender is also the sensible way to deal with mailing lists. Most of the professional list service providers have already set their systems up to deal with rejected and bounced messages. They have a variety of strategies for dealing with this, ranging from message hold & re-try, through to account suspension.

This responsible approach is wasted if their messages are simply blocked and deleted without notification.

▶ Exclaimer's anti-spam solution

Exclaimer takes a different approach to eliminating spam.

It does not believe the answer lies in analyzing each email's content. Instead it uses a variety of tests to determine whether a message is spam or genuine. For example:

- check for spoofed information.
- reference DNS lists to check for known spam sources.
- match the fingerprint of the message against recognized spam outbreaks identified by our detection centre.

To perform many of its tests, Exclaimer needs access to some simple information from each incoming message². In particular, Exclaimer needs to be able to accurately determine the IP (Internet Protocol) address of the client MTA (Message Transfer Agent) that is attempting to submit a message into the queue for delivery. This information is at the root of many of Exclaimer's anti-spam tests and is fundamental to the successful classification of messages as spam.

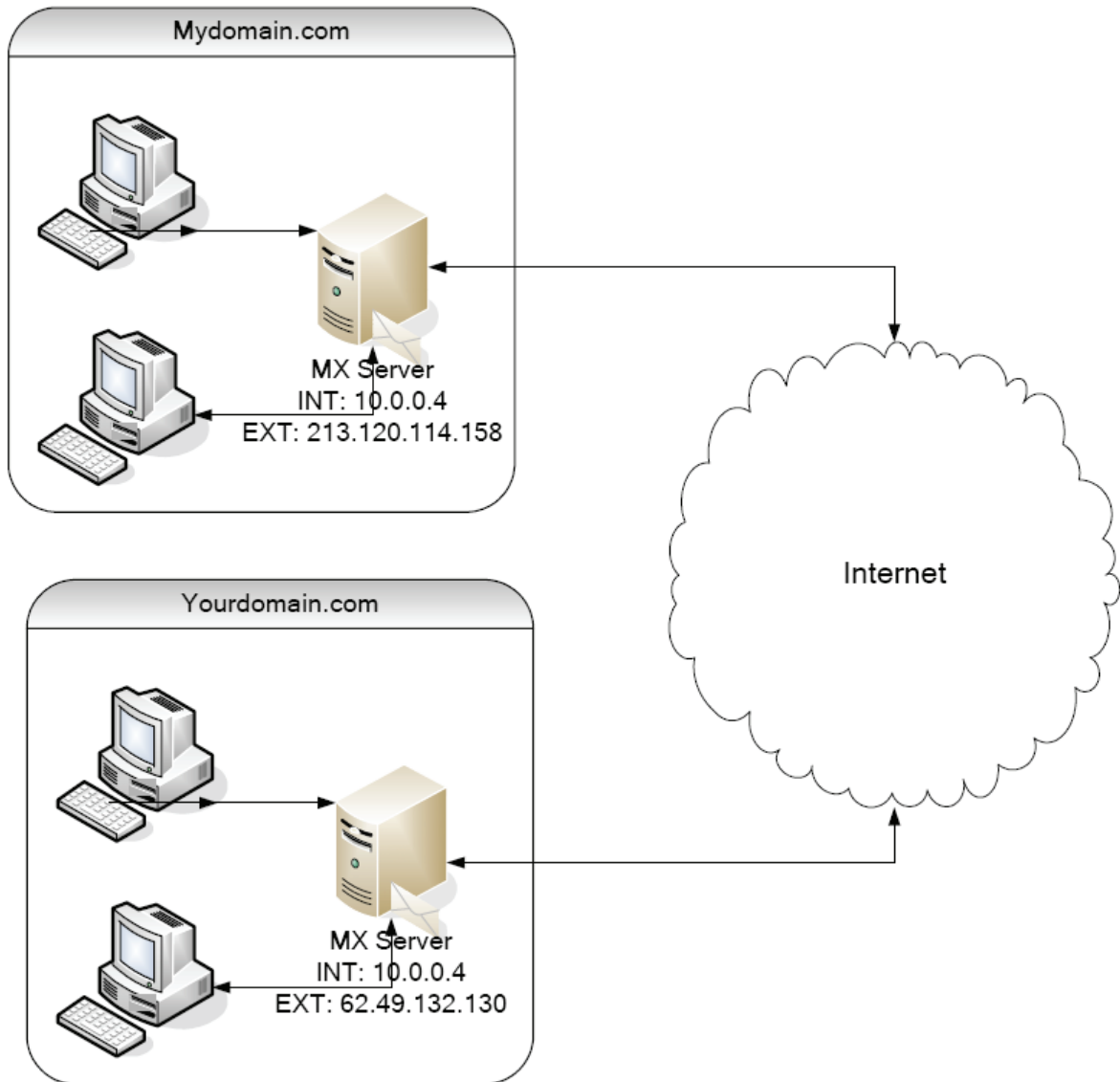
If you are not familiar with the acronyms used in the last paragraph, or perhaps didn't understand it at all, now might be a good time to find out what it means. We are going to be referring to them (and other concepts) later, which will require you to at least understand what they are.

² If this information is incorrectly determined, perhaps because of a configuration error, the test results may become ineffective. So, it's important to the success of your anti-spam solution that you configure the system correctly.

Identifying the Client IP Address

In a very straightforward environment, the client IP address is available directly to Exclaimer from the TCP/IP conversation with the client MTA. Simply put, this information is available as a direct consequence of the client opening a connection to the SMTP server (normally on TCP port 25).

FIGURE 1



In a more typical environment it is not usually this easy to identify the connecting IP address. For example:

- a backup mail server may be specified in the DNS data for your domain. This server allows mail to be delivered if your local system is unavailable for any reason. It follows that when Exclaimer is talking to the backup (or secondary) MX (mail-exchanger) it cannot use the connecting IP address because this is the address of the backup server, not that of the original MTA (and possible spammer).
- there may also be other trusted servers, or a DMZ, or simply a corporate mail gateway that is not in the local domain.
- you may have a front end service that receives your email and pushes it back to you after some cleansing or analysis campagne.

So, to find the connecting IP address, Exclaimer establishes a network of trust that contains the IP addresses of servers that are known to be under its control or influence. Armed with this information, Exclaimer can then determine the connecting address of the real MTA by carefully and judiciously analyzing the message headers.

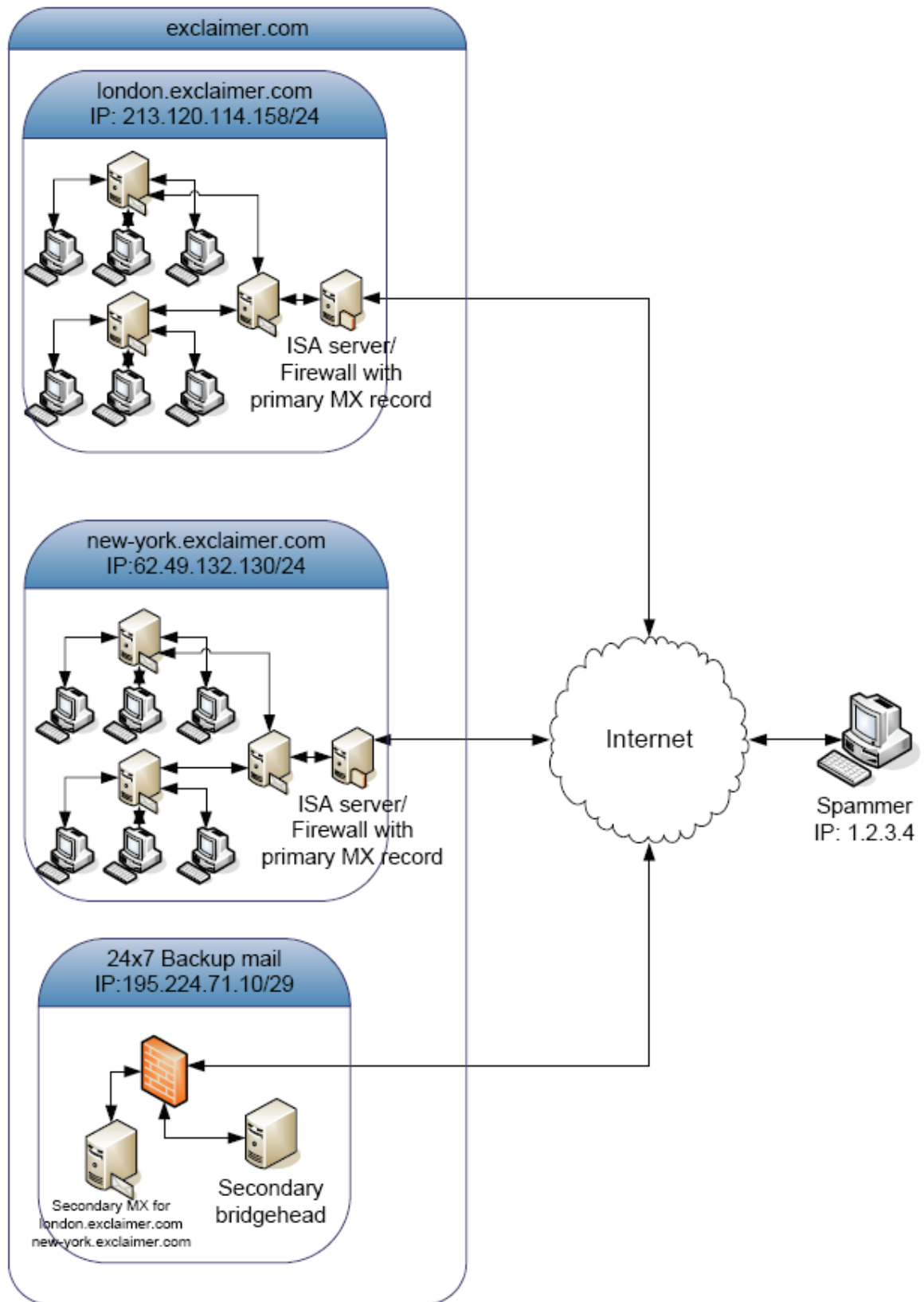
This list of trusted IP addresses is formed from the following sources, as outlined below:

Source of trusted IP addresses	How the IP addresses are identified
The private network range (RFC 1918)	These are hard coded in the product
IP sub-net of all IP addresses on the network adaptors installed in the system on which Exclaimer is installed	Generated by querying the Windows operating system to request the IP addresses bound to the network adaptors installed in the server.
IP addresses listed against the MX servers for domains that you have indicated that you are authoritative for	The MX servers are determined by making a DNS query for the domains that you are authoritative for. These can be altered using the control panel applet.
IP addresses listed as your trusted relay servers	The trusted relay servers are loaded from the Exclaimer control file and can be changed using the control panel applet.

Using the example in the following diagram, this might result in the following IP's being trusted:

- 127.0.0.0/8 - Local host
- 10.0.0.0/8 - RFC 1918 Class-A private network
- 172.16.0.0/12 - RFC 1918 Class-B private network
- 192.168.0.0/16 - RFC 1918 Class-C private network
- 213.120.114.158/32 - Primary MX for london.exclaimer.com
- 62.49.132.130/32 - Primary MX for new-york.exclaimer.com
- 195.224.71.10/32 - Secondary MX for exclaimer.com

FIGURE 2



Spam classification at protocol time or transport time?

Exclaimer's anti-spam is best installed at the infrastructure boundary, i.e. a system that has a direct connection to the Internet and that is also an MX server. This allows you to get the best bandwidth reductions and best overall end-user experience.

Ideally Exclaimer's Anti-Spam solution should be installed on all MX servers. However, we recognize that this is not always possible, particularly if you rely on an ISP to provide backup/secondary MX servers.

Installing Exclaimer on a boundary server

When Exclaimer is installed on a boundary server, a connection with the client MTA can be rejected with an RFC compliant message during the protocol (RFC281/2821) stage.

If the connecting MTA was a spammer, this puts the onus of dealing with a broken connection back onto them. It also allows upstream providers to see that the connection with the remote client has been dropped and, provided that the data is not already "on the wire", they can block the packets and save the bandwidth.

Alternatively, if the connecting MTA was a legitimate mail server, then rejecting the message issues an NDR (non-delivery report) to the sender. We feel that taking this positive step to inform the sender is better than the approach taken by other Anti-Spam solutions, which is to simply bin the message or move it to a junk mail folder. When this happens, the sender, who may be an important customer, friend or acquaintance, is left waiting for a reply to an email that no longer exists!

Even though many users ignore NDRs or at most contact their IT support with the typical "my email is not working again" message, this is still the most constructive way to reject an email. At least the sender is made aware that the email was not delivered and can make contact another way if necessary (after all telephones do still exist and almost always work).

Installing Exclaimer on an internal server

If Exclaimer is installed on an internal server, the message must be delivered in its entirety so that the headers can be interrogated to locate the boundary server and its connecting IP address.

Once this information has been identified, the anti-spam tests are carried out to classify the message. If it is classified as potential spam, the connection with the client MTA can be terminated. In this scenario, it falls to that server to issue a bounce to the remote address and this is where some problems need to be addressed.

If the ultimate connecting server was a legitimate mail server, then in all likelihood the NDR message will be delivered to the sender without any problem, in a timely fashion. Then, as in the boundary installation scenario, the sender knows that his email was not delivered, which as we have established, is the most desirable outcome. If a spammer has hijacked the domain of some poor unsuspecting soul to make the connection, that person will receive the NDR message and typically their mailbox will overflow with these rejection messages.

In many cases, the sender's email address is fictional and when your server attempts to deliver the NDR message, the remote server rejects it. What does your server do with this undeliverable NDR? It normally remains in your outbound mail queue while the server repeatedly tries to deliver it, until it finally drops it into your bad mail directory. (Have a look. You may find thousands of messages in there!)

Scanning the Internet Header for the Client IP

Exclaimer's anti-spam tests are generally performed at the earliest possible opportunity. If Exclaimer is installed on a boundary server, then many of these tests are performed during the SMTP protocol phase of the message transfer. However, if Exclaimer is installed on an internal server, or is receiving messages from a secondary MX server, these tests can only be performed after the entire message has been received.

If Exclaimer is installed on an internal server it determines the client IP by scanning the Internet headers of the message. It searches the Received: header lines and identifies the first IP address that is not in the list of trusted or internal IPs.

Using the following message header, we can see this in operation:

```
1. Microsoft Mail Internet Headers Version 2.0
2. Received: from mail1.exclaimer.com ([10.0.0.2]) by member1.dev.exclaimer.local with Microsoft SMTPSVC(6.0.3790.0); Wed, 8 Dec 2004 00:22:49 +0000
3. thread-index: ActCvAxR6yI5Jn4EQlQm2VoifQU6nA==
4. Content-Transfer-Encoding: quoted-printable
5. Received: from mail2.exclaimer.com ([195.224.71.32]) by mail1.exclaimer.com with Microsoft SMTPSVC(5.0.2195.6713); Wed, 8 Dec 2004 00:22:48 +0000
6. Content-Class: urn:content-classes:message
7. Priority: normal
8. X-MimeOLE: Produced By Microsoft MimeOLE V6.00.3790.181
9. Importance: normal
10. Received: from ynmmail.com ([217.216.204.232]) by mail2.exclaimer.com with Microsoft SMTPSVC(6.0.3790.0); Wed, 8 Dec 2004 00:27:45 +0000
11. Received: from hotmail.com ([65.54.190.7]) by ynmmail.com; Wed, 8 Dec 2004 00:17:45 +0000
12. Received: from yahoo.com ([206.190.36.244]) by hotmail.com; Wed, 8 Dec 2004 00:07:45 +0000
13. From: <*****@ynmmail.com>
14. To: <lygby2317964a@exclaimer.com>
15. Subject: Cheapest meds you'll find.
16. Date: Wed, 8 Dec 2004 02:19:20 -0500
17. MIME-Version: 1.0
18. Content-Type: text/plain;
19. charset="us-ascii"
20. Return-Path: <*****@ynmmail.com>
21. Message-ID: <MAIL2sJ3Ih7jtrSByh100002009@mail2.exclaimer.com>
22. X-OriginalArrivalTime: 08 Dec 2004 00:27:54.0734 (UTC) FILETIME=[C270A0E0:01C4DCBC]
```

This message was received on server "member1.dev.exclaimer.local" which is an internal server.

Exclaimer then scanned each Received line extracting the IP addresses:

These are 10.0.0.2, which is an internal address, 195.224.71.32, which is our secondary MX, and finally 217.216.204.232, which is the first alien system and must therefore be the Client IP.

The other received header lines (11 & 12) are both irrelevant (and often spoofed) lines. We've already found the first server that we know nothing about (and therefore don't trust) and that is good enough for us.

Armed with this information, Exclaimer can then proceed to analyze the message.

▶ **Exclaimer's anti-spam individual tests**

Exclaimer applies a collection of tests and techniques to identify spam messages. These tests can be broadly broken down into the following eight categories.

1. Auto Whitelist

This automatically whitelists any email addresses that you send mail to. After all it would only seem logical to expect that if you were sending an email to someone you would want them to be able to send you back a reply.

Auto Whitelisting works by maintaining a list of correspondents you send mail to. Should the list become too large the oldest entries will be removed first. This ensures that those clients you correspond with most regularly with are automatically whitelisted and therefore won't have their email bounced back as spam.

This technique can be especially useful if you receive email from a large client base where manual whitelisting would be impractical and difficult to manage.

2. Trusted/Blocked IPs

A trusted IP can be used to whitelist an individual server or range of servers that you are confident do not send you spam. You must be careful using this feature if you are not in control of these servers because if the servers are compromised you may be inundated with spam.

A blocked IP can be used if you are getting persistent spam from a single source that is not being caught by the other tests. This might be because the spam is not being directed at other people (so the detection centre/DNS lists don't notice it). It could be a malicious sender, or someone that you particularly don't want to communicate with. Again, we advise caution using a blacklisting strategy like this as the IP address will never be permitted to send email to you and this may be a little too aggressive.

3. Spoofed domains/IPs

These two tests are really only present because they offer a potential performance improvement. All the email that would be permitted to continue if these tests were not present would be blocked by other tests. Nevertheless, this is such a massive opportunity to throw these emails away without any further interrogation that we could not resist! Basically, during the SMTP protocol command HELO (or EHLO), the sending server is supposed to provide an FQDN

and can also supply an IP address. So the command may look something like:

```
HELO customer.com  
HELO 1.2.3.4
```

For historic reasons, many pieces of spam and virus software will use your domain or IP address in this command to try to spoof you into thinking that it is a legitimate or internal email message.

However, since no correctly configured MTA would ever do this, it's a very simple to just reject a message if it uses your IP address or any domain that you have authority over.

4. SPF tests

SPF is a technique used to ensure that the sender of a message is authorized to send messages on behalf of a domain.

The DNS system lists mail exchanger records (MX) for inbound messages to a domain but historically it has not listed outgoing mail senders. SPF is an attempt to redress this balance (see spf.pobox.com for details of this). A recently formed council is attempting to take this technology to RFC status.

A domain that publishes an SPF policy is making a statement concerning their outgoing mail servers. This allows receivers (such as Exclaimer) to enforce this policy.

A domain policy may look like one of the following:

```
v=spf1 mx -all  
v=spf1 ip4:195.224.71.10 mx ?all  
v=spf1 a mx ip4:200.100.50.0/24 ~all
```

The first example states that email from the MX server is permitted but email received from all other servers should be treated as failures.

The second examples states that email from IP address 195.224.71.10 and email from the MX server is legitimate, but that it is unsure about emails received from other servers.

The final example states that email from any DNS listed host, the MX server, and the Class-C network starting at IP address 200.100.50.0 is permitted, but emails received from other servers should be treated with caution. The SPF documentation terms the '~' modifier as soft-fail.

It is recommended that messages which fail the stated SPF policy should be rejected outright and that any messages that soft fail should be marked as such and delivered. Interestingly, an SPF pass does not necessarily mean that the message is not spam. It only means that the domain

owner has authorized the IP address of the sender (which in many cases is a spammer!).

We encourage you as a domain owner to publish your own SPF policies, as this will help others who enforce SPF policy from fallout of spoofs of your domain.

SPF and Microsoft's SenderID are similar technologies as they protect different parts of the message. SPF protects the SMTP protocol MAIL FROM: whereas SenderID protects the PRA (Purported Responsible Address). This is commonly the From: address, but can be other pieces of information in certain circumstances.

Exclaimer will support SenderID in a future release when the protocol has become a little more solid.

5. DNS black & white lists

Many DNS based lists exist that contain the IP addresses of known spam sources, and these have varying rates of success and aggressiveness. After a long period of analysis, we recommend Spamhaus; a UK based organization that we believe takes a professional and sensible approach. See www.spamhaus.org for details.

Exclaimer is pre-configured to use this service. If you receive millions of messages a day, you might want to consider a domain transfer from them to deal with Internet bandwidth issues. Please contact spamhaus for details.

If you have other preferred providers, you can utilize them using this feature.

Exclaimer uses the Client IP and performs a PTR type query against the listing host.

6. DNS RHS black & white lists

RHS lists are those that list the right-hand-side of email addresses, or to put it simply, the domain.

Other than this difference, they operate in the same way as the DNS black & white lists.

Exclaimer uses the domain of the MAIL FROM: command to validate against this. Some lists prefer this test to be done against the From: field. Exclaimer does not offer this capability at present.

7. Local black & white lists

The local black and local white lists test the information in an incoming email against the entries in each list. In order to do this, the entire message must have been received. So, these tests can only be performed if the message is

received in its entirety i.e. after the DATA phase of the SMTP protocol.

If Exclaimer is installed on a boundary server some spam emails will be rejected before this stage is reached. Therefore, it is not possible to whitelist emails that trigger these protocol based tests as the message has already been rejected.

However, local black and white lists are still used by Exclaimer.

The whitelist is particularly useful for testing messages that get marked by the anti-spam tests as BULK. These may be messages received from an online store or a mailing list, or someone who sends hundreds of emails a day to you. The whitelist can be used to accept selected bulk emails. The Exclaimer whitelist wizard provides a very simple "query by example" approach to finding the common elements of the acceptable messages and building a whitelist from this.

The blacklist may be suitable for getting rid of messages that are persistent in nature but which don't trigger other anti-spam tests. For example if an individual is being targeted with unwanted persistent email from a source that is not being identified as a spammer.

8. Detection centre fingerprint test

Exclaimer's detection centre is the last line of defense against messages that have not been trapped by the preceding anti-spam techniques. At this stage our tests will have already removed some 75% of the spam messages. However, to achieve the 96% or so that we need to be an effective solution, we use the detection centre.

Exclaimer scans the incoming messages and generates a DNA style fingerprint from the message. This fingerprint is passed to our detection centre and if it is recognized, the message is classified as spam accordingly.

The detection centre is constantly on the lookout for an outbreak of spam. It often recognizes an outbreak by the exceptional level of activity a spam outbreak generates. As we have mentioned earlier, the spammer's business model relies on getting a response rate of at least 15 in 1,000,000. To achieve this, the spammer usually operates a run of 200-300 million messages. Since they have hosted their website on a commercial service, the ISP will receive a barrage of complaints from people who receive spam email advertising the site. Often these sites are only up and running for 24-48 hours.

This means that the spammers need to get the 200-300 million messages out in pretty short order and it is precisely this level of activity that condemns them, not the fact that they are selling Viagra or cheap Rolexes. Heaven knows, we don't have a problem with either product; we

just have a problem with 10,000 emails telling us about them!

It is this level of activity that is identified by our detection centre and used to recognize the spam outbreak. This is a far more flexible and robust method of detecting spam than scanning the message content for the latest hot topic. It is a reliable and robust method of detection that is not language or content dependent.

Note: We recommended earlier that the rejection message that is applied to the spam messages blocked by the detection centre provide the sender with a method of bypassing this block. This ensures that legitimate messages that have been blocked can be handled in a coherent fashion.

Proxy server settings

Exclaimer's real-time detection centre stores the fingerprints³ of spam messages and monitors the outbreak of spam runs. Exclaimer communicates with this service via HTTP.

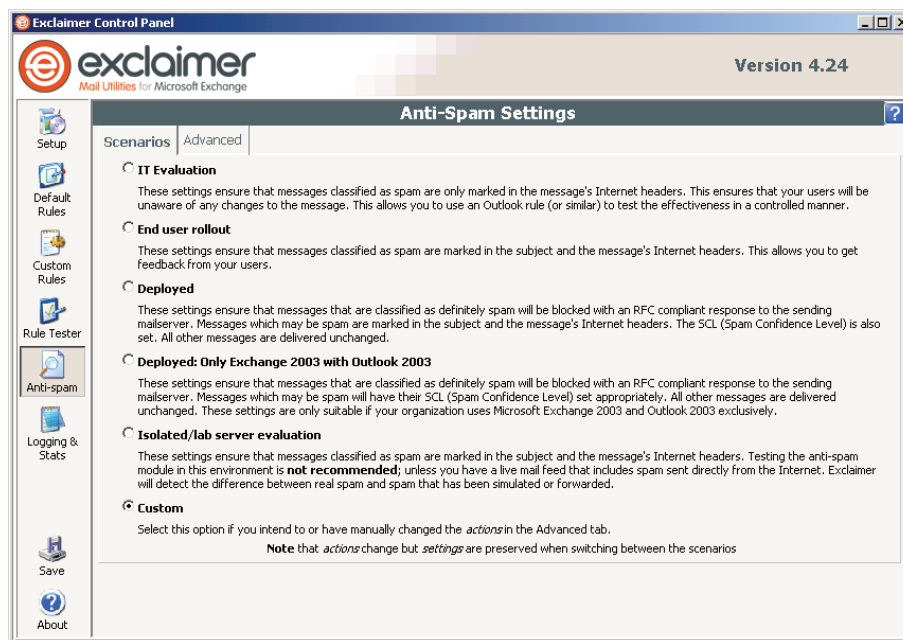
If you talk to the Internet using a proxy server, you will need to configure this information so that Exclaimer can talk to the detection centre.

If Exclaimer cannot talk to the detection centre, then the effectiveness of the anti-spam tests will be compromised and you won't get the same rate of classification.

³ At no time is any confidential information passed to our detection centre. The fingerprint is computed on your server and it is this that is sent to our detection centre.

▶ Exclaimer's anti-spam deployment strategy

Exclaimer recognizes that the implementation of a new anti-spam product requires careful evaluation and deployment. To this end, Exclaimer anti-spam includes a powerful deployment scenario wizard that assists you in this deployment process.



Evaluation Deployment

Initially you will probably want to install Exclaimer in the most unobtrusive fashion possible during your evaluation phase. This is provided by the IT Evaluation scenario option.

This scenario option allows you to evaluate Exclaimer's anti-spam usefulness without affecting the majority of your users (which means you can avoid having to deal with lots of questions as to why messages are being marked in the subject as "[SPAM]").

You select a number of key users (yourself included if you suffer from a big spam problem) to trial the anti-spam features.

You must create a custom rule in Outlook on the appropriate computers to move messages with the Internet header "X-Exclaimer-Spam" to another folder for interrogation later.

End User Rollout

Following a successful evaluation, the next step is the End user rollout scenario phase.

In this scenario Exclaimer identifies messages that it has classified as spam by adding "[SPAM]" to the message subject but still continues to deliver the messages. This allows your users to quickly identify possible spam and track how effectively these messages are being caught. Your users must be involved in this process. They need to be aware that you are implementing a new anti-spam solution because they need to check that their messages are being classified correctly.

Have any spam messages escaped detection? Have any legitimate messages been classified as spam i.e. the dreaded false positives? They need to notify you in either case, in a timely fashion, so that the reason the messages have been wrongly classified can be determined and corrected.

Exclaimer marks the message header to let you know why a message was condemned. This may be because the message has been classified as spam by one of the tests. Alternatively, it may have been condemned because it was sent from a mailing list, in which case [BULK] is added to the message header.

Using a whitelist wizard to deal with mailing lists
Emails received from mailing lists are not always spam. Exclaimer uses a whitelist to solve the problem of identifying those that are not spam.

Whitelists are not new but have often become hard work, difficult to maintain and often unwieldy. Many other anti-spam products operate an automatic whitelist policy that works by whitelisting anyone you've ever emailed. Whilst this may suit their particular brand of anti-spam, these one-to-one messages are rarely categorized as spam.

Exclaimer maintains a message tracking log that you can interrogate to find the best way to whitelist a particular type of message. It also provides a whitelist wizard that interrogates the message log and helps you to maintain an accurate, targeted whitelist.

As an example, assume one of your users subscribes to a car list and that list always sends messages with the subject "[car list]". You can enter this identifying information into the whitelist wizard. The wizard searches the message tracking log and lists any messages that meet this criterion. You then select several of these messages that are representative of the list. The wizard analyses these messages and identifies their common characteristics which it then displays. If these details are correct, you can save them as part of the whitelist.

Final Deployment

Finally, once you are happy that your messages are being classified correctly, you can move to "Deployed" mode of operation.

In this mode, messages that are classified as spam are rejected or cause an NDR message to be generated. This is the safest and, contrary to most people's initial reaction, the best way to deal with spam because the originator of the message gets an immediate notification that his email wasn't delivered and he can act upon this information.

The other scenarios deliver the message but may place it into a public folder or private folder where it may never be read.

We suggest that you update the default rejection message assigned by the "Detection Centre - Spam" test to say something like:

```
SPAM email not accepted. Please include the code  
"333373" in the subject to bypass this anti-spam  
test.
```

We have used Exclaimer's telephone number in this example, but you can use any suitable text. You need to whitelist any message that contains the suggested text in the subject and you can use the whitelist wizard to do this.

Incidentally, we've never had a message erroneously blocked by the detection centre so that someone needed to use this technique, but we feel it better to be safe than sorry.

▶ Conclusion

Exclaimer's Anti-Spam feature uses just a handful of very effective techniques to ensure that the maximum quantity of spam is trapped, whilst at the same time minimizing false positives.

Exclaimer does not use techniques that other anti-spam vendors use. Content analysis, along with Bayesian filters and content heuristics, all have transient success. Whilst they seem to be working well to start with, after a while, the spamming community adapts and learns to bypass the filters (often getting more mail through than before the filter was installed).

Exclaimer does not use these methods, and does not check the content of the message. This allows Exclaimer to be language independent and require less ongoing maintenance and administration.

Exclaimer's front line of defense is a collection of tests that stop huge swaths of spam, viruses, Trojans and the like by simply refusing to communicate with them. Despite their attempts to have an unhindered conversation with your mail server, they can't help identifying their nature and they are immediately blocked.

The next line of defense comes from checking that the incoming mail is not coming from a known spam source, or from a machine that has been compromised by being infected with a Trojan or other spamming tool.

Exclaimer has a dedicated detection centre that provides the last line of defense. After all other tests have been performed directly on your mail system, the email's unique fingerprint is passed to the detection centre to see if similar messages have been delivered elsewhere. If they have, then they may be spam and the message is subjected to closer scrutiny, which may result in it being classified as spam.

Where is it from and where is it sending you?
This question is at the heart of Exclaimer's Anti-Spam technology. We are simply concerned with where the message came from and where it's trying to send you. If it came from a known spam source, then it's condemned. If it's sending you to a website that has been advertised in a spam email (spamvertised), then it's condemned too. This seems like a simplistic approach. And it is.

Simple solutions are always the best.

▶ Glossary

Blacklist

A list of email addresses or IP addresses that individuals and organizations can use to filter out unwanted emails.

Class-C network

A type of network where the first 3 bits, or the *high-order* bits, are always `110.` The next 21 bits are used to define the Class-C network, and the final eight bits are used to identify the host.

DMZ

De-militarized Zone – the area between a domain's firewall and the outside world.

DNS

Domain Name Service - used on Internet for translating hostnames into Internet addresses. For example, <http://www.exclaimer.com>

EHLO

Extended HELLO – used to open a transmission with ESMTP clients.

FQDN

Fully Qualified Domain Name – an unambiguous domain name that specifies the node's position in the DNS tree hierarchy absolutely. For example, `somehost.example.com`.

HELO

HELLO - used to open a transmission with ESMTP clients.

HTML

HyperText Markup Language – used predominantly for web page creation.

HTTP

HyperText Transfer Protocol – used to transfer information on the World Wide Web.

IP

Internet Protocol - a unique number that devices use to identify and communicate with each other on a computer network.

ISA server

Microsoft Internet Security and Acceleration Server – a server which runs on Windows 2000 and Windows Server 2003.

ISP

Internet Service Provider - a business or organization that provides access to the Internet and other Internet related services.

MTA

Mail Transfer Agent – a Mail Server or Mail Exchange Server.

MX

Mail Exchange record - a type of resource record by the Domain Name Service (DNS) specifying how Internet email should be routed.

NDR

Non-Delivery Report - a message tells the sender that their message could not be delivered.

PRA

Purported Responsible Address - usually the From: address, but can be other pieces of information in certain circumstances.

PTR

Pointer Record – used to map an IPv4 address to the canonical name for that host.

RFC 1918

A published paper that specifies Internet Best Current Practices for the Internet Community.

RFC 281

A published paper suggesting an addition to File Transfer Protocol.

RFC 2821

A published paper that specifies Internet standards track protocol for the Internet community.

RHS list

Right Hand Side – a list of email address domains.

SMTP

Simple Mail Transfer Protocol – used as the standard protocol for email transmissions across the Internet.

SPF

Shortest Path First - a routing method able to eliminate loops.

TCP/IP

Transmission Control Protocol/Internet Protocol – also known as the Internet protocol suite. A set of communications protocols that implement the protocol stack on which the Internet and most commercial networks run.

Trojan machine

A computer that has been infected and taken over by a Trojan computer virus. Often used by spammers to send unsolicited email.

Whitelist

A list of email addresses or IP addresses that individuals and organizations use to identify senders they wish to receive email from.

 **References**

Graham, Paul – A plan for spam
(<http://www.paulgraham.com/spam.html>)

RFC 1918 - Address Allocation for Private Internets

RFC 821 - Simple Mail Transfer Protocol

RFC 822 - Standard for the format of ARPA Internet text messages

RFC 2821 - Simple Mail Transfer Protocol

RFC 2822 - Internet Message Format

SPF - Sender Policy Framework –Meng Weng Wong et al.
(<http://spf.pobox.com>)