Stopping Outgoing Spam by Examining Incoming Server Logs

Richard Clayton
Computer Laboratory, University of Cambridge,
15 JJ Thomson Avenue, Cambridge, CB3 0FD, U.K.

Abstract

Processing server logs for the email arriving at an ISP can be used to detect remote sites where machines are infected by email viruses or have been hijacked and used for sending spam. Simple heuristics distinguish the patterns of such traffic from those of legitimate email. Stopping this material being sent is matter for the remote site. Nevertheless, this paper shows that processing can also detect if any of the ISP's own customers have problems, because their email is logged when it is sent to other customers (or even back to themselves). Experimental results from a medium-sized ISP show that the scheme is successful in detecting customer problems. Unfortunately, if the spam or virus is not sent to anyone local then the problem remains undetected. Estimates of worldwide rates of compromise of end-user machines are used to give an indication of the likely overall effectiveness of the detection scheme.

1 Introduction

Previous work, dubbed "Extrusion Detection" (Clayton, 2004), showed how automated processing of email server logs for *outgoing* email was an extremely effective method for an Internet Service Provider (ISP) to detect when customers were sending unsolicited bulk email (spam) or were infected with virus/worm malware that spreads via email. Many ISPs provide email "smarthosts" for customers so that relatively simple SMTP clients on end-user machines are spared the complexity of arranging for delivery to remote sites. Some straightforward heuristics for smarthost log processing, concentrating on delivery failures, can pick out customers whose systems are compromised and are being used to relay spam or to send out email viruses.

Quite obviously, a system based on outgoing email logs remains entirely unaware of email that, by accident or design, is sent directly to its destination, because it bypasses the smarthost. Unfortunately, this means that many recent virus outbreaks and modern spam sending engines fail to be detected. However, this email may be being sent not only to remote sites elsewhere on the Internet, but also to other customers of the same ISP, or even (because the sending programs are not very sophisticated) back to the sender themselves. Because ISPs provide "store and forward" services for incoming email, delivering incoming email into POP3 or IMAP mailboxes, the email that is being sent to the ISP customers will arrive at the ISP's incoming mail servers, the "MX hosts". This provides an opportunity to process the traffic logs from these servers as well and thereby detect the underlying problem.

In this paper we consider what heuristics can be applied to the processing of the logs for an ISP's *incoming* email server. We present some encouraging initial results to show that this is a successful method of detecting problems with the ISP's own customers as well as giving a wealth of evidence about problems elsewhere on the Internet. This leads to some estimates of the scale of the global problem and hence indicates, for the ISP being studied, how many problem machines are not being detected by this system.

In Section 2 we present a description of how ISP email handling systems are organised. In Section 3 we outline the types of heuristics that have proved to be useful and in Section 4 we present the results from using these heuristics, which demonstrates that, even with the current rough tuning of "trigger levels", they are successfully detecting customers with problems. In Section 5 we look at data for incoming email from the rest of Internet, which allows us to estimate the total number of customers who are likely to be having problems, and hence deduce how many remain undetected. Finally, in Section 6 we discuss where this type of log analysis may lead.

2 ISP Email Handling

ISPs supply a variety of email services to their customers. A common arrangement is to provide a "smarthost" for outgoing SMTP connections, which simplifies the sending of email by the ISP's customers. For incoming email a POP3 (or IMAP) server is offered, and there will also be an SMTP server – the "MX host" for the customer domains – that accepts email from elsewhere on the Internet and stores it into the POP3 mailboxes. For performance reasons, the three logical components, outgoing smarthost, POP3 service and incoming SMTP server are often split across separate machines or clusters of machines.

Customers may be compelled, by port 25 blocking, or by "transparent" redirection of SMTP connections, to use the smarthost. If so, then the type of log processing described in earlier work (Clayton, 2004) will be effective in detecting customers whose systems are being hijacked to send spam, or who are infected by viruses that are attempting to propagate via email. However, many ISPs continue to allow direct SMTP connection to the wider Internet because they do not believe the benefit of blocking is worth the inconvenience to their customers. These ISPs will therefore be unaware of spam and viruses issuing from their customers until the receivers report it to the ISP's abuse team.

However, where email, whether spam or a virus, is sent to another customer of the same ISP then it will be handled by the incoming email system. Even email sent by customers to themselves will follow this path. Spammers may try and avoid local deliveries, but viruses are seldom so clever, especially where customers have multiple email domains, owning `example.co.uk` as well as `example.com` and also processing email addressed to a subdomain of their ISP, such as `example.isp.co.uk`. In addition, spammers often send to low priority (fallback) MX hosts, hoping to avoid blocking systems at the primary site, and again this can mean that the ISP receives the email, rather than the customer receiving it directly[1].

3 Log Processing Heuristics

Just three very simple heuristics are currently being trialled on a live system at an ISP. The first and simplest is to check whether the incoming content detection system (at the ISP considered, this was supplied by Brightmail) believes the content of the email to be "spam". Although at first sight, this is very significant, it can result in false positives when the email involved originated elsewhere and is merely being forwarded from one customer site to another – perhaps to a remote office or to offspring away at University. At present, following some initial testing to establish suitable parameters, forwarding is deemed to be occurring if more than 4 emails (that have been flagged as spam) are sent to an identical destination address. To further reduce false positives, no report is made unless more than 20 "non-forwarded" emails have been flagged as spam. Clearly these numbers are fairly arbitrary, but they seem to give reasonable results at present. Unfortunately, the ISP devolves virus checking to its customers and so there is no equivalent content-based heuristic test that can be used for virus infected emails.

The second heuristic is to count variations of the HELO (or EHLO) message used by the email sender as part of the SMTP protocol (Klensin, 2001). This text should give the name of the sending machine and hence, in most common cases, it will be constant for any given IP address. However, the HELO is commonly forged by spammers and viruses. Spam often has a forged source address which varies from one email to the next, in an attempt to evade detection and fool filtering systems, and the HELO is then forged to match. Alternatively, some spammers and many viruses forge the HELO to match the destination address – perhaps hoping that this will act as an authenticator for access. False positives arise for dynamic IP address usage, when multiple machines share a single IP address (using Network Address Translation (NAT)), and in other situations where machines legitimately use multiple HELOs. Nevertheless, in practice, good results have come from generating reports as soon as 3 different HELO strings have been used.

The third and final heuristic is to consider customers who are attempting to send email to remote sites via the incoming email system, which, at the trial ISP, refuses to relay the email. Correctly configured customers will of course use the smarthost for outgoing email, but viruses (and spammers who have not done their homework) assume that looking up the MX record for a host will yield the name of a machine that will accept outgoing email. But the MX record points at the incoming server, and a telltale pattern of failures is created in the logs. Spammers are reported (Spamhaus Project, 2005) to be trying to send more email via ISP servers because of widespread blocking of customer address ranges. This can only be welcomed, because if they use the smarthost then log processing will detect their activity, and if they use the incoming email system then not only will they be de-

[1] We are not aware of formal studies of how prevalent is the use of secondary MXs by spammers, however there is a wealth of informal evidence that it is widespread. For example, Linfoot has tracked deliveries to his systems for several years and in August 2004 was reporting that 100% of email sent to his secondary MX was spam (Linfoot, 2004). Later, this led him to deliberately block deliveries there.

tected, but their email will be entirely rejected. False positives for this third heuristic will only occur when customers misconfigure their systems, so the threshold can be set very low.

4 Experimental Results

We examined the email logs for Demon Internet, a medium sized ($\sim$ 200 000 customer) ISP in the United Kingdom. The customers connect via a mixture of dialup, ADSL and leased lines. We considered the four week period from 20 February to 19 March 2005, which, since there were no national holidays, covered 20 working days and 4 weekends. During this time, customers sent just 4 204 828 emails to the incoming mail system, from 9 521 different source addresses. The overwhelming majority of Demon's customers use static IP addresses, which considerably simplifies the processing of historical data, and so for our analysis, hopefully without loss of generality, we excluded from the data the few (employing both dialup and ADSL) who used dynamically allocated addresses – leaving 8 445 customers who sent 3 665 883 emails.

Table 1 gives the results of the log processing analysis program, which uses the heuristics given above. The false positives were determined by manual inspection of the reports to ensure that they were correct. The false negatives – the reports that were not made, but should have been – were determined by using much more aggressive settings for the heuristics and running them over the whole dataset at once, rather than processing one day's logs at a time. These two changes to the analysis meant that if customers only occasionally sent a problematic email then an anomalous pattern would still be spotted.

Table 1: Customers Detected by Log Progressing

Problem type	Valid reports	False positives	False negatives
Virus infected	318	5	88
Sending spam	78	6	52

It can be seen that the current tuning ensures a low number of false positives but that quite a few problems are being overlooked. Most of the virus incidents that were missed involved only a handful of virus emails spread over several days, viz: there was no "locality of access" to the spread of the malware and it is unclear that it is possible to improve these figures substantially without a virus detection system to report on the email content. The majority of the false negatives for the spam also involved very low volumes of traffic – and here there was undoubtedly an attempt being made by the spammers not to generate local traffic, viz: they were attempting to hide and being pretty successful at doing so.

Table 2: Top 20 Virus Sources (by AS)

Count	AS	Description and Country
295	2856	BTnet (UK)
136	4134	CHINANET (CN)
107	5089	NTL (UK)
70	5462	Telewest (UK)
66	3352	Telefonica (ES)
55	9105	Tiscali (UK)
54	3269	Telecom Italia (IT)
47	9121	TTnet (TR)
45	4837	CNC (CN)
34	20959	Telecom Italia (IT)
32	2529	Demon Internet (UK)
32	4766	Korea Telecom (KR)
31	3215	France Telecom (FR)
28	3320	Deutsche Telekom (DE)
27	4538	CERNET (CN)
25	9498	Bharti Infotel (IN)
22	4589	Easynet (UK)
21	4788	TM Net (MY)
21	6871	Plusnet (UK)
20	3462	HiNet (TW)

Note that where a customer is misbehaving, but through chance or design avoids ever sending email to the ISP's smarthost, then they will not be detected at all and such customers do not occur in table 1, even as a false negative. In the next section we consider how we might estimate the number of customers who might be being missed entirely.

5 Incoming Email From Remote ISPs

Similar heuristics were used to examine incoming email traffic from remote sites (in fact slightly less aggressive HELO detection was used, 5 different strings were required, rather than 3). Logs for a single day, Wednesday 16 March 2005, were considered, during which 6 612 496 emails arrived from 413 728 different IP addresses. Of these, 2 527 were detected to be sending virus traffic and 35 615 were detected to be sources of spam.

The ISP that is ultimately responsible for each of these addresses was established by looking up which AS (Autonomous System) was announcing the IP address; 593 different ASs were sending viruses and 1 822 were sending spam.

Table 3: Top 20 Spam Sources (by AS)

Count	AS	Description and Country
3 416	4134	CHINANET (CN)
3 036	4766	Korea Telecom (KR)
2 883	4812	China Telecom (CN)
1 711	9318	Hanaro Telecom (KR)
831	6478	AT&T (US)
707	12322	Proxad (FR)
603	9277	Thrunet (KR)
574	3356	Level 3 (US)
549	22909	Comcast (US)
444	3786	Dacom (KR)
405	7738	TeleBahia (BR)
374	3215	France Telecom (FR)
364	27699	Telesp (BR)
355	5617	TPNet (PO)
327	3269	Telecom Italia (IT)
319	3320	Deutsche Telecom (DE)
289	7132	SBC Internet (US)
285	4837	CNC (CN)
271	16338	Auna (ES)
265	7015	Comcast (US)

As can be seen tables 2 and 3, listing the top 20 ASs in each category, there is a noticeable difference between the sources of the two types of traffic. UK and European ISPs dominate the virus traffic table, but Asian ISPs host the major sources of spam, with US and European ISPs trailing behind. One should not read into this that Asian and US sites have no viruses, but merely that the infected machines are mainly unaware of the email addresses of Demon Internet customers, whereas the spammers mine global sources of information for their target lists. There are undoubtedly other biases present as well, for example the count is of IP addresses, and dynamic allocation may skew the counts higher for dialup (where multiple short connections will be allocated multiple addresses) as opposed to rather more long-lived ADSL connections.

Further examination of the server logs shows that these figures significantly underestimate the sources of both viruses and spam. This is because most IP addresses only sent a very small number of emails during the single day that was considered – and this was insufficient to trigger the heuristics. This can be seen by considering the traffic from particular subnets. For example, 81.156/16, a BT subnet of 65 536 addresses used for ADSL customers, had 7 IP addresses within it reported for spam and 4 for viruses, but there were 121 sources of email, only 3 of which looked as if they could possibly be genuine. Performing this counting exercise (albeit without taking a view as to what might be genuine) for all the /16 subnets allocated to the ISPs in the "top 20" tables gave the results in table 4 (note that virus and spam reports have been combined).

The last column of the table is the ratio between the number of IP addresses reported and those which sent any email at all. There is a striking difference between values for Asian networks (where the ratio is much less than 100) and for UK networks (where it is generally from two to five times higher). It is difficult to draw any firm conclusions from this, but the most likely explanation is that the higher rates of sending from the Asian networks have made detection more likely.

Looking specifically at UK networks, the detection ratio averages about 300. It seems reasonable to believe that a similar ratio will apply to Demon Internet's customers, since they are also in the UK and are therefore likely to have a similar profile. On the day in question, 42 reports relating to Demon Internet customers were received. This suggests that there could be more than 12 000 customers with problems, a very great many more than the 530 who were detected or found to be "false negatives" over the month that was studied (and more even than the 8 445 who sent any email to the incoming system at all). Put simply, the detection rate of about 11 customers a day is making limited inroads into the overall problem. Of course this does not make the system valueless, but it does suggest that there are distinct limits to what an ISP can detect on its own email systems, however carefully it looks.

6 Conclusions and Future Directions

Processing email server logs continues to prove extremely useful in detecting customers who are infected with viruses or who are unwittingly sending out spam. Simple heuristics can detect this behaviour and distinguish it from other activity such as relaying messages to other sites.

It is possible to get reasonably accurate detection (limiting the false negatives) without incurring the considerable expense of manually discarding a large number of false positives.

The same log processing can also be used to detect the problems of other ISPs' customers. This enables an estimate to be made of the number of customers whose problems are not being detected at all. Unfortunately this suggests that quite a large number of customers have problems that will only come to light when remote sites complain to the abuse team.

Clearly, it would be valuable to use the system to report problems with customers at other ISPs. This is easier to say than to do, but it is being actively considered. There are legal hurdles to overcome, such as

Table 4: Ratio of Abuse Detection to Email Senders

Description and Country	/16s	Sending addresses	Percentage sending	Detected problems	Detection ratio
China Telecom (CN)	21	88 968	6.46%	2 907	30
CHINANET (CN)	179	166 220	1.42%	3 487	47
CERNET (CN)	20	2 369	0.18%	40	59
TeleBahia (BR)	19	26 435	2.12%	402	65
TTnet (TR)	13	13 642	1.60%	194	70
HiNet (TW)	58	14 092	0.37%	198	71
Hanaro Telecom (KR)	69	126 475	2.80%	1 740	72
Dacom (KR)	49	39 437	1.23%	491	80
TM Net (MY)	13	8 076	0.95%	98	82
CNC (CN)	98	32 207	0.50%	386	83
Thrunet (KR)	19	48 547	3.90%	566	85
TPNet (PO)	41	35 961	1.34%	366	98
Telecom Italia (IT)	24	6 889	0.44%	69	99
Comcat (US)	13	26 784	3.14%	261	102
Comcast (US)	77	66 251	1.31%	596	111
Bharti Infotel (IN)	10	9 204	1.40%	81	113
Level 3 (US)	72	68 234	1.45%	589	115
Korea Telecom (KR)	184	359 357	2.98%	3 084	116
Telesp (BR)	13	37 945	4.45%	324	117
Deutsche Telekom (DE)	136	45 033	0.51%	347	129
Telecom Italia (IT)	93	49 971	0.82%	383	130
AT&T (US)	60	111 857	2.84%	817	136
Proxad (FR)	51	109 254	3.27%	716	152
Telefonica (ES)	85	49 657	0.89%	321	154
BTnet (UK)	58	76 031	2.00%	474	160
France Telecom (FR)	76	69 401	1.39%	406	170
Auna (ES)	18	68 847	5.84%	280	245
NTL (UK)	74	61 360	1.27%	220	278
Telewest (UK)	37	69 940	2.88%	243	287
Tiscali (UK)	14	32 602	3.55%	87	374
Plusnet (UK)	6	10 585	2.69%	27	392
SBC Internet (US)	158	138 570	1.34%	304	455
Easynet (UK)	11	25 779	3.58%	27	954

Data Protection legislation and expectations of confidentiality. There is a need to standardise reporting formats so that remote ISPs can rely upon the integrity of reports and feed them into their automated systems. Most significantly perhaps, is the need to educate abuse teams into understanding that traffic data extracted from logs can be an extremely accurate indicator as to the content of email.

Acknowledgements

We wish to recognise the vital support of Demon Internet in providing email logs and the patience of their abuse team in dealing with the automated reports that were generated. We also acknowledge the financial assistance of the Cambridge MIT Institute (CMI) through the project: "The design and implementation of third-generation peer-to-peer systems".

References

Clayton, R. (2004). Stopping spam by extrusion detection. *Proceedings of the First Conference on Email and Anti-Spam (CEAS)*. http://www.ceas.cc/papers-2004/172.pdf.

Klensin, J. (2001). Simple mail transfer protocol. IETF RFC 2821. http://www.rfc-editor.org/rfc/rfc2821.txt.

Linfoot, C. (2004). Spam/virus stats for august 2004. Blog Entry, Chris-Linfoot.net. http://chris-linfoot.net/plinks/CWLT-64FDZF.

Spamhaus Project (2005). Increasing spam threat from proxy hijackers. Press Release, SpamHaus, 2 Feb 2005. http://www.spamhaus.org/news.lasso?article=156.