Mail Gateway Virus Scanning Proposal

Gary Scwartz
Michael D. Sofka

April, 2002

Sophos Anti-Virus, under a demonstration license, has been checking email attachments on mail.rpi.edu since the mail system upgrade on February 16, 2002. Since February 20th actions taken by Sophos, and separately the actions that would have taken by heuristic virus detection methods alone, have been logged.

This document is a summary of the results the test, as well as a proposal for institute-wide email and AFS virus scanning using Sophos. This document also summarises blocking actions taken by sendmail, many of which are due to virus infections on the connecting client or relay machine.

Background and History

Email at Rensselaer is delivered by sendmail. Starting with version 8.10 sendmail could communicate with external programs via the Milter API. A Milter can request information from sendmail about the message, including the sender, recipient, connecting mail relay and the message body. A Milter can then alter the message, re-route it, reject it, and so on. Since December 20th mail.rpi.edu has run a Milter called MIMEDefang1, which had previously been installed on lists.rpi.edu after a series of wide-spread virus incidents in late November 2001. On lists, MIMEDefang was configured to block executable attachments, and reject suspicious messages.2

The experiment on lists.rpi.edu was successful, and MIMEDefang was installed on mail.rpi.edu prior to Christmas break, 2001. Executables were allowed on mail.rpi.edu, but a list of the names of known virus attachments was added to the MIMEDefang rule set.3 Since MIMEDefang was installed there


2. For example, an attachment named resume.doc.exe, which is an executable that may appear to a user as a Microsoft Word document.

3. For example: notepad.exe, thought.com, comical_story.doc, .... This is crude tests, since the attachment may be virus free. It is also not effective against viruses which generate
have been no major virus outbreaks at Rensselaer. Viruses such as Magistr, which sends a plain executables with an enticement to open the attachment, and SirCam, which sends a random attachment from the infected machine, remained endemic.

When the new mail machine was put into service there was sufficient CPU and I/O to run Sophos Anti-Virus’s (SAV) sweep program to search for viruses in email attachments. Sophos is the preferred virus detection software for many mail hubs because it is fast, lightweight, allows for quick (real-time) virus pattern updates by http, and runs on a variety of UNIX platforms including Linux, and AIX (PowerPC). Sophos is being run on all message attachments by MIMEDefang. The results of both SAV sweep, and the heuristics used in the two months prior to SAV, are logged. This allows a comparison of Sophos (and by extension other virus scanners) with heuristic anti-virus methods. Since mid-March all mail to newsfeeds.rpi.edu and lists.rpi.edu has been redirected to mail.rpi.edu for SAV scanning.

Mail and Mail Anti-Virus Procedures

With the installation of MIMEDefang and Sophos Anti-Virus, along with sendmail’s relay and address checks, a virus sent to mail.rpi.edu can be intercepted at many levels:

- Sendmail may reject connections from machines known or suspected of being infected with a virus. In the case of mail relays users who have sent email from a virus infected machine by be blocked individually. Prior to installing MIMEDefang this was our only option for stopping viruses a random attachment name, or select an attachment from stored email. Finally, there is a limit to how many attachment names can be blocked without inconveniencing users.

4. under a 30 day demo license. The installed binary, however, is good for 90 days of virus updates, and Sophos is flexible with educational institutions on tight budgets.

5. When scanning for viruses in email it is necessary to unpack the attachments, which are encoded in base-64 or other ASCII encoding, prior to scanning. MIMEDefang in it’s default installation unpacks attachments into a working directory where they can be scanned. The results are reported back to MIMEDefang, which can then take action such as rejecting the message, sending an warning to the message sender, quarantining viruses, and so on. There are other UNIX mail scanners, and if Sophos alone can catch all viruses a lighter-weight scanner such as MailScanner (http://www.smg.ece.soton.ac.uk/mailscanner/) may be used instead.

6. Both lists and newsfeeds apply additional rules post-SAV to reject all executables attachments. Both of these machines distribute messages to a large set of recipients on and off campus. As such, even a small virus window can result in numerous infections. Rejecting executables prevents most viruses.

7. The goal, as always, is minimise imposition to users. Blocking a single client machine is preferable to blocking a user account, since users may send mail from multiple machines. But, blocking a single account is preferable to blocking a mail relay, since multiple users send mail through relays.
in real time. The overhead of running a MIMEDefang and SAV is high enough, however, that relays or accounts that repeatedly send viruses are still blocked to ensure good mail performance.

- MIMEDefang may find a suspicious attachment. For example, an attachment ending in `.doc.exe` is an executable on the Windows platform, but it’s name suggests the sender is attempting to pass it off as a Word document file. Likewise, the W32/Klez-G virus is an executable sent with a MIME-type of “application/x-wav”. Normally, a WAV file ends with the `.wav` attachment, and not `.exe`. MIMEDefang also checks for the names of known virus attachments.

- Sophos Anti-Virus scans attachments for known viruses. This includes executables which MIMEDefang would allow.

- If the above methods fail (possible for new, fast-spreading viruses not recognised by SAV) the already delivered virus must be removed from the mail spool. This requires checking each of the 10,000 mailboxes on mail.rpi.edu for a pattern matching the virus. This may be an attachment name, a section of ASCII encoded virus, or scanning with SAV sweep. This is a disk intensive operation which adversely affects mail performance. Prior to MIMEDefang mail spool scanning was an ongoing, 7x24 process. Often one scan did not finish before the next scan (with a new virus pattern) was started. Spool scanning is less common now.

As you can see, professional virus scanning software fills a niche by catching viruses that heuristics cannot detect. The question is how many such viruses are there, and can we catch them by other means.

**Results of Sophos Test**

MIMEDefang has been running SAV in parallel with heuristic virus detection rules. All attachments are first scanned by SAV, and then checked for inconsistent MIME types, hidden executables, and known virus attachment names (as outlined above). Any message containing a virus or suspected virus is rejected or defanged.8 The results of both tests are logged for later analysis.

Between February 20th and April 3rd action was taken by MIMEDefang with SAV, or by sendmail on 508,186 out of 3,168,101 messages. This means a little over 15% of the messages sent to mail.rpi.edu were rejected for some reason. SAV found 49 virus types in 40,046 messages. 30,886 of these, however, were

---

8. Defanging means the attachment is encoded as a non-executable binary. A text attachment is inserted explaining why the attachment was defanged, along with generic instructions for recovering the original attachment name. The goal is to avoid defanging, since it is annoying to users, but use it when caution is required.
a copy of the W32/Goner-A virus sent from a single source over one weekend. Excluding Goner-A, SAV found 9,160 viruses.9

MIMEDefang Results

MIMEDefang alone is able to block any other virus that uses a fixed attachment name, or a small set of fixed attachment names. During the 6 weeks of data gathered for this report, Goner-A, Gibe-A and a single copy of Magister-A are the only viruses blocked by attachment name. It is too easy for virus writers to select random attachment names, or to change attachment names, for this to be an effective method of blocking viruses. It also requires active vigilence to survey virus reports looking for possible attachment names. In addition, blocking by attachment name is no mean task, since virus reports often give conflicting information, or leave out needed details. There is also the practical consideration of how many attachment names can be blocked, and for now long.

MIMEDefang found 2,511 hidden executables, and an additional 2,845 executables with questionable or risky MIME types. Of these, SAV failed to detect a virus in 688 of the attachments. On closer inspection, some of these messages were intended viruses that for some reason did not include the actual virus.10 The bulk of these misses consisted of 577 attachments sent by Microsoft Outlook that were defanged by MIMEDefang, and likely represent false alarms. That is, using the heuristics “executables sent by Outlook are suspect,” MIMEDefang took action. But, according to Sophos, no action was required. One may question this heuristic, but during the 6 weeks of the test Sophos found a virus in 52% (704 out of 1,347) attachments sent by Outlook. An additional 61 attachments were corrupted and could not be scanned by Sophos.

Comparison

MIMEDefang detected viruses in 3,934 attachments. Of these, 688 were likely false-alarms. That is, MIMEDefanged detected a virus, but Sophos Anti-Virus did not detect a virus in the attachment. Of these, 577 were executables mailed from a Microsoft Outlook client, that is, they are MIME type application/msdownload, and were “defanged” by MIMEDefang.

MIMEDefang failed to detect a virus in 5,914 attachments in which Sophos detected a virus. About 2,159 of these represent potential false-alarms by SAV. That is, the attachments were password protected (1,335 attachments), were corrupted (795 attachments), or were in an unsupported file format (29 attachments). This number is inflated, however, by 1,221 attempts to send a single

9. Goner represents a clear outlier, which is safely ignored in all comparisons because MIMEDefang by itself is able to block this virus.

10. The virus may have been stripped by an upstream mailer, or an error in the virus code sent an incomplete message. Viruses sometimes include Simple Mail Transport Protocol code that is less than 100% reliable.
password protected excell file. Without more information it is not possible to
know how many of the corrupt files were partial viruses or miss-labeled types.
The following table summarise the test results:

<table>
<thead>
<tr>
<th></th>
<th>MIMEDefang</th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>virus</td>
<td>3,246</td>
<td>5,914</td>
<td>9,160</td>
</tr>
<tr>
<td>no virus</td>
<td>688</td>
<td>3,127,367</td>
<td>3,128,055</td>
</tr>
<tr>
<td>Total</td>
<td>3,934</td>
<td>3,133,281</td>
<td>3,137,215</td>
</tr>
</tbody>
</table>

It is clear from the results that Sophos Anti-Virus was more accurate than
MIMEDefang. SAV both detected more viruses (fewer misses), and, discounting
the excell password protected file, it was more specific (lower false alarm rate).

Summary of sendmail results

Sendmail checks the connecting mail relay, the message sender address and the
message recipient address against a database of blocked hosts and addresses.
The majority of its checks are to deny relaying of non-Rensselaer email through
mail.rpi.edu, and to block possible or actual Unsolicited Commercial Email
senders. However, starting with the SirCam virus last summer, an increasing
number of the blocks were to deny mail service to virus infected machines, or
to users whose desktop is virus infected. These are course-grained tests in that
all mail from the user or machine is blocked, and not just mail containing a
virus. Further, once the user has taken action to remove a virus intervention is
required to remove the block.

In all, sendmail blocked 196,101 messages because the sender or machine is
or was infected with a virus. It is unknown how many of these represent false
alarms. The number of blocks for any one machine/user range from 1 to 176647
with 19 machine/users being blocked over 100 times.

The actions taken on the remaining 272,000 messages represent sendmail
anti-relay, anti-spam, or unknown sender/recipient domain blocks.

Mail Virus Scanning Proposal

It is clear that running a commercial virus scanning program is superior to the
combination of sendmail and MIMEDefang alone. MIMEDefang has reduced
virus infections on campus, and it or something like it will remain part of any
anti-virus strategy. But, viruses have many vectors, and the most effective
method remains sending an innocent looking executable with a message designed
to trick users into opening the file. This is how the Mellisa and ILOVEYOU
viruses spread, and it is the method used by the MyParty virus (with the twist
that the executable was named to appear as a .com web page).
A commercial virus scanner also allows us to take advantage of other people’s expertise. While writing this summary several more virus patterns were posted by Sophos Anti-virus. Most are simple variants of previous viruses, such as the LoveLet-DO virus (a variant of the original Visual Basic ILOVEYOU virus purporting to be a collection of jokes about Government and FBI secrets.) Commercial virus scanners allow for automated update of these virus patterns, night or day, weekday or weekend.

To get the most from a commercial virus scanner, it would be desirable for other campus mail servers relay their email through the campus mail hub. This can be done by using the Domain Name Service MX (Mail Exchange) records to divert incoming email to mail.rpi.edu. Once scanned, the messages can be forwarded to the final destination machine.

In order to scan all email, however, it is important to consider the separate needs of alternative campus mail machines, and to maintain high-throughput. One way to accomplish this is to separate POP3 Service from SMTP services. POP3 is the Post Office Protocol and is used by clients to pick up email. SMTP is the Simple Mail Transport Protocol and is used to route and deliver email. These both currently reside on one machine, and when one of the other is busy both services are degraded. Because POP deals with specific user mailboxes it is more difficult to share the service among different machines without using a shared file system (which is slower), or by dynamically redirecting connection requests. SMTP service, however, can be provided by any machine, or any series of machines.

We would like is to run a pool of SMTP servers on Linux machines. This will save hardware costs, introduces truly redundant mail delivery, and improve uptime and machine availability. Currently, if the mail mail server is being upgraded most users will get an error message when they attempt to read mail. A rotary of SMTP servers however, will allow service to continue uninterrupted.

The current mail machine would become the POP3 (and KPOP) server, and would be just one of the many destinations to which mail can be sent. Others destinations would include the Exchange server, and the various departmental servers.

Mail is not the only method by which viruses spread. Some viruses, for example, exploit shared file systems to spread copies. Sophos Anti-Virus can be used by a Samba add-on to scan files in AFS space as they are retrieved. This will restrict another popular virus vector.

This raises a number of technical and political issues:

**Question:** What if a departmental server overwhelms one or more of the SMTP relays?

**Answer:** This happens now, and when it does the connecting relay is blocked. With a pool of SMTP servers, however, the alternates can continue to deliver email, and the POP server would be unaffected. Under the current architecture PC clients would get connection errors, and mail from Unix hosts or off-campus relays collects (un-delivered) on the
backup MX host later delivery In short, the proposed new architecture is more robust to an aggressive or miss-configured campus mail machine.

**Question:** Would we require all mail machines on campus to have their email filtered?

**Answer:** That is the plan, using the carrot approach—“free” virus scanning, and a pool of backup MX hosts in the event a departmental server is busy or unavailable.

**Question:** What if a department wants to send very large attachments?

**Answer:** While we would prefer they use a more efficient method of sharing large attachments, as long as the final destination is not our POP3 server there is no reason to place a limit on message size. The current limit is because the POP3 server has limited storage space, and hard real-time response expectations. We may have to restrict the number of such connections taking place simultaneously, and it would be best to skip virus scanning for extremely large attachments, but these are Quality of Service issues to negotiate.

**Question:** Would this eliminate the need for desktop virus scanners?

**Answer:** No, viruses will still get in by one means or another. Scanning email would only stop known viruses from being spread through our mail servers. Laptops, for example, can pick up viruses while. Desk top virus scanning would be needed to prevent and repair such infections. Viruses can also spread through Instant Messenger, Microsoft ISS, and so on.

### Summary

Viruses cost money—lots of money. They cost money in for for of lost user time, in lost and damaged files, in CPU cycles to filter messages, and in lost reputation when your machine or site is the source of a virus outbreak. Virus scanning software also costs money, but it is a bargain by comparison. Without the Sophos Anti-Virus demo license chances are the Gibe and Klez viruses would have spread widely within Rensselaer. The Magistr, SirCam and Hybris virus remain endemic, and are detected daily by Sophos. However, in contrast to the pre-SAV days, most of these are now detected on non-Rensselaer machines.

Not running a commercial virus scanner would requires we accept a high false alarm rate. We currently disallow .exe, .com, .scr, .bat, etc on the listproc machine. .doc and .xls are allowed as a convenience to users, but they too can spread viruses. Without a good virus scanner it is only a matter of time before a sufficiently virulent and polymorphic .doc virus requires some or
all .doc files be blocked. Relying only on updated user desktop to stop viruses has failed to date. It is time to take virus scanning to the next level.

The Sophos Anti-Virus test showed that a commercial product detects more viruses, with a lower false alarms rate. What this test didn’t show, and can’t show, is what would have happened were we not running Sophos. We can only speculate about how severe the outbreak of Give-A might have been before a new check were added to MIMEDefang. Since Sophos was installed, however, even minor virus outbreaks have stopped. And now the overwhelming majority of viruses come from off-campus, and only a handful of campus machines have had to be blocked by sendmail, and then only because the virus’s SMTP code caused a high load on mail.rpi.edu.

By comparison, blocking campus machines was a weekly or even multi-daily activity prior to SAV. This, presumably, means that fewer Rensselaer based machines are sending viruses to non-Rensselaer accounts. Running gateway anti-virus software is not only good for Rensselaer users, it is good for Rensselaer.