Spotting spam in sampled sFlow

Richard Clayton

University of Cambridge, Computer Laboratory, William Gates Building,
JJ Thomson Avenue, Cambridge CB3 0FD, United Kingdom
richard.clayton@cl.cam.ac.uk

Extended Abstract

Email spam is not just a technical problem, but a market failure compounded by regulatory deficiencies. However, at present, it is technology that is giving us most of the countermeasures against spam, with filtering being key to keeping spam out of our inboxes and hiding the worst of the problem. Two (rather contrary) trends in filtering can be distinguished: highly personalised Bayesian filtering, and corporate out-sourcing of filtering decisions to specialist companies. It must be expected that both techniques will become far less effective over time, as spammers perfect methods of making spam look just like "real" email. However, spam can never be quite "the same" as real email – and the area where it will be hardest for the spammers to blend in will be their communication patterns.

The spammers are doing things which normal senders of email do not. They are sending in bulk with few replies coming back; they are sending at all hours of the day; and, they are sending with limited discrimination in their targets. Since they are not doing the same thing as legitimate senders, their activity should now (and will continue) to be distinguishable by means of traffic analysis that picks out the distinctive patterns that spamming exhibits.

Traffic analysis is already known to be effective in detecting spam. Many customers send their email via their ISP's "smarthost" outgoing email server. Processing the activity logs from this server can pick out the customer machines that have been hijacked by spammers [2]. However, the email that is sent "direct" to its destination does not show up in the smarthost logs (for the obvious reason that it doesn't use those machines). The email will, of course, show up in the logs at the destination; where spam-like patterns can be detected. Some good results have already been obtained from traffic analysis of recipient logs for the special case where an ISP customer is sending email to other customers of the same ISP [1]. The open question is whether traffic analysis will be as effective at lower levels in the protocol stack, and in particular when there is no "content" available, even of the SMTP protocol, just a record of the source and destination of TCP packets.

The LINX is the major UK Internet Exchange Point (IXP) where almost all UK ISPs interconnect, so as to pass Internet traffic to each other's networks. LINX generates sFlow data, which is packet header information about a statistically sampled subset of the packets flowing through the IXP switches. The sFlow data does not contain any packet contents, but, because email uses a fixed port (`tcp/25`) it can be used to determine which machines are sending email and to estimate volumes. However, the sampling (only 1 in 2048 packets is recorded) is a significant impediment to transforming this data into useful metrics that can form the basis of reports to ISPs about problems.

As an experiment, sampled port 25 sFlow data was collected from one of the two LINX peering rings for one day (13.9 million packets). At the same time, the incoming email server logs for a medium size (150 000 customer) ISP were analysed, combining the traffic analysis techniques mentioned above with the results from a commercial spam filter which processed the incoming email. The results from the log processing were used to label the IP addresses in the sFlow data as "good" (not sending spam), "bad" (sending spam) or "unknown" (no data). The patterns of sFlow detected traffic for some typical "bad" source addresses are shown in figure 1. This indicates spammers are currently attempting to hide from naïve blocking systems by sending in small bursts.[1]

[1] It is not possible to say whether the quiet periods are actually quiescent, or whether during those periods the traffic is being sent over routes that that do not traverse the LINX peering LAN where the sFlow data was collected.

Since this bursty pattern fitted preconceptions as to the current nature of spam sending activity we processed the data to look for non-uniform activity. As will be noted, the sFlow sampling was only detecting a handful of packets per hour, so a very course-grained method was used. We considered only email sources that were "good" or "bad" from a single multi-million customer UK ISP (which we expected to have a mix of business and consumer customers). We determined how many 1-hour-long periods exhibited activity and how many 4-hour-long periods and then looked for correlations. Fig 2 shows that how the proportion of sites sending spam is around 80-90% when only the total active hours are considered, but there are significant differences when the sources are sub-divided into six groups based on how many 4-hour periods showed any activity.

This work is clearly at an early stage, and we will be continuing this analysis with a view to refining the technique; and considering other discriminators such as the number of destinations email is sent to, the presence of two-way traffic and so forth.

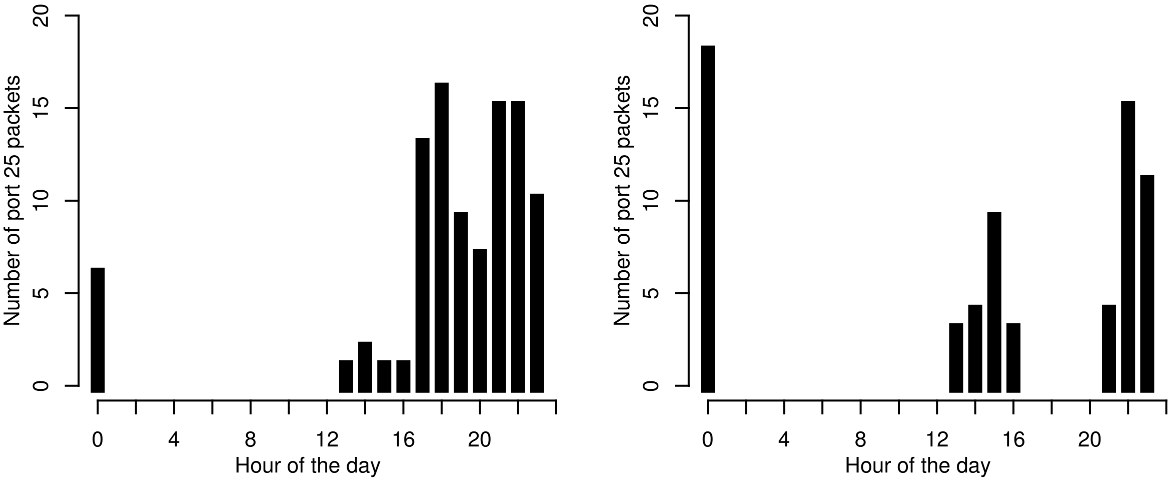

Figure 1: Email traffic volumes in sampled sFlow data for two typical spam sending sources

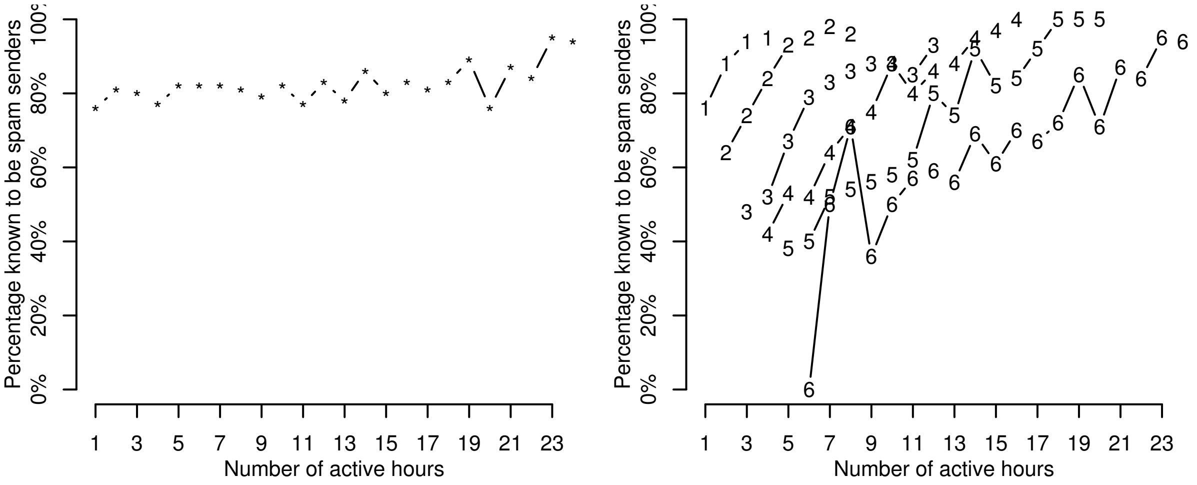

Figure 2: Left: proportion of sources sending spam collated by total hours active; Right: these sources are split into six groups depending on the number of 4-hour periods during which they are active

References

[1] Richard Clayton: Stopping Spam by Extrusion Detection. First Conference on Email and Anti-Spam (CEAS 2004), Mountain View CA, USA, July 30–31 2004. `http://www.cl.cam.ac.uk/~rnc1/extrusion.pdf`

[2] Richard Clayton: Stopping Outgoing Spam by Examining Incoming Server Logs. Second Conference on Email and Anti-Spam (CEAS 2005), Stanford CA, USA, July 21–22 2005. `http://www.cl.cam.ac.uk/~rnc1/incoming.pdf`