the Availability Digest

www.availabilitydigest.com

Spamalytics

October 2009

Masses of spam can bring your email service to its knees. To protect against this, complex spam filters are used on email servers and on browsers. But how effective are the filters?

To answer this question, a team of scientists at the Department of Computer Science and Engineering, University of California, and the International Computer Science Institute became spammers of sorts. They reasoned that the best way to measure spam is to be a spammer.

The scientists orchestrated a parasitic infiltration of an existing spam botnet's infrastructure and caused it to modify some of the spam it was sending so as to redirect the spam to web sites under the control of the team. These web sites presented "defanged" versions of the spammer's own web sites.

The infiltration was accomplished by masquerading as bots under the control of the infiltrated botnet.[1] The team modified spam so that it would cause no harm to anyone. All it did was to allow the team to measure the conversion rate of its modified spam campaigns.

The team defined "conversion rate" as the probability that a spam email will result in a desired action. There were two desired actions that they studied – the purchase of a product and the infection of a browser.

Armed with these results and with the cost of spam mailings, the team then considered the value proposition of spamming. They call this science "spamalytics."

The Subterfuge

The Storm Botnet

The research team selected the Storm botnet as its victim. Storm is one of the largest spam service providers. It spreads itself by encouraging recipients of some of its email spam to visit a web site under its control. Those that visit and then download an executable become infected with the Storm Trojan. The Storm Trojan is a piece of malware that contacts a Storm master server and reports for duty as a bot.

Storm organizes its bots into a hierarchy. At the top are the master servers that are controlled directly by the botmaster. The master servers communicate with proxy bots that are infected browsers. Proxy bots communicate with worker bots that form the mass of the Storm army. The

[1] Chris Kanich, Christian Kreibich, Kirill Levchenko, Brandon Enright, Geoffrey Voelker, Vern Paxson, Stefan Savage, Spamalytics: An Empirical Analysis of Spam Marketing Conversion, *Communications of the ACM*; September, 2009.

worker bots harvest email addresses from the address books of their local browser hosts for use by the botnet.

A spam campaign is defined by an email message and a recipient count (typically in the millions of recipients). To initiate a spam campaign, a master server distributes directives to a set of its proxy bots with the spam email message and a target address list for each proxy bot. The total of the target email addresses equals the recipient count for the campaign.

Each proxy bot then sends directives to its designated list of worker bots with the email message and a subset of the target address list. Each worker bot sends the spam email message to its specified addresses. A spam email can be sent to millions of unsuspecting users in minutes. Upon completion, each worker bot sends a completion report to the master server via its proxy bot.

The Infiltration

The team's infiltration strategy was to pose as a set of eight proxies to a Storm master server by infecting the team's hosts with the Storm malware. This would cause the team's hosts to contact a Storm master server, which would appoint the hosts as proxy bots. When one of the team's proxy bots received a directive for a spam campaign from a master server, the proxy bot would modify the URL in the email message before sending it to its worker bots so that the email message pointed to a defanged web site managed by the team. Two web sites, identical in appearance to the spammer's web sites, were developed:

- A web site that attempted to sell pharmaceutical products to the recipient. It collected no sensitive information such as name, address, or credit card number. A "sale" occurred if the recipient accessed the web site and went to its purchase page.

- A web site that appeared to deliver an electronic postcard. This web site was designed to propagate a harmless Trojan to the browser. All the Trojan did was to report back to the team's proxy botr that it had successfully infected the browser.

To evaluate the effectiveness of different email-filtering methods, the team set up several test email addresses and included them in the target email address lists sent to the worker bots. By noting the number of messages that were actually received, the proportion of messages rejected by the filters could be determined.

The Campaigns

Three spam campaigns were designed. Two of these, the postcard campaign and the April Fool's campaign (sent as postcards on or about April 1), were structured to propagate a Storm Trojan. The third campaign was one designed to sell pharmaceuticals.

The campaigns lasted almost a month. During this time, worker bots were instructed to send 348 million pharmacy spams, 75 million postcard spams, and 37 million April Fool's spams

Together, over 469 million emails were sent. Under the direction of a Storm master server, up to 2.5 million emails per hour were sent. The average number of worker bots assigned to a proxy bot for a campaign day averaged 23 but peaked at 500. The pharmacy campaign used more than 31,000 worker bots over its life. Most worker bots attached to a proxy bot only once over the length of the campaigns. It is estimated that these campaigns used only about 1.5% of Storm's worker bots.

The Goal

The goal of this exercise was to measure several spam parameters:

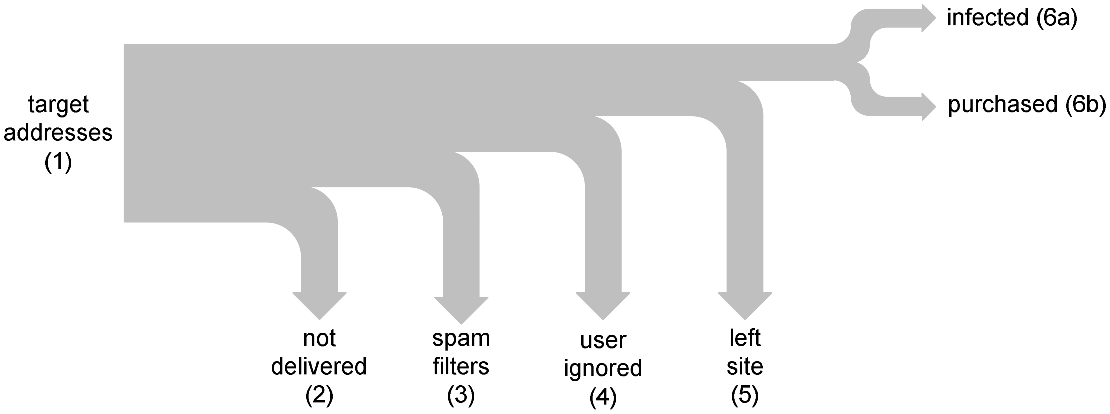
- The proportion of spam that was filtered by the antispam solutions.
- The proportion of spam that was successfully delivered.
- The proportion of spam that resulted in click-throughs.
- The proportion of spam that resulted in sales.
- The length of time that a spam web site needed to be up to garner maximum sales following a spam email campaign.

Conversion Analysis

The Spam Conversion Pipeline

The spam conversion pipeline is shown in Figure 1. It starts with a set of email addresses targeted by the spam campaign (1). Some of this email is not delivered for a variety of reasons, such as no-longer existing email addresses or blacklisted worker bots (2). Of the valid email addresses, some (hopefully most) get blocked by spam filters at either the email servers or at the users' browsers or get put into a spam folder (3). Of the email that gets through the spam filters, most will be ignored by the users (4).

Some emails will result in click-throughs to the spam web site, but the users will leave the site after taking no action (5). The remaining users who visit will convert. If the web site is a malware site, users will become infected if they download and run an executable (6a). If it is a purchase site, some users will make a purchase (6b). Actions (6a) and 6b), infection or purchase, are successful spam conversions.



Spam Conversion Pipeline
Figure 1

Conversion Results

As noted above, almost a half-billion spam emails for the three campaigns made up the target email addresses. Of this amount, 25% was reported as being delivered by the worker bots.

The team did not have a method for measuring the amount of spam email that was passed by various spam filters. However, tests of three popular free email services (Gmail, Yahoo, and Hotmail) indicated that only about .001% to .007% of the team's email messages got past the spam filter. That means that all but one to seven out of 100,000 spam messages were missed by the spam filters. About 15,000 emails of the half-billion sent resulted in click-throughs.

Of the 125 million postcard and April Fool's spam emails sent, 541 users were infected. The 350 million pharmacy emails that were sent resulted in 28 purchases (27 of which were for male enhancement products).

20% of the conversions were made in the first hour of the campaign. 70% were made in the first day following the spam campaign. 90% were made in the first week, and almost 100% were made in the first month. Therefore, a bogus web site must be around for a while to get full benefit from a campaign.

The team found that most spam email was directed at U.S. residents. However, India, Pakistan, and Bulgaria had the highest response rates. The U.S. had the lowest.

The conversion rate of spam is almost negligible;[2] but if it is inexpensive enough, it can still be profitable. This is the question tackled next by the team.

The Value Proposition

The team was quick to point out that their experiment was just one point on the curve, and other similar experiments might yield significantly different results. However, using the experiment's data, some interesting though perhaps speculative observations can be made about the economic viability of spamming.

In the case of the postcard and April Fool's campaigns, these were targeted at growing the Storm bot base and did not have a specific profit motive. It is estimated that Storm self-propagation campaigns can generate 3,500 to 8,500 new worker bots per day.

However, the pharmacy campaign was justified strictly on the basis of the money that it could make. There are three basic parameters in the profitability value proposition equation:

- The cost of a spam campaign.
- The conversion rate.
- The profit on a sale.

In the pharmacy campaign, the average sale that would have resulted was about $100. This means that the campaign would have brought in $2,800 in revenue to the party that contracted for it. The party's profit would have been a fraction of this due to possible affiliation fees, web site costs, and other expenses.

Anecdotal evidence places the cost of a spam campaign at about $80 per million emails, with substantial discounts for large volumes. Thus, the pharmacy campaign of 350 million emails would have cost about $28,000. Clearly, this particular campaign, which would have brought in only $2,800, would have been a big loser. In fact, because of the discrepancy between revenue and cost in this experiment, the authors speculate that the pharmacy web site was in fact owned and run by Storm itself.

However, clearly there are campaigns that do make money, as evidenced by the continuing spam that gets through to our mailboxes. Spam services continue to be attractive to the disreputable community of businesses.

[2] A *Wall Street Journal* investigation of Howard Carmack, the "Buffalo Spammer," reported in an op-ed article on June 23, 2003, revealed that he had converted 10,000,000 spam emails into 3,600 sales for an herbal stimulant. It appears that spam filters have come a long way.

Summary

The team concluded from this experiment that the profit margin for spam may be meager enough that spammers must be sensitive to the details of how their campaigns are run. So how can this information be used to reduce or eliminate spam?

Trying to shut down spammers under anti-spamming laws hasn’t worked. Spammers are fast on their feet, continually changing their domain addresses before law enforcement can catch up. Furthermore, many operate in countries with no anti-spamming laws.

Imposing a postage fee on email won’t work. A postage fee of only a penny per hundred emails could seriously alter the economics of spamming if the spammer paid the postage. But as seen with the Storm botnet, the spammer doesn’t pay the postage. The infected users do.

However, as seen by the minimum amount of spam email that gets through the spam filters, the spammers are economically sensitive to spam defensive measures. If we can improve these defenses faster than the spammers can react, they might dry up and go away. This requires that anti-spam defenses to stop incoming spam as well as outgoing spam must be continually improved (which they are) and that everyone must use good spam filtering tools (which they do not).

Alternatively, advertise the “delete” key.