94 percent of spam-advertised online scams are hosted on individual Web servers

August 6 2007

Geoff Voelker and Stefan Savage (left to right) are computer science professors from the UC-San Diego Jacobs School of Engineering. They found striking differences between the infrastructure used to distribute spam and the infrastructure used to host the online scams advertised in these unwanted email messages. This discovery should aid in the fight to reduce spam volume and shut down illegal online businesses and malware sites. Credit: UC San Diego

Computer scientists from UC San Diego have found striking differences between the infrastructure used to distribute spam and the infrastructure used to host the online scams advertised in these unwanted email

messages. This discovery should aid in the fight to reduce spam volume and shut down illegal online businesses and malware sites.

While hundreds or thousands of compromised computers may be used to relay spam to users, most scams are hosted by individual Web servers, computer scientists from the UCSD Jacobs School of Engineering have found. Based on an analysis of over one million spam emails, 94 percent of the scams advertised via embedded links are hosted on individual Web servers, according to new peer-reviewed research to be presented@the USENIX Security 2007 conference in Boston on August 09, 2007.

Using new Internet monitoring approaches developed@UCSD, the computer scientists studied a spam feed over the course of a week. They analyzed spam-advertised Web servers hosting online scams that either offer merchandise and services (e.g., pharmaceuticals, luxury watches, mortgages) or use malicious means to defraud users (e.g., phishing, spyware, rootkits). The researchers followed the URLs embedded in spam back to the hosting servers, probed the servers and analyzed the Web pages advertised in the spam.

"A given spam campaign may use thousands of mail relay agents to deliver its millions of messages, but only use a single server to handle requests from recipients who respond. A single takedown of a scam server or a spammer redirect can curtail the earning potential of an entire spam campaign," write the UCSD computer scientists in their paper accepted for publication@USENIX Security 2007 conference.

These new insights on the Web server infrastructure for online scams pertain to the scams advertised via spam that contain embedded links.

In 2006, industry estimates suggest that spam comprises over 80 percent of all Internet email with a total volume up to 85 billion messages per

day. What drives spam are the various money-making scams (legal or illegal) that are advertised in email messages.

"The availability of scam infrastructure is critical to spam profitability. Our findings suggest that the current scam infrastructure is particularly vulnerable to common blocking techniques such as blacklisting," said Geoff Voelker, a computer science and engineering professor@the UCSD Jacobs School involved in the study.

Through the Collaborative Center for Internet Epidemiology and Defenses (CCIED) funded by the National Science Foundation, the UCSD researchers are continuing their efforts to measure and understand the infrastructure used to support the active underground market for illegal online goods and services as a basis for developing controls and defenses against them.

Using their new "spamscatter" approach, the computer scientists studied over 1 million spam messages from a live feed (all the messages sent, over the course of a week, to any email address@a four-letter top-level domain that has no active email accounts). Spamscatter allows researchers to mine emails, identify URLs in real time and follow these links through any redirection mechanisms and on to the Web page on the destination server.

"Spamscatter provides a mechanism for studying global Internet behavior from a single vantage point," said Voelker.

The computer scientists recorded the server locations and captured screenshots of the spam URL destination Web pages. From these screen shots, the researchers grouped the scams using a technique called "image shingling." This approach matches visually similar Web pages based upon images rendered in a Web browser rather than on HTML source, URL text, or spam email contents. Image shingling enables spamscatter

to foil common scammer techniques for avoiding detection in which, for example, the scammers compose their Web sites entirely with images.

"Our image shingling approach breaks new ground in determining which servers are running the same scams," said Chris Fleizach, the second author on the USENIX security paper who recently earned a Master's degree from the Computer Science and Engineering Department@the UCSD Jacobs School of Engineering.

Using this approach, the computer scientists identified scams across servers and domains and reported on distributed and shared infrastructure, lifetime, stability, and location.

By clustering the Web pages that were visually equivalent and integrating this information into the other data collected from the spam feed, the computer scientists determined that about 94 percent of the scams advertised in spam emails with embedded URLs were hosted only a single web server.

Of the 6 percent of scam servers that were distributed across multiple servers, a few used more than ten IP addresses, and one scam used 45 servers.

"Scams might use multiple hosts for fault-tolerance, for resilience in anticipation of administrative takedown or blacklisting, for geographic distribution, or even for load balancing," the authors write, noting that most scammers are not currently taking this precaution.

The computer scientists also found that more than half of the scam servers identified in the live spam feed were in the United States, 14 percent in Western Europe and 13 percent in Asia. This finding is particularly interesting given that only about 14 percent of spam relays used to send spam to the feed used in this study were located in the

United States, while 28 percent of the spam relays were located in Western Europe and 16 percent in Asia.

"The strong bias of locating scam hosts in the United States suggests that geographic location is more important to scammers than spammers," the authors write. There are a number of possible reasons for this bias, including the perceived enhanced credibility of scams hosted in the United States. Another possibility, the authors say, is that scam hosts benefit more from stability than spam relays do, and that hosts and networks within the United States can provide this stability.

"We're learning about the hosting infrastructure of online scams from the networking point of view. We also took an inventory of what kinds of Web sites are advertising with spam," said David Anderson, the first author on the USENIX security paper. Anderson recently earned a Master's degree from the Computer Science and Engineering Department@the UCSD Jacobs School of Engineering.

Scams fell into more than 60 categories. The most prevalent scam category was Information Technology, which includes click affiliates, survey and free merchandise offers and some merchandise for sale (e.g., hair loss, software). Just over 2 percent of the scams were labeled as malicious sites (e.g., containing malware such as phishing, spyware, rootkits).

Source: University of California - San Diego