stopBADware.org
May 2008 Badware Websites Report

Abstract

Using data from Google's Safe Browsing initiative, StopBadware.org analyzed over 200,000 websites found to engage in badware behavior. The analysis found that over half of the sites were based on Chinese network blocks, with a small number of blocks accounting for most of the infected sites in that country. The U.S. accounted for 21% of infected sites, and these were spread across a wide range of networks. Compared to last year, the total number of sites was much higher, likely due both to increased scanning efforts by Google and to increased use of websites as a vector of malware infection. Several U.S.-based network blocks that were heavily infected last year, including that of web hosting company iPowerWeb, whose network block topped last year's list, no longer host large numbers of infected sites.

Data source & methodology

StopBadware.org maintains in its Badware Website Clearinghouse an updated copy of the list of active badware websites generated by Google's Safe Browsing initiative. Using the current list of active sites as of late May 2008, which totaled 213,575, StopBadware.org resolved each site's URL to an IP address and used data from Team Cymru to identify the autonomous system (AS) block name, autonomous system block number (ASN), and the network's registered country of origin. An AS represents a set of interconnected networks and routers that are operated by a single entity. The entity that operates the network block may, but does not necessarily, own or operate the servers hosting the infected websites. Some URLs (8,556) did not resolve to a valid IP address and are treated as "unknown" in our analysis. A smaller number of sites returned no data for AS block (252) or country code (2,332), and these are also treated as "unknown" in our analysis.

Sites that are identified as exhibiting badware behavior may be deliberately participating in the distribution of malware or may be compromised through manual or automated means. This report does not distinguish among sites based on intent, but rather treats all of the sites equally as vectors for malware infection.

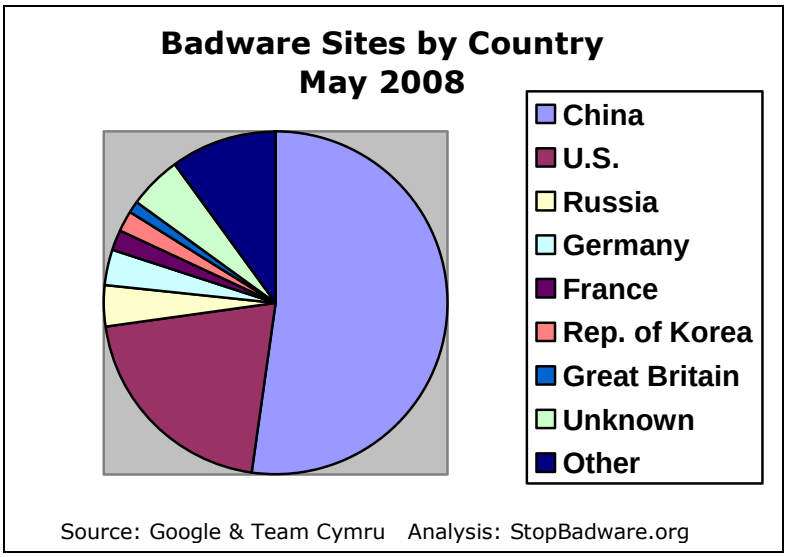
Known limitations

The data from Google limit the analysis in two ways. First, the methods Google uses to identify badware websites are limited to capturing certain common varieties of

badware behavior[1]. Therefore, this report is limited to sites that contain these particular badware traits. Second, while Google has an extensive cache representing much of the web, it is unlikely to be completely comprehensive. Furthermore, it is not necessarily the case that Google scans all web content for badware with the same frequency. It is therefore possible that the findings are skewed somewhat towards those networks, websites, or types of content that are scanned most aggressively. We present these findings acknowledging the possibility of skewed and/or incomplete results but believing that they are reasonably representative of the overall web ecosystem.

Country findings

With 52% of identified badware sites, China hosts far more sites than any other country. The U.S. is second with 21%. No other country hosts more than 4% of the world's badware sites, though a total of 106 countries host at least one infected site and 38 countries host at least a hundred.



StopBadware.org also analyzed, for the seven countries topping the list of infections, the relationship between a country's Internet-using population and its number of badware sites. It is difficult to find current numbers of Internet users by country, but using the most recent data (ranging from 2005 to 2008) from the CIA Fact Book[2], StopBadware.org calculated the badware sites per million Internet users for the world (210 sites per million) and for each of the seven countries, as shown in the table on the following page.

[1] For more info, see http://www.usenix.org/event/hotbots07/tech/full_papers/provos/provos.pdf
[2] https://www.cia.gov/library/publications/the-world-factbook/rankorder/2153rank.html

Country	*Badware sites per million Internet users*
China	689
Russia	307
United States	212
Germany	135
France	128
Republic of Korea	115
Great Britain	60

These numbers reinforce the dominance of China as a malware host, with an infection rate over three times that of the world average. Russia also stands out with a disproportionately high rate (possibly skewed by rapid growth in Internet use not reflected in the CIA's 2006 numbers), while the United States is just about average. Relative to their populations, the western European countries and the Republic of Korea are far less likely to host badware sites than other nations.

Network block findings

The top ten network (AS) blocks hosting badware websites were:

Network block name & description	*Country*	*Number of infected sites*
CHINANET-BACKBONE No.31,Jin-rong Street	China	48,834
CHINA169-BACKBONE CNCGROUP China169 Backbone	China	17,713
CHINANET-SH-AP China Telecom (Group)	China	9,445
CNCNET-CN China Netcom Corp.	China	6,058
GOOGLE - Google Inc.	U.S.	4,261
DXTNET Beijing Dian-Xin-Tong Network Technologies Co., Ltd.	China	3,604
SOFTLAYER - SoftLayer Technologies Inc.	U.S.	3,507
THEPLANET-AS - ThePlanet.com Internet Services, Inc.	U.S.	3,166
INETWORK-AS IEUROP AS	France	2,878
CHINANET-IDC-BJ-AP IDC, China Telecommunications Corporation	China	2,357

The network blocks and their owners play different roles in the Internet ecosystem. Google and iEurop, for example, use their networks to provide hosted blogs and websites, respectively, indicating that the companies have direct control over the infected servers. Google reportedly disables infected blog sites as its systems detect badware behavior[3], and iEurop has expressed a willingness to take action as badware sites are brought to its attention. SoftLayer and ThePlanet.com offer data center services and/or dedicated, self-managed hosting, indicating that they do not control the content of many systems operating on their networks. Both companies, however, have acceptable use policies for their customers and have expressed an interest in investigating potential violations of these policies. The Chinese companies in the top 10 list operate as Internet service providers (ISPs) or backbone providers, offering bandwidth to customers who may use the bandwidth for a variety of purposes. Some may also offer direct hosting services. StopBadware.org has not had success in contacting these companies.

In China, 68% of the country's infected sites are hosted on just three AS blocks, while in the U.S., the top three blocks account for only 25% of infected sites. This is likely reflective of a more centrally-controlled Internet infrastructure in China, in contrast to the highly distributed infrastructure in the U.S.

Discussion & conclusions

Web-based malware is a global problem, and nowhere is the problem more pronounced than in China. This analysis does not identify the reason for China's disproportionate share of infected sites. Additional research by StopBadware.org[4], however, has postulated that part of the reason for this could be the lack of economic incentives for Chinese hosting providers and site owners to inform their users of infected sites and/or to take action to clean or remove these sites.

In the U.S. and Europe, cooperation and data sharing among multiple links in the Internet chain have proven to be effective strategies in addressing the issue. For example, after StopBadware.org released a similar list of top infected AS blocks last

[3] Google, a sponsor of StopBadware.org, tells StopBadware.org that when a Blogger site is identified as badware by their Safe Browsing initiative, the site is immediately reported to Google's Blogger group and the site is disabled. However, the URL for the site remains listed as badware until the Safe Browsing systems rescan the site, which means that there is a lag from the time the site is rendered harmless to the time at which it no longer appears in the data used by StopBadware.org for analysis.

[4] http://weis2008.econinfosec.org/papers/Greenstadt.pdf

year[5], the owner of the most infected block, iPowerWeb, used Google's data and cooperated with StopBadware.org to clean and protect thousands of infected sites. Similar success was seen more recently on a smaller scale with U.K. hosting provider Byet Internet Services[6]. Neither hosting provider had enough infected websites on its network block at the time of our analysis to rank in the top 250 infected networks.

It is clear that further research is needed to identify the parties that are most likely to be willing and able to take action against badware sites, especially in China, and to engage in conversation with these parties. To this end, StopBadware.org intends to deepen its analysis to include Whois network data, which is more specific than AS block data, and to continue its outreach efforts to all of the network block owners identified in this report.

[5] http://blogs.stopbadware.org/articles/2007/05/04/stopbadware-identifies-hosting-providers-of-larged-numbers-of-sites-in-badware-website-clearinghouse

[6] For more info, see http://blogs.stopbadware.org/articles/2008/05/07/taking-a-byet-out-of-badware