Comments in response to Department of Commerce's Internet Policy Task Force inquiry
Ref: docket #100721305-0305-01
Submitted by Maxim Weinstein on behalf of StopBadware, Inc.
September 20, 2010

StopBadware (SBW) is a 501(c)3 non-profit anti-malware organization that originated as a project of Harvard University's Berkman Center for Internet & Society. One focus of SBW's work over the past few years has involved one particular form of cybersecurity threat, badware websites[1]. We do not identify badware sites ourselves, but rather depend upon corporate data providers to share the URLs of these sites with us. These data providers include Google, Sunbelt Software, and NSFOCUS. We use the data to:

- Identify and communicate trends in website infections and their impact on Internet users
- Educate webmasters about how to protect their sites and how to remove badware if their sites have been compromised
- Influence web hosting providers to take action when badware is prevalent on their networks
- Provide transparency into which sites have been identified as badware by various providers
- Offer "due process" for site owners that believe their sites are unfairly identified as badware

We have had limited success—punctuated by our share of challenges—collecting and analyzing the data, as well as educating and influencing the relevant parties. Our hope is that our experience in the area of badware websites provides insight into the potential challenges and opportunities implied by the Task Force's broader inquiry.

Challenge: Organizations are reluctant to share data

Many organizations that identify badware sites are reluctant to share the list of sites with StopBadware, or will do so only if StopBadware agrees not to share the data with other organizations. Some reasons:

- The "commons" problem (organization A invested resources in finding badware sites and doesn't want organization B to benefit without A getting something of equal value in return)
- Competitive advantage (the organization with the "best" list of bad sites can protect its customers more effectively than its competitors)
- Market value (the organization intends to sell/license its list of badware sites to others)
- Relationships (the organization believes it can use its knowledge of a badware site to establish a valuable relationship with the site owner, hosting provider, or other relevant party)
- Security concerns (if the "bad guys" can gain easy access to what the "good guys" know, they will use this to improve their ability to evade detection)
- Backlash and support concerns (the organization that identified the badware site doesn't want to deal with irate or confused website owners whose sites are on the list)

[1] This includes websites that host or distribute malware, including drive-by downloads, rogue anti-virus products, and malicious executable files. Some of these websites are set up for this purpose, while others are legitimate sites that have been compromised.

- Time and energy (the organization doesn't want to invest the resources in developing the systems and tools necessary to share the data or participate in SBW's review process)
- Lack of data sharing standards (each organization has different systems, different data formats, and different methods for exchanging the data with other organizations)

Despite these challenges, we have succeeded in developing a database with three, soon to be four, data providers contributing lists of badware sites. Some factors that made this possible:

- StopBadware, despite a desire for greater openness, agreed to substantially restrict the ways in which we would share providers' data.
- By offering education, community support, and an independent review process to site owners, we have reduced the backlash and support burden for the data providers.
- We offer each data provider aggregate data that compares that provider's detections with the collective information in our database.
- Data providers gain public recognition for their contributions and their support of StopBadware.

Challenge: Data is difficult to interpret

At any given time, StopBadware has over 400,000 active badware URLs in its database. Analyzing and reporting trends can be challenging for several reasons:

- Data comes in different forms (one provider might submit host names like example.com, while another might submit exact URLs like http://example.com/foo/bar.html)
- There's no consistency in URLs (example.com/a and example.com/b might be two different websites run by different people, or they might be two directories within a single site)
- Time is a moving target (new badware sites appear, get cleaned, get taken down; delays between infection/cleanup/takedown and detection vary from source to source)
- Sites all look the same (it's difficult to programmatically distinguish commercial from non-commercial, malicious from compromised)
- Identifying responsible parties is hard (whois data may be false or outdated, the hosting company associated with the IP address may not be the company actually interacting with the site owner)
- We don't know the denominator (how many websites are there in the country/world? how many on a particular hosting company's network or a specific IP address?)
- It's difficult to gauge impact (how many users visit the site? how many would have visited if not for warnings from search engine or browser? how many actually got infected? what damage was done to the end user? what did remediation cost for the owner of the compromised site?)

StopBadware and others in the community have started to identify partial solutions:

- We accept data from our providers in multiple forms and then normalize it within our database into a standard format.

- We try to provide as much context as possible when reporting data, so the value and limitations of the data are clear.
- We make use of publicly available data and tools, such as Team Cymru's IP-to-ASN service, to aggregate data in meaningful ways.
- Researchers Tyler Moore and Richard Clayton have identified a method for estimating the traffic to a compromised website, using publicly readable statistical data.[2]
- Security consulting firm HostExploit has developed a model for estimating the "badness" of a network by manipulating aggregate data from multiple sources. Although imperfect, the model demonstrates the potential of combining disparate data.[3]

Challenge: Influencing change

StopBadware, working with the Anti-Spyware Coalition and the National Cyber Security Alliance, commissioned a report earlier this year to map out the parties and interactions that are most critical to badware websites. This report, "A Broader Look at Web-Based Malware: Mapping the Threat to Better Fight It," is attached. It offers a model, known as the "chain of trust," for mapping the parties and interactions that affect the security of a particular type of transaction. Once these parties and interactions are identified, it becomes easier to develop levers for influencing change. It also becomes possible to repeatedly use the same map as a starting point for conversation with all of the relevant parties.

Even before the chain of trust report, StopBadware and other security researchers recognized the important role played by web hosting companies. These companies have several opportunities to reduce the prevalence of badware websites:

- Securing their own servers and infrastructure
- Educating their customers
- Providing technical support to find, remove, and prevent badware
- Notifying their customers when badware is reported on those customers' websites
- Taking down websites and/or suspending customer accounts when badware is not addressed

In May 2007, StopBadware released a list of the web hosting companies whose networks contained the greatest number of badware websites[4]. The top network at the time, iPowerWeb, was responsible for over 10,000 badware sites. The negative publicity that resulted from the report led to a quick response from iPowerWeb. Within a week, iPowerWeb had cleaned badware from thousands of sites and updated their servers to prevent reinfection.

Negative publicity has been effective in other cases, as well. In September 2008, a report from HostExploit identified Atrivo as a company that failed to take action against known badware activity.

---

[2] http://portal.acm.org/citation.cfm?id=1299016

[3] http://hostexploit.com/downloads/view.download/4/25.html

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

After publicity from Brian Krebs at the Washington Post[5], among others, Atrivo's network providers dropped their service to Atrivo, effectively taking the company's network offline.

Negative publicity, of course, is not always the only or best way to influence change. StopBadware has worked extensively with two of its partners, Google and Mozilla, to effectively influence website owners to clean up sites that have been compromised by badware. Following guidelines developed with StopBadware, Google uses automated methods to detect badware sites. Google and Mozilla then use this list of bad sites to warn users in Google search results, the Chrome and Firefox browsers, and elsewhere. These warnings, in addition to protecting users, often cause a substantial drop in traffic for the affected websites. This drop in traffic becomes a trigger that encourages the site owner to resolve the issue. StopBadware has worked to educate site owners on *how* to resolve the issue, through publicly available educational content and a volunteer-driven online community. In addition, we have worked with Google to ensure that sites that have been cleaned up can be removed from the badware list quickly, ensuring that site owners benefit from taking prompt action.

Closing Thoughts

Over the past few years, StopBadware and others working to combat badware websites have shown that there are sensible ways to collect data, interpret data, and influence change in the realm of cyber security. Still, many challenges remain in all three of these areas. We hope that the experiences related here will prove useful to the task force as it works to identify solutions to other problems in cyber security. If anyone from the task force would like to speak with us for additional information, please contact us using the information below.

Respectfully,
Maxim Weinstein
Executive Director

StopBadware
PO Box 380295
Cambridge, MA 02238

[5] http://voices.washingtonpost.com/securityfix/2008/08/report_slams_us_host_as_major.html