

Fools Download Where Angels Fear to Tread

An evident chasm exists between the time a new malware specimen is discovered and the time updated virus signatures for this malware are generally available, and in this chasm, 0-day malware lurks. We define 0-day malware as any

malicious software that antivirus programs can't detect due to a lack of existing virus signatures or other malware detection techniques.

Much has been said about the dangers posed by 0-day malware, but our search of the literature yielded few hard facts with respect to what quantity of 0-day malware can currently be found "in the wild." We therefore constructed an experiment in which we exposed a set of computers to several sources that potentially contain malware, to see how many 0-day malware specimens our computers could contract. Our aim wasn't to mimic any particular type of user, but rather to perform a set of actions that we believed would maximize the probability of being infected.

Method

To carry out our experiment,¹ we set up a laboratory of five computers with fresh Windows XP installations, including the latest service packs and updates. We set each PC up with different antivirus software (see Table 1); all the software automatically updated itself with the latest virus definitions when detecting an Internet connection. Our main purpose in using desktop antivirus programs was to avoid having our comput-

ers infected with known malware. We installed all the software with default settings but disabled the built-in personal firewall in Norton Antivirus 2008 to make the different software installations as similar as possible.

We installed Spybot Search & Destroy (www.safer-networking.org) on all machines to protect against spy- and adware. Although we didn't originally intend to do this, a pre-study "dry run" surfing various download sites quickly made it apparent that antispymware protection was necessary: the machines used in the pre-study became so crammed with spy- and adware that they were unusable for the intended activity, and we had to clean them up and reconfigure them.

In addition to the installed antivirus packages, we obtained two offline antivirus programs, F-PROT for Linux and the Avast! Bootable Antivirus and Recovery (BART) CD. By "offline" we mean that the host operating system isn't booted, but the antivirus software is instead run from a live CD. We used the offline antivirus packages to perform the control, baseline, and final malware scans, as we describe.

As the following time schedule

shows, we actively exposed the computers to suspicious Web sites and file-sharing networks during a two-week period in 2008, then shut them down for about a month before turning them on again to perform antivirus scans and analyses:

- 10 September—we connected all computers in the laboratory to the Internet.
- 15 September—we performed control malware scans and experiment start-up.
- 1 October—we performed baseline malware scans before shutting the computers down for a month.
- 3 November—we conducted our final malware scans.

Monday, 15 September, we actively started to expose the computers to Web sites, file-sharing systems, and so on. We had prepared an initial list of suspicious Web sites containing warez (illegally copied software), screensavers, codecs (add-on software modules for playing various audio and video files), MP3s, and other free downloads. The actual number of visited Web sites ended up being much larger because we clicked on many advertisements and visited partner sites. Although we tried to follow the same click path on all computers, the dynamic generation of banner ads, popups, and so on made a divergence between the computers inevitable. Because Web sites from Romania, Hong Kong, and Russia are considered the most risky,^{2,3} we tried to include some sites from those countries as well.

We also had a list of search

MARTIN GILJE
JAATUN AND
JOSTEIN JENSEN
SINTEF
*Information and
Communication
Technology*

HÅVARD VEGGE,
FINN MICHAEL
HALVORSEN,
AND RUNE
WALSØ
NERGÅRD
*Norwegian
University of
Science and
Technology
(NTNU)*

Keywords Used for Data Collection

avg, antivirus, adaware, limewire, frostwire, winrar, winzip, mirc, irc, player, real, media player, zip, free edition, youtube, downloader, irfanview, google, chrome, adobe, firefox, virtualdj, vlc, iso, cleaner, msn, live, nero, divx, spyware, torrent, activex, flash, trillian, norton, mp3, 2008

Table 1. Computers and antivirus packages.

COMPUTER NAME	ANTIVIRUS SOFTWARE
Gustav	Norman
Ivan	Norton
Katrina	F-Secure
Mitch	Avast!
Andrew	AVG

keywords, which we applied when using file-sharing programs. We compiled the list from the names of the 50 most popular Windows downloads at Download.com on 15 September. We configured each test system with the same search keywords, but we discovered that identical searches frequently yielded different results on different machines. We therefore collected all the downloaded programs in a joint pool and subsequently installed the same programs on all the test systems from that pool.

During the two weeks of exposure, we spent some days on only one source and others using several sources. Even on days when we weren't actively using the computers, they were still connected to the Internet, and Internet Explorer, μ Torrent (a popular client for using the BitTorrent protocol), and LimeWire (a popular client for the Gnutella network) were all running.

File-Sharing Networks

File-sharing networks are a significant source of malware,⁴ so we found it important to expose the experiment to them. We used the keywords from the sidebar ("Keywords Used for Data Collection") to search for candidate files. Although the LimeWire client accesses a defined file-sharing network (Gnu-

tella), there is, strictly speaking, no such thing as a BitTorrent network because users share files via individual "tracker" hosts. We decided to use btjunkie (<http://btjunkie.org>), an advanced BitTorrent search engine that uses a Web crawler to search for torrent files from other BitTorrent trackers. Btjunkie's search mechanisms made it easier to search for new files.

Surfing the Web

A disturbing amount of Web sites contain adware, viruses, and other threats.^{2,3} Armed with this knowledge, and with the help of search engines—such as Google and Yahoo—fed with popular search phrases, we came up with a list of possibly malicious sites. The list was a mixture of popular ordinary Web sites and sites that claim to offer downloadable items, such as warez, screensavers, and MP3s. When visiting Web sites, the integrated browser Internet Explorer 7 was a natural choice because it's still the most widely used.

We employed the following strategy:

- We started at the top of our list and visited the Web sites one by one.
- Because our goal was to be exposed to as much malware as

possible, we acted like a naive (but enthusiastic) user, uncritically clicking OK to everything that popped up.

- We installed missing plug-ins, such as Flash, on demand, without regard to whether they'd been signed by a trusted entity.
- If the particular Web site had partner sites or other tempting links, we paid them a visit, too.
- When visiting warez sites, where we could download software, for instance, we typically chose a few of the most popular items and saved them to a directory on the computer for later analysis.

We admit that it's unlikely that a regular user would behave exactly like we did; on the other hand, we're confident that each of our actions, in isolation, could have been performed by some user out there on the Internet.

Results

Our experiment resulted in 124 0-day malware instances—that is, malware that neither F-PROT nor Avast! BART found immediately after the exposure period but that the software detected in a new scan with updated signatures after the one-month dormant period. All the files Avast! BART and F-PROT detected were subsequently uploaded and scanned at VirusTotal (www.virustotal.com), which uses 36 different updated antivirus engines.

From the VirusTotal results, we could check whether other antivirus engines detected the same files as Avast! BART and F-PROT. We chose to study the results from F-Secure and Symantec further because they're big antivirus solution vendors, and they also had good descriptions of the different malware types.

Based on the dates associated with each signature on VirusTotal, we can verify that neither F-Secure, Symantec, nor both used together would have detected

many of the malware instances at the end of the exposure period. If F-Secure reports that it's added the signature at some date in October, then it didn't detect the malware during the baseline scan (which we did on 1 October). We gathered all files in our experiment during September, which means they'd existed for more than one month at the time of our final scan. It's disquieting that so many antivirus engines didn't detect that these files were malicious, even though they'd been in the wild for such a long time.

Most of the 0-day malware came from using the btjunkie search engine and BitTorrent to download and run (possibly pirated) programs. Although we estimate the percentage of 0-day malware specimens that originate from btjunkie/BitTorrent at 47 percent, such malware from the Gnutella network constitutes only 7 percent.

Forty-two percent of the discovered 0-day malware specimens have unknown origins. Some might have come from Web surfing or clicking on different pop-ups and ads, and some might have been created when we installed downloaded files. We determined that at least 4 percent of the 0-day malware came from files downloaded while Web surfing.

Table 2 estimates what percentage of files downloaded through btjunkie, Gnutella, and the Web contained 0-day malware. Because we configured all desktop antivirus software to remove known malware instances as the software detected them, we don't have the exact number of downloaded files, and these percentages are a rough estimate. We downloaded roughly 400 files using BitTorrent because we had 40 keywords that, on average, resulted in 10 downloads apiece. The Gnutella number is more difficult to estimate, but the percentage of 0-day malware was still clearly significantly higher in the

Table 2. Approximate percentage of 0-day malware from different sources.

SOURCE	DOWNLOADED FILES	0-DAY MALWARE	%
Btjunkie (BitTorrent)	~400	58	14.5
Gnutella	~6,000	9	0.15
Web downloads	~80	5	6.25
Unknown source	?	52	?

torrents included in the btjunkie search engine than in the Gnutella network. This is partly due to the difference in the search mechanisms in these networks.

It's worth mentioning that the files downloaded from the Web are typically from sites Google and Yahoo search engines have identified as suspicious. So, if we had looked at files downloaded from the Web in general, certainly a lot fewer than 6 percent would have contained 0-day malware; what the table indicates, however, is that such malware exists in all these areas. Also, note that we found a significant number of files with malware elsewhere on the test computers; we can only assume that (some of) these files were downloaded by spyware contracted during the experiment.

Discussion

Although 0-day malware's existence is well known, few can refer to actual numbers regarding its prevalence. As mentioned, we identified 124 unique files as infected with 0-day malware after exposing the laboratory PCs to a broad range of suspicious material. A normal user would probably not manage to expose his or her computer to the same amount in such a short time frame, but a normal user has a much longer exposure period (that is, continuous and never ending).

Our study illustrates that the risk of getting infected by malware

that antivirus protection doesn't detect is alarmingly high. New malware that the antivirus engines don't have signatures for is likely to escape detection by a desktop antivirus solution. Taking precautions while using the Internet can protect users only to a certain extent. If they visit the wrong Web site or download a file infected with 0-day malware, they probably won't be protected from infection.

The malware specimens that our antivirus packages didn't detect during our two-week exposure period suggest to us that signature-based antivirus software doesn't provide sufficient protection for users who live on the bleeding edge with respect to where they obtain their software. Coupled with the exponential growth of new malware variants (see Figure 1), our findings suggest that antivirus vendors have major problems keeping the signature lag within acceptable limits. Continued research will be needed to combat the virus threat in the years to come. □

Acknowledgments

This installment of Attack Trends is based on the results of a minor thesis at the Norwegian University of Science and Technology (NTNU).

References

1. H. Vegge et al., "Where Only Fools Dare to Tread: An Em-

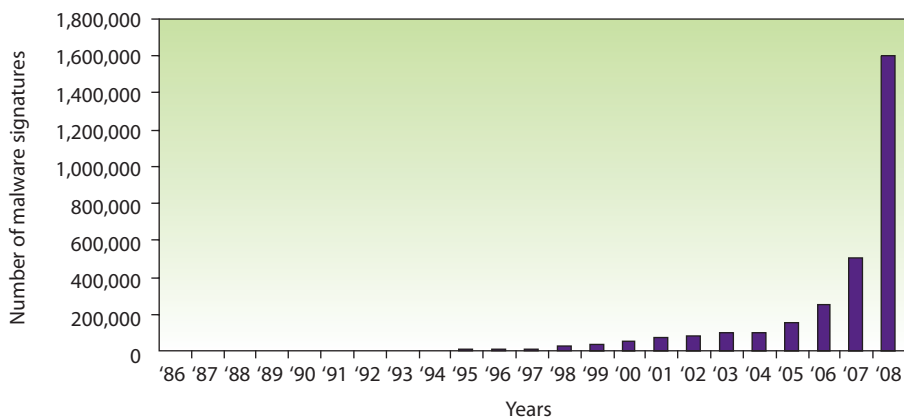


Figure 1. Accumulated number of malware signatures in F-Secure's databases from 1986–2008. The number of new malware signatures has risen exponentially in the past few years, and this is becoming a major challenge to antimalware vendors. (This figure is based on "F-Secure IT Security Threat Summary for the Second Half of 2008," www.f-secure.com/2008/2/index.html.)

pirical Study on the Prevalence of Zero-Day Malware," to be published in *Proc. 4th Int'l Conf. Internet Monitoring and Protection (ICIMP 09)*, 2009.

2. D. Nunes and S. Keats, "Mapping the Mal Web," McAfee

SiteAdvisor, 12 Mar. 2007; www.siteadvisor.com/studies/map_malweb_mar2007.html.

3. S. Keats, "Mapping the Mal Web Revisited," McAfee SiteAdvisor, 4 June 2008; www.siteadvisor.com/studies/map_malweb_jun2008.pdf.

4. S. Shin, J. Jung, and H. Balakrishnan, "Malware Prevalence in the Kazaa File-Sharing Network," *Proc. 6th ACM SIGCOMM Conf. Internet Measurement (IMC 06)*, ACM Press, 2006, pp. 333–338.

Martin Gilje Jaatun is a research scientist at SINTEF Information and Communication Technology, where he heads the Information Security Research Group in the Department of Software Engineering, Safety and Security. His research interests include malware, software security, and communications security. Jaatun has an MSc in telematics from the Norwegian Institute of Technology. Contact him at martin.g.jaatun@sintef.no.

Jostein Jensen is a research scientist at SINTEF Information and Communication Technology. His research interests include secure software development and security in service-oriented architectures. Jensen has an MSc in communication technology from the Norwegian University of Science and Technology. Contact him at jostein.jensen@sintef.no.

Håvard Vegge is an MSc student in communication technology at the Norwegian University of Science and Technology. He's currently working on his master's thesis on secure multiparty computations. Contact him at havardv@stud.ntnu.no.

Finn Michael Halvorsen is an MSc student in communication technology at the Norwegian University of Science and Technology. He's currently working on his master's thesis on cryptanalysis of the Temporal Key Integrity Protocol. Contact him at finnmich@stud.ntnu.no.

Rune Walsø Nergård is an MSc student in communication technology at the Norwegian University of Science and Technology. He's currently working on his master's thesis on the security of administrator passwords on popular operating systems. Contact him at runewals@stud.ntnu.no.

ADVERTISER INFORMATION • MARCH/APRIL 2009

Advertiser	Page	Advertising Sales Representatives	Product:
Usenix 2009	Cover 4	Midwest/Southwest Darcy Giovingo Phone: +1 847 498 4520 Fax: +1 847 498 5911 Email: dg.ieeemedia@ieee.org	US East Joseph M. Donnelly Phone: +1 732 526 7119 Email: jmd.ieeemedia@ieee.org
Advertising Personnel Marion Delaney IEEE Media, Advertising Dir. Phone: +1 415 863 4717 Email: md.ieeemedia@ieee.org		Northwest/Southern CA Tim Matteson Phone: +1 310 836 4064 Fax: +1 310 836 4067 Email: tm.ieeemedia@ieee.org	US Central Darcy Giovingo Phone: +1 847 498 4520 Fax: +1 847 498 5911 Email: dg.ieeemedia@ieee.org
Marian Anderson Sr. Advertising Coordinator Phone: +1 714 821 8380 Fax: +1 714 821 4010 Email: manderson@computer.org		Japan Tim Matteson Phone: +1 310 836 4064 Fax: +1 310 836 4067 Email: tm.ieeemedia@ieee.org	US West Lynne Stickrod Phone: +1 415 931 9782 Fax: +1 415 931 9782 Email: ls.ieeemedia@ieee.org
Sandy Brown Sr. Business Development Mgr. Phone: +1 714 821 8380 Fax: +1 714 821 4010 Email: sb.ieeemedia@ieee.org		Europe Hilary Turnbull Phone: +44 1875 825700 Fax: +44 1875 825701 Email: impress@impressmedia.com	Europe Sven Anacker Phone: +49 202 27169 11 Fax: +49 202 27169 20 Email: sanacker@intermediapartners.de
		New England John Restchack Phone: +1 212 419 7578 Fax: +1 212 419 7589 Email: j.restchack@ieee.org	
		Southeast Thomas M. Flynn Phone: +1 770 645 2944 Fax: +1 770 993 4423 Email: flynntom@mindspring.com	