

A Crawler-based study of Spyware on the web

Presented at NDSS 2006

Alex Moshchuk, Tanya Bragin,
Steve Gribble, Hank Levy
(University of Washington)

By Shvetank Jain
Tuesday, October 17, 2006
ECE 1776

Spyware today

- Most Internet PCs have, or have had, spyware
- Privacy of victims compromised
- Little quantitative data on extent of spyware

Study Plan

- Two methods in which spyware infects:
- Spyware piggy-backed on executables
 - E.g., Kazaa ships bundled with multiple spyware programs
- Drive-by download installation
 - Malicious web content exploits browser flaws to install spyware

Crawling for executables

- Defined 10 interesting Web categories
 - E.g., games, news, celebrities, pirate, wallpaper
- For each category:
- Used Google to identify several hundred domains
- Crawled each domain (to depth 3) to find executables
- Downloaded executables for offline analysis
- Crawled about 20 million URLs over 2,500 domains
- Collected 20,000 executables
- 19% of domains had downloadable executables

Analyzing executables

For each executable:

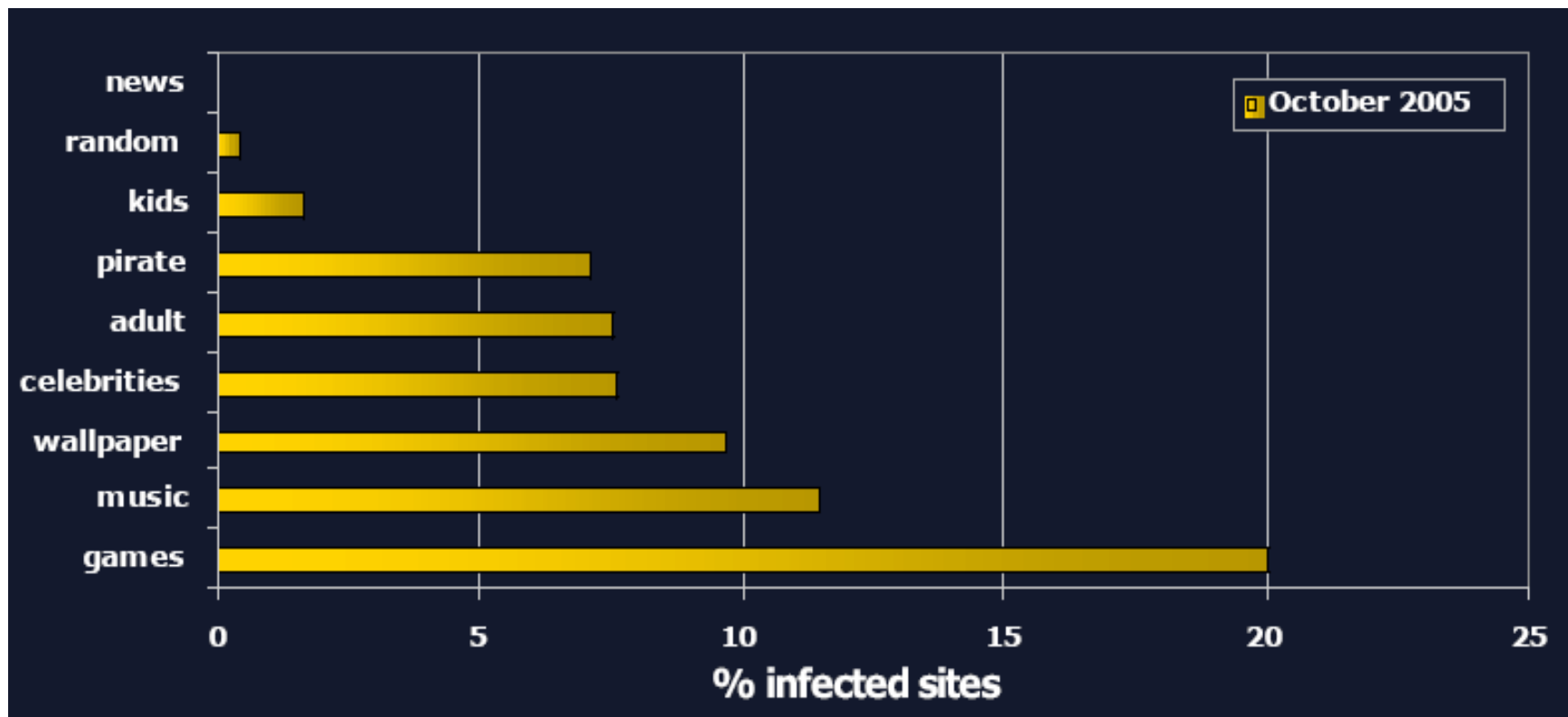
- Cloned a clean WinXP virtual machine (VMware)
- Automatically installed the executable into the VM
- Ran an anti-spyware tool (Lavasoft Ad-Aware) to look for infections
- Automating installation required some heuristics
 - E.g., pressing “Next,” agreeing to EULAs, ...
- An executable is *infected* if Ad-Aware finds spyware

High-level results

- Found a lot of piggy-backed spyware
 - 1 in 20 executables contained spyware
 - 1 in 25 domains were infectious
- Observed few spyware variants
- Encountered 1,294 infected executables but only 89 spyware programs
- No significant change in amount of piggy-backed spyware from May 2005 to October 2005

Where is the spyware found?

- Spyware is concentrated on specific popular Web zones

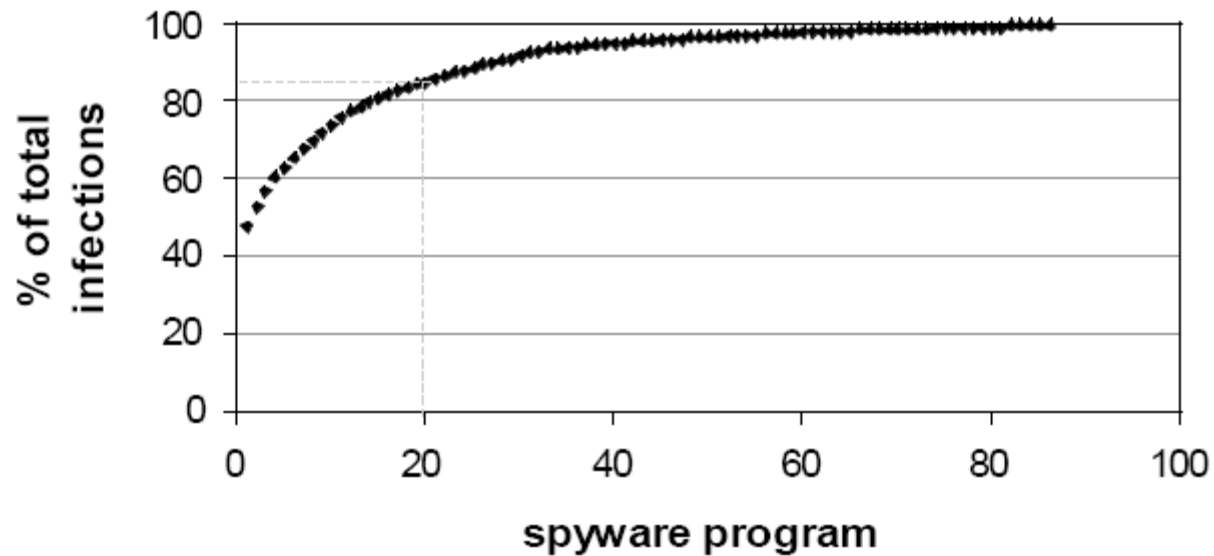


How is spyware distributed across sites?

- A small no. of sites have a large no. of infected executables
- Easy to detect and blacklist
- Top spyware sites (# infected executables)
 - scenicreflections.com (503)
 - gamehouse.com (164)
 - screensavershot.com (137)
 - screensaver.com (107)

Distribution of spyware programs

- A few offenders are responsible for most infected executables
- Signature-based detection should be effective



What kinds of spyware do we find?

- Measured the prevalence of five spyware functions:
 - Adware and browser hijackers are most common (86%)
 - Trojan downloaders pose a risk (13%)
 - Keyloggers and dialers are more rare (1%)

Drive-by download study

- Web content exploits browser flaws to install spyware
- Victims are infected just by visiting a malicious page
- Detect attacks as they happen in practice
 - Crawl the Web categories
 - Render each page in an unmodified Web browser inside a clean VM
 - Internet Explorer (6.0, unpatched)
 - Mozilla Firefox (1.0.6)
 - Run anti-spyware check to look for spyware

Using Event Triggers

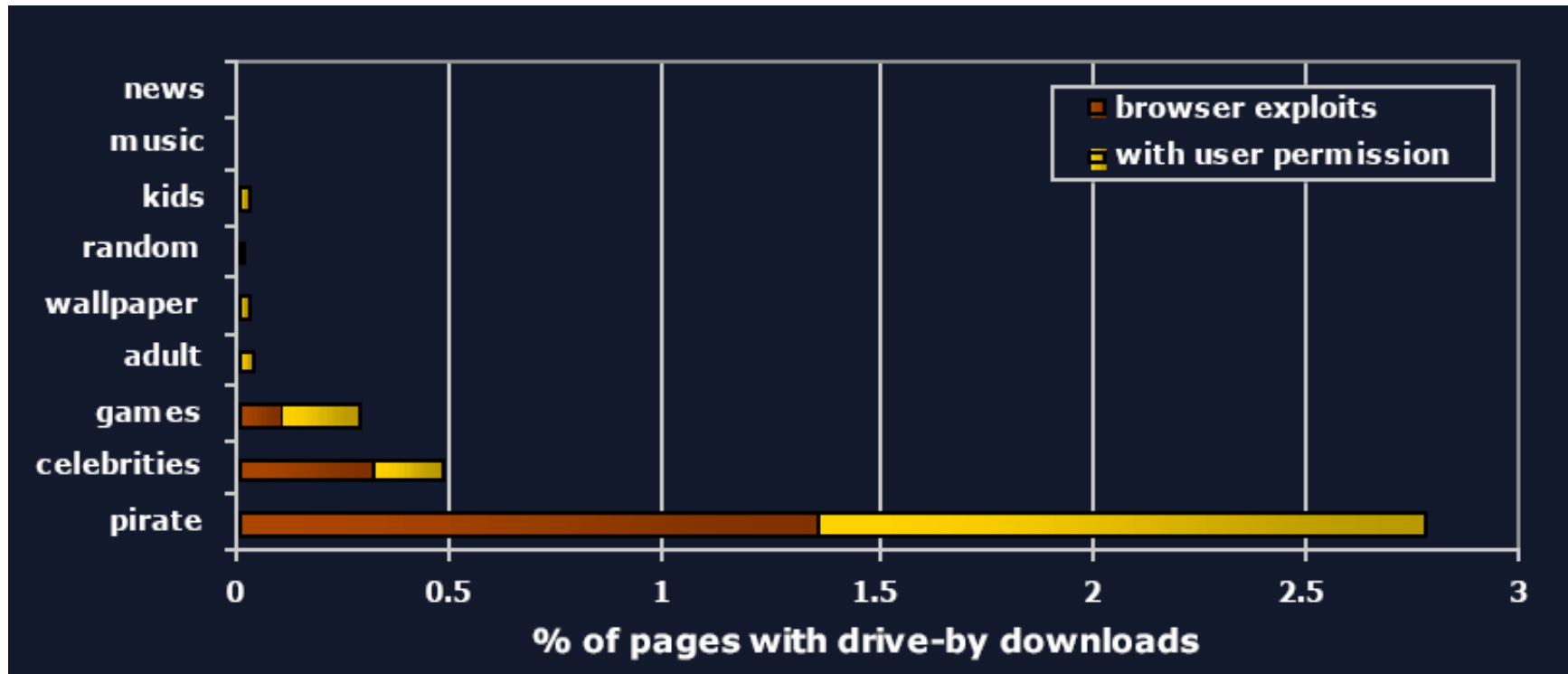
- Event triggers are a performance optimization
- Triggers detect suspicious activity
 - Process creation
 - Suspicious registry modifications
 - Files written outside browser temp. folders
- Run Ad-Aware check *only* when a trigger fires
 - No false negatives
 - 41% false positives
 - Benign software installations
 - Background noise
 - Spyware not detected by Ad-Aware

High-level results

- There are many Web pages(0.4%) with drive-by downloads
- 50% of attacks exploited browser flaws
- Little variation
 - Only 36 spyware programs responsible for 186 attacks
- The number of pages with drive-by downloads is decreasing

Where are drive-bys found?

- Non-uniform distribution



Is the Firefox browser susceptible?

- Successful drive-by downloads appeared on 0.08% of pages
 - All require user consent
 - All are based on Java
- Firefox is *not* 100% safe, but it is *safer to use* than IE
 - Found 13 times more attacks for IE than for Firefox

Related Work

- Honeypots
- Strider HoneyMonkey
 - Tool to find Web pages with browser exploits
 - Method similar to the trigger-based VM approach
- Webroot Phileas, Sunbelt
 - Automated web crawling for new spyware variants
- SiteAdviser
 - Upcoming commercial service to rate safety of Web sites

Pros

- First realistic attempt to quantify the prevalence of spyware on the web (although its only a sample)
- Some of the results obtained are interesting
 - A substantial number of pages exploited browser flaws
 - Web infection is diminishing (in the sample)
 - A lot of popular spyware programs, infectious sites were singulated
 - Signature based techniques may be effective

Pros

- Blacklist approach may not work well
- Spyware attack sites have become sophisticated – browser sniffing

Cons

- Decline of spyware from April to October may be due to different reasons
 - Different versions of Anitspyware tools
 - Increased adoption of anti spyware tools by sites
 - Increased usage of automated patch install tools to manage patch upgrades
 - Social reasons: lawsuits discouraging spyware distributors

Cons

- Spyware is ignored if it reveals after a long time
 - What if JavaScript in the browsed page triggers a spyware after time $t(15s)$?
 - What if Javascript calls some spyware when the page closes?
 - What if JavaScript pops up some windows which in turn can lead to spyware

Cons

- The study focuses mainly on windows executables.
 - Study deliberately choose unpatched versions of XP to capture majority of existing exploits
 - Hence it may not be appropriate to say Firefox is more secure than IE
 - “Study found 13 more attacks for IE than Firefox”
- Study failed to extrapolate the relationship between the density of spyware on the web and the presence of spyware on the desktop.

Cons

- For drive-by download attacks, triggers limit how much spyware is missed
 - Upper bound: 41% false positives when a trigger fires
 - Benign software installations
 - Background noise
 - Spyware not detected by AdAware
- Adaware dependency
- The study did not crawl the entire web.
 - Though crawl didn't make use of any false positives, it missed some real executables.
 - Also missed dynamic URL executables constructed by Javascript



Q & A