Spyware

Steven Gribble

Department of Computer Science and Engineering
University of Washington

kingsofchaos.com

- A benign web site for an online game
  - earns revenue from ad networks by showing banners
  - but, it relinquishes control of the ad content

kingsofchaos.com

- A benign web site for an online game
  - earns revenue from ad networks by showing banners
  - but, it relinquishes control of the ad content

banner ad from
adworldnetwork.com
(a legitimate ad network)

inline javascript loads
HTML from ad provider

Incident

- kingsofchaos.com was given this “ad content”

```
<script type="text/javascript">document.write(‘
\u003c\u0062\u006f\u0064\u0079\u0020\u006f\u006e\u0055
\u006f\u0077\u0050\u006f\u0070\u0075\u0070\u0028\u0029
\u003b\u0073\u0068\u006f\u0077\u0048\u0069 …etc.
```

- This “ad” ultimately:
  - bombarded the user with pop-up ads
  - hijacked the user’s homepage
  - exploited an IE vulnerability to install spyware

What's going on?

- The advertiser was an ex-email-spammer
- His goal:
  - **force** users to see ads from his servers
  - **draw revenue** from ad "affiliate programs"
    - Apparently earned several millions of dollars
- Why did he use spyware?
  - control PC and show ads even when not on the Web

Take-away lessons

- Your PC has value to third parties
  - spyware tries to steal this value from you
    - adware: eyeballs and demographic information
    - spyware: sensitive data, PC resources
- Web content should never be trusted
  - even if its direct provider is
- Consumer software and OSs are weak
  - browsers are bug-ridden
  - OSs do not protect users from malicious software
    - yet, this is increasingly the world we live in

Outline

- Background
- Measurement study
- Discussion on spyware mitigation

Outline

- Background
  - definitions
  - trends
  - defenses
- Measurement study
- Discussion on spyware mitigation

What is spyware?

- Incredibly difficult to define “spyware” precisely
  - no clean line between good and bad behavior
- Spyware is a software parasite that:
  - collects information of value and relays it to a third party
  - hijacks functions or resources of PC
  - installs surreptitiously, without consent of user
  - resists detection and de-installation
- Spyware provides value to others, but not to you

How one becomes infected

- Spyware piggybacked on executables
  - model for profiting from free software
  - e.g., Kazaa installed 2-7 adware programs
- Drive-by downloads
  - Web site attempts to install software through browser
  - may involve exploiting browser vulnerabilities
- Trojan downloaders / “tricklers”
  - spyware that fetches additional spyware
  - snowball effect

Types of spyware

Class	# signatures
Cookies and web bugs	47
Browser hijackers	272
Adware	210
Keyloggers	75
Dialers	201
Backdoors / trojans / tricklers	279

From the “Spybot S&D” database, Feb. 2005

Spyware trends

- Most Internet PCs have, or have had, it
  - 80% of Internet-connected PCs are infected
  - *[AOL/NCSA online safety study, Oct. 2004]*
- Much of the Web has it
  - 1 in 8 executables on Web piggyback spyware
  - 0.1% of random Web pages try “drive-by” installs
  - *[UW study, Oct. 2005]*
- Convergence of threats
  - worms, viruses, spyware, botnets are fusing
  - e.g., many spyware programs now install spam relays

Industrial responses

- Anti-spyware tools
  - predominantly signature based
  - e.g., AdAware, Spybot S&D, Microsoft AntiSpyware
- Blacklisted URLs in firewalls, NIDS
  - e.g., UW tipping point machine
- Sandboxes for isolating untrusted content
  - e.g., GreenBorder

Legislative responses

- Federal “SPY ACT”
  - Oct. 6: passed in House, received in Senate
  - lists prohibited software functions
    - e.g., “Modifying settings related to use of the computer or to the computer's access to or use of the Internet by altering (A) the Web page that appears when the owner or authorized user launches an Internet browser or similar program used to access and navigate the Internet, (B) …”
  - requires user consent to “information collection programs”
    - required functions for such programs, e.g., easy to disable
  - list of exclusions
    - law enforcement, ISPs, diagnostic and security software/services, good samaritan protection, manufacturers and retailers providing third party branded software
  - has big teeth
    - up to $3,000,000 penalty per violated provision

Outline

- Background
- Measurement study
  - "A Crawler-based Study of Spyware in the Web"
    - Alex Moshchuk, Tanya Bragin, Steven D. Gribble, and Henry M. Levy. To appear, NDSS 2006.
- Discussion on spyware mitigation

Measurement study

- Understand the problem before defending against it
- Many unanswered questions
  - What's the spyware density on the web?
  - Where do people get spyware?
  - How many spyware variants are out there?
  - What kinds of threats does spyware pose?
- Answers give insight into what defenses may work

Approach

- Large-scale measurement of spyware on the Web
  - crawl “interesting” portions of the web
  - download content
  - determine if content is malicious

- Two parts
  - Executable study
    - Find executables with known spyware
  - Drive-by download study
    - Find web pages that attempt drive-by download attacks

Analyzing Executables

- Web crawler collects a pool of executables
- For each:
  - clone a clean virtual machine
    - 10-node VM cluster, 4 VMs per node
  - scripted install of executable
  - run analysis to see what changed
    - currently, we use an anti-spyware tool (Ad-Aware)
- Average analysis time – 90 sec. per executable

Analyzing Drive-by Downloads

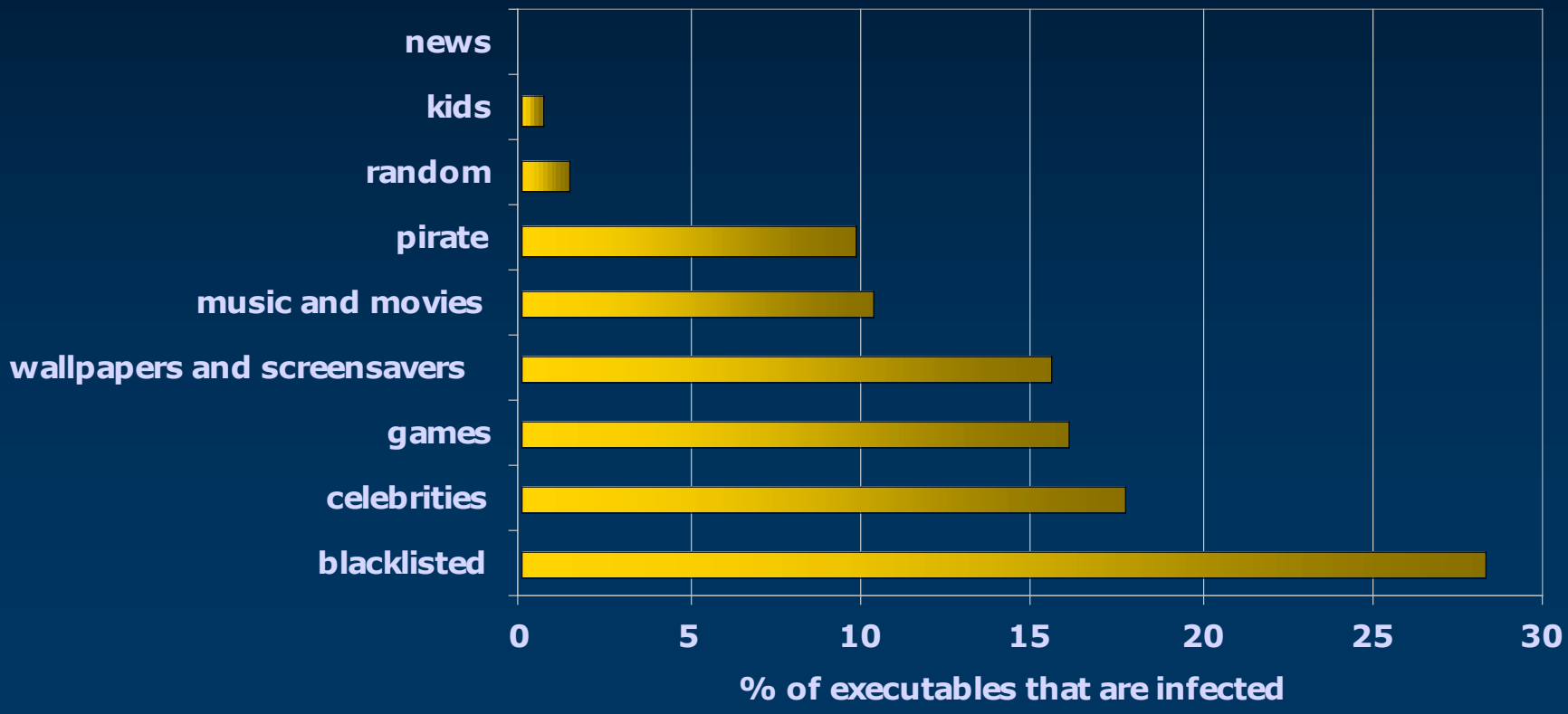
- Evaluate the safety of browsing the web
- Automatic “virtual browsing”
  - render pages in a real browser inside clean VM
    - unpatched Internet Explorer on unpatched Windows XP
  - define triggers for suspicious browsing activity
    - process creation
    - files written outside browser temp. folders
    - suspicious registry modifications
  - run anti-spyware check only when trigger fires
- (c.f. Honeymonkey work, concurrent with ours)

Executable Study Results

- Crawled 32 million pages in 10,000 Web domains
- Downloaded 26,000 unique executables
- Found spyware in 13.5% of them
  - most installed only one spyware program
    - 6% installed three or more spyware variants
  - 142 unique spyware threats

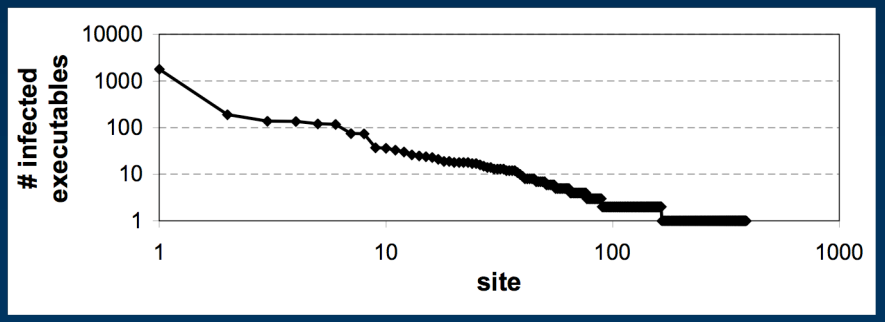
Infection of Executables

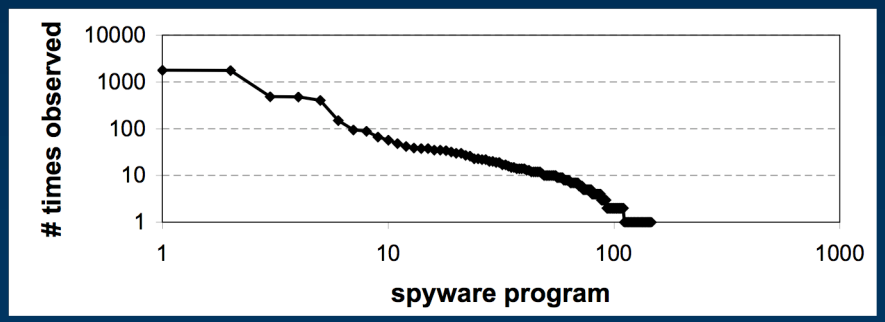
- Visit a site and download a program
- What's the chance that you got spyware?



Spyware popularity

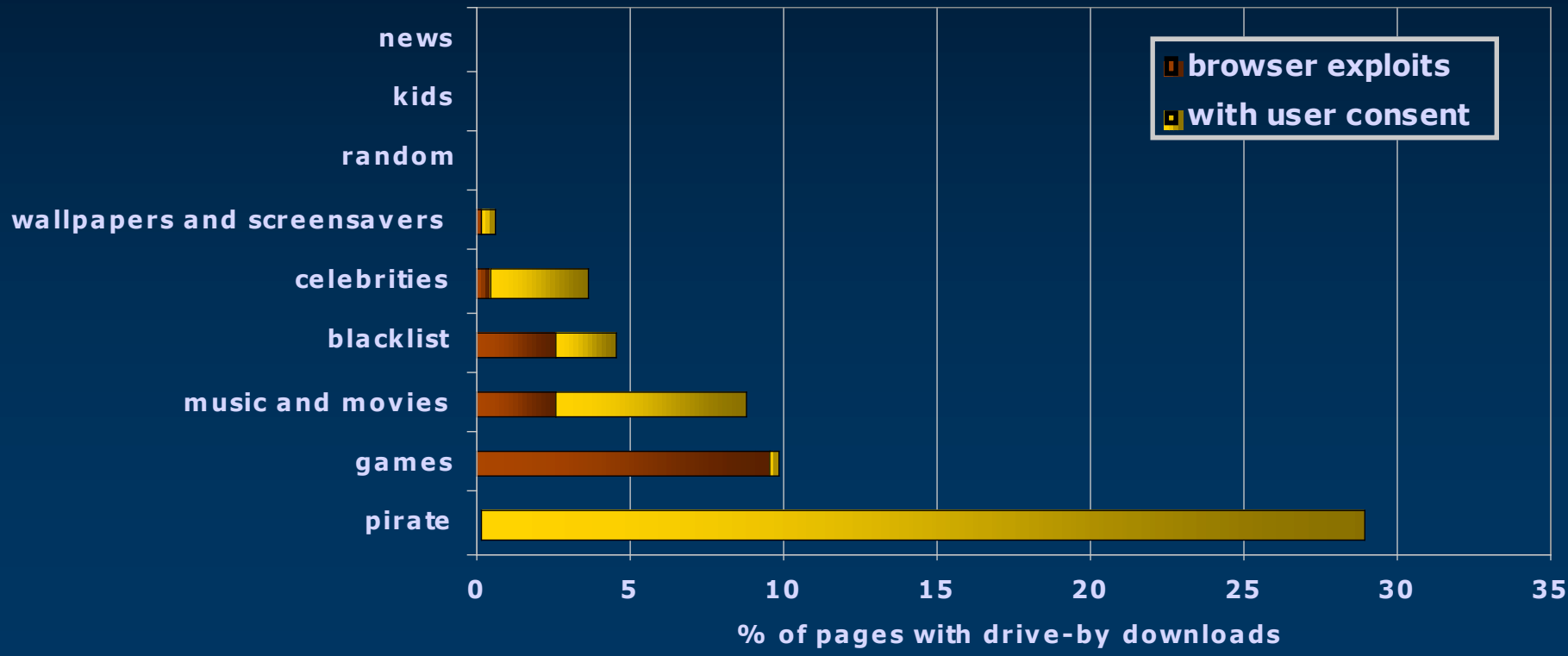
- Spyware popularity is (surprise, surprise) Zipfian
- A small # of spyware variants are found frequently
  - top 28 variants account for 90% of infected execs.
  - WhenU, eZula, 180Solutions at top of list
- A small # of sites have large # of infected execs.





Drive-by Download Results

- 5.5% of pages we examined carried drive-by downloads
  - 1.4% exploited browser vulnerabilities



Types of spyware

- Five oft-discussed spyware functions
  - What's the chance a spyware program contains each function?

	Executables	Drive-by Downloads
Keylogger	0.05%	0%
Dialer	1.2%	0.2%
Trojan Downloader	12%	50%
Browser hijacker	62%	84%
Adware	88%	75%

Summary

- There is plenty of spyware on the web
  - 1 in 8 programs is infected with spyware
- Spyware targets specific popular content
  - 0.1% of random web pages try drive-by downloads
  - 5% of “celebrity” web pages try drive-by downloads
- Most spyware is just annoying (adware)
  - but a significant fraction poses a big risk
- Few spyware variants are encountered in practice

Outline

- Background
- Measurement study
- Discussion on spyware mitigation
  - the “opinion” part of this talk

My view on the problem

- Spyware separable into two “classes” of problem
- Shucksters out for a quick buck
  - taking advantage of current blurry legal status of spyware
  - tweak and distribute off-the-shelf adware
    - rarely engineer new code
    - goals: “throw it far and wide, make it stick”
  - responsible for most of what’s out there
- Determined criminals
  - phishers/pharmers looking for credit card numbers
  - keyloggers for personal/corporate espionage
  - may be willing to engineer boutique spyware software

How to stop the shucksters

- Legislation helps take away incentive
  - makes it clear what is illegal
  - legit companies will clean up their act
- Anti-spyware tools deal well with remainder
  - you're really paying for the top ~50 signatures
  - new threats emerge from time to time
    - need engineers to keep rules fresh
    - seems no different than antivirus signature problem

The criminals

- We're not well prepared for this threat
  - regular users have poor model of safe vs. risky
    - and savvy users don't have good tools for coping
  - OSs built as single trust domain; if compromised, lose
    - no firewall between Internet-facing code and your stuff
- Maybe we just need "street smart" mechanisms
  - help users avoid sketchy parts of the Web
    - Blacklists? Reputation-based schemes?
  - help users keep valuables locked up
    - Lampson's "red vs. green" VMs, GreenBorder

Advanced techniques

- Rejigger OS so harder for users to add new code
  - \+ less likely to get unwanted code
  - — makes it hard to add legitimate apps
  - — doesn’t help with scripts / bytecode
- Semantic analysis (look for spyware-like behavior)
  - \+ fewer signatures needed, higher leverage in arms race
  - — too many ways to do the same thing in today’s systems
  - — prone to false positives

Questions?