The Ghost In The Browser
Analysis of Web-based Malware

Niels Provos
Dean McNamee
Panayiotis Mavrommatis
Ke Wang
Nagendra Modadugu
Overview

• Introduction
• Detecting Malicious Pages
• Content Control
• Malware Trends
• Conclusion
Introduction

- Internet essential for everyday life: ecommerce, etc.
- Malware used to steal bank accounts or credit cards
  - underground economy is very profitable
- Internet threats are changing:
  - remote exploitation and firewalls are yesterday
- Browser is a complex computation environment
- Adversaries exploit browser to install malware
Introduction

• To compromise your browser, we need to compromise your web server

• Very easy to set up new site on the Internet

• Very difficult to keep new site secure
  
  • insecure infrastructure: Php, MySql, Apache
  
  • insecure web applications: phpBB2, Invision, etc.
Detecting Malicious Websites

- Malicious website automatically installs malware on visitor’s computer
  - usually via exploits in the browser or other software on the client (without user consent)
- Using Google’s infrastructure to analyze several billion URLs.
Detecting Malicious Websites

Web Page Repository

MapReduce Heuristical URL Extraction

Virtual Machine

Internet Explorer

Monitor Execution Analysis

URL

Result

Malicious Page Repository
Processing Rate

- The VM gets about **300,000** suspicious URLs daily
- About **10,000** to **30,000** are malicious
Content Control

• what constitutes the content of a web page?
  • authored content
  • user-contributed content
  • advertising
  • third-party widgets

• ceding control to 3rd party could be a security risk
Web Server Security

- compromise web server and change content directly
- many vulnerabilities in web applications, apache itself, stolen passwords
- templating system
Advertising

- by definition means ceding control of content to another party
- web masters have to trust advertisers
- sub-syndication allows delegation of advertising space
- trust is not transitive
Third-Party Widgets

- to make sites prettier or more useful:
  - calendaring or stats counter
- search for **praying mantis**
  - linked to free stats counter in 2002 via Javascript
  - Javascript started to compromise users in 2006

http://expl.info/cgi-bin/ie0606.cgi?homepage
http://expl.info/demo.php
http://expl.info/cgi-bin/ie0606.cgi?type=MS03-11&SP1
http://expl.info/ms0311.jar
http://expl.info/cgi-bin/ie0606.cgi?exploit=MS03-11
http://dist.info/f94mslrfum67dh/winus.exe
Malware Trends and Statistics

- Avoiding detection
  - obfuscating the exploit code itself
  - distributing binaries across different domains
  - continuously re-packing the binaries

```javascript
document.write(unescape("\x3CHEAD\x3E\x0D\x0A\x3CSCRIPT\x20
LANGUAGE\x3D\x22Javascript\x22\x3E\x0D\x0A
\x20\x20\x20\x20\x20\x20\x20\x20\x20/*%20criptografado%20pelo%20Fal%20-%20Deboa%E7%E3o
\x20gr%Eltis%20para%20seu%20site%20renda%20extra%0D
\x0D%0A\x20\x20\x20\x20\x20...\n\x3C/SCRIPT\x3E\x0D\x0A\x3C/HEAD\x3E\x0D\x0A\x3CBODY\x3E%0D%0A
\x3C/HTML%0D%0A"));
//-->
</SCRIPT>
```
Malware Classifications

![Graph showing unique URLs discovered over time for Adware, Unknown, and Trojan categories. The x-axis represents dates from January 11 to March 21, and the y-axis represents the number of unique URLs discovered in logarithmic scale.]
Remotely Linked Exploits

- Exploits are leveraged across many sites
- Popular exploits are linked from over 10,000 URLs
Discussion

- increase of web-based exploitation over time
- installed malware allows for remote control
- observed botnet like structures:
  - pull-based: frequently checking for new commands
  - observed user agents such as: DDoSBotLoader
  - binary updates can be interpreted as command & control
Conclusion

• Web-based malware is a real problem
• millions of potentially infected users
• Automatic detection of malicious web pages to secure web search results
• Identified four areas of content control
• Observed botnet-like structures