SAND2011-5024C

Detection of PDF Based Malware

8/4/11

Jesse Cross
CCD Intern

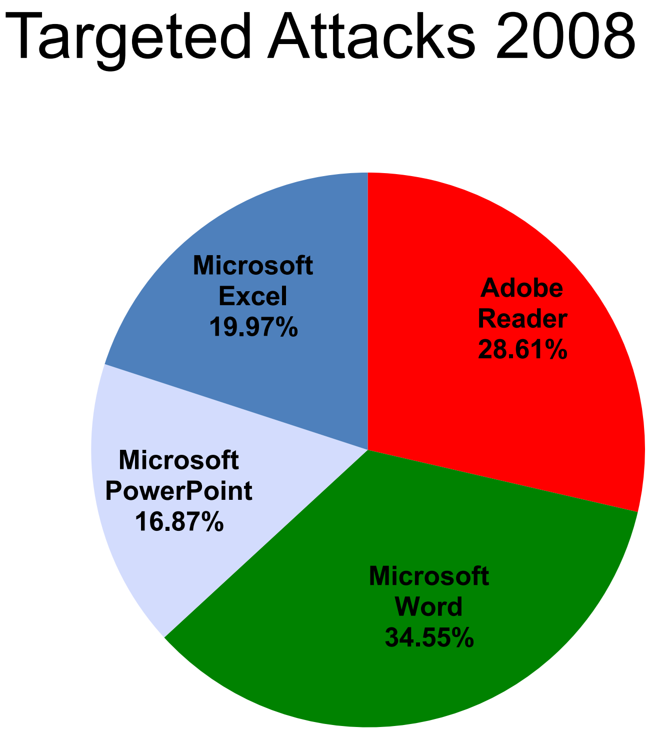
Art Munson
Mentor

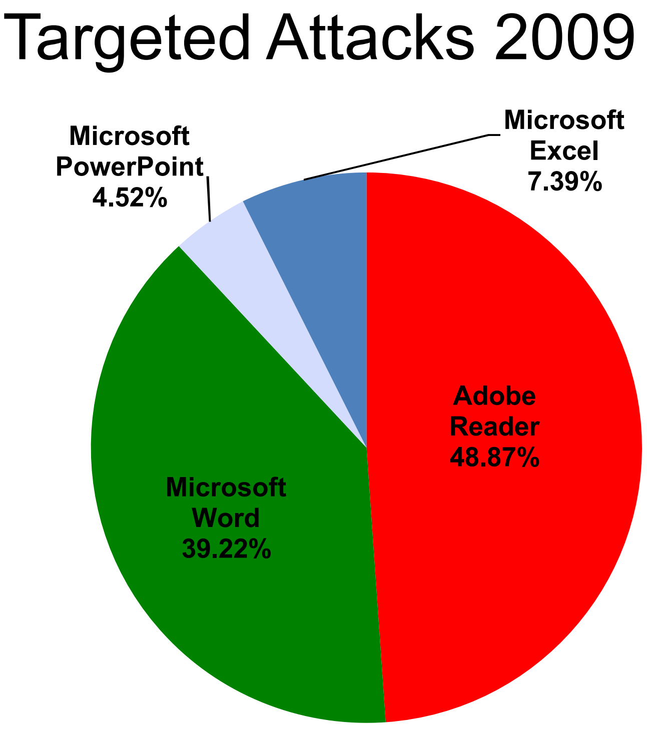
Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation,
a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's
National Nuclear Security Administration under contract DE-AC04-94AL85000.

Introduction

- **Portable Document Format**
- **Widely used**
- **The number of malicious PDFs has been on the rise for the past few years.**
- **The PDF format now contains many vulnerable multimedia functions.**

Rise of PDF-Based Malware





Source: F-Secure. Accessed 7/18/2011.
http://www.f-secure.com/weblog/archives/00001676.html

Overview

- Define and collect features from a group of benign and malicious PDFs in order to train a machine learning algorithm to recognize malicious PDFs.

- Corpus Data:
  - 3490 benign PDFs were captured from Sandia's network traffic
  - 1573 malicious PDFs from offensivecomputing.net

Current Malware Dataset

- **The number of actual malicious PDFs is very small (≈18/1573)**

- **Reports describe much more difficult obfuscation than my malware examples use.**

- **The current malicious PDF samples all use basic filters and JavaScript based obfuscation**

Antivirus Evasion

- Morphologic Manipulation
  - /OpenAction is the same as /Open#41ction
- File Cloaking
  - Adobe Reader ignores data before the header and after the end of file marker
  - Possible to disguise PDF as other file type
- Encryption
  - A null password will allow Adobe Reader to transparently decrypt a PDF
  - Will stop any anti-virus that has no decryption support

Antivirus Evasion (Cont.)

- Filters
  - Streams are compressed to reduce file size
  - Hamper signature based detection
- Forward Compatibility
  - PDF readers are required to ignore anything not recognized so that future versions of PDF may be partially rendered
  - Can be used to hide code

Code Execution

- /Launch
  - – Executes an external script
  - – .exe files are blacklisted
  - – Others such as .py files are not
- /OpenAction, /AA, /Names
  - – Can be used to execute code or a /Launch without explicit user authorization
- /JavaScript and /JS
  - – Used to execute embedded JavaScript code

Work in Progress

- Locate more malicious PDFs
- Instrumenting an open source PDF reader to extract features that may describe a malicious PDF
- The re-instrumented PDF reader will flag more well hidden malicious features than current tools.

Conclusion

- PDF files are very susceptible to malware
- Analysis and detection can be difficult due to malformed and cloaked PDFs
- Will PDF parsing be quick enough to use when scanning network traffic?
- Turn off JavaScript, multimedia functions, and internet access.