Attacks/analysis over PDF files

Javier Diaz Santiso
Juan González Iglesias

CONTENTS

Introduction

Created in 1993 by Adobe Systems

Exchange and print documents regardless of hardware, software and OS

ISO 32000-1

Architecture

Header

Version and format identification

Body

It contains several objects

- Direct and indirect objects
- Dictionary
- Stream object

Cross-reference

Indexes all locations of objects in the file

Trailer

- Specifies how the application reads the document
- File creation and modification dates
- Metadata

How to read a PDF

1. Reads the first bytes of a document
2. Verify PDF version in the header
3. Read the possition of the cross-reference table from the trailer
4. The table provides access to all body objects
5. From these objects the rest of the PDF is processed

Attack vectors in PDF files

- **OpenAction**

  Allows malware to run when opening the pdf.

- **Launch action**

  Allows special commands to be launched. Requires user approval.

- **Embedded files**

  Hide malicious files.

- **GotoEmbedded action**

  Hide malicious pdfs and avoid antivirus detection.

- **URI action**

  Allows access to remote resources.

Obfuscation techniques

Javascript obfuscation

Streams

String manipulation

Recent cases

Advanced RAT functioning as a keylogger and information stealer

Agent Tesla

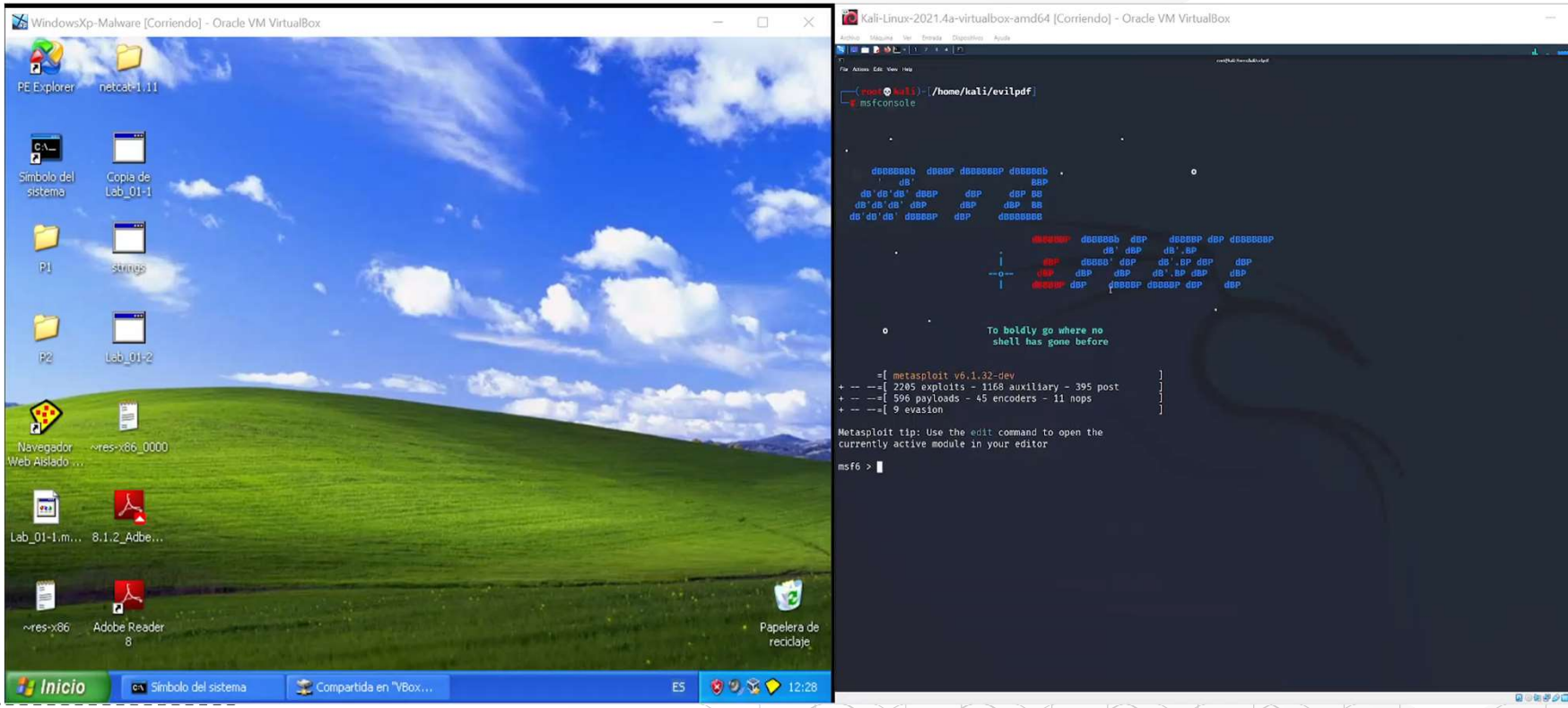
Download the malware via embedded world document macros

Locky ransomware

Malware distributed via email through fake pdfs, encrypting files with long RSA key

CryptoLocker

PDF backdoor demo



PDF malware detection

Static analysis

- Pdfid: Analyze header information
- Pdf-parser: Analyze contents in raw format

Hardware Malware Detection

- Opcode frequency
- Opcode sequence

Dynamic analysis

- API and systems calls
- Dtrace or Strace: locate traces

Machine-learning based techniques

- Using an artificial neural network
- Detect malicius PDFs that have never been seen

Specific PDF analysis

Keyword-based Analysis

- Extract keywords to identify the actions performed by the file.
- PDFiD as a forensic tool

Tree-based Analysis

- Reconstruct the PDF file tree, iterconnections among its objects.
- PeePDF or Origami

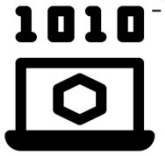
Code-based Analysis

- Analyze embedded scripting code
- PhoneyPDF

PDF malware tools

Pdf Tools

Scan and parse



Peepdf

Command-line Js and shellcode decoder

Pdf Stream Dumper

Explore, decode, deobfuscate

Origami

Automate pdf analysis

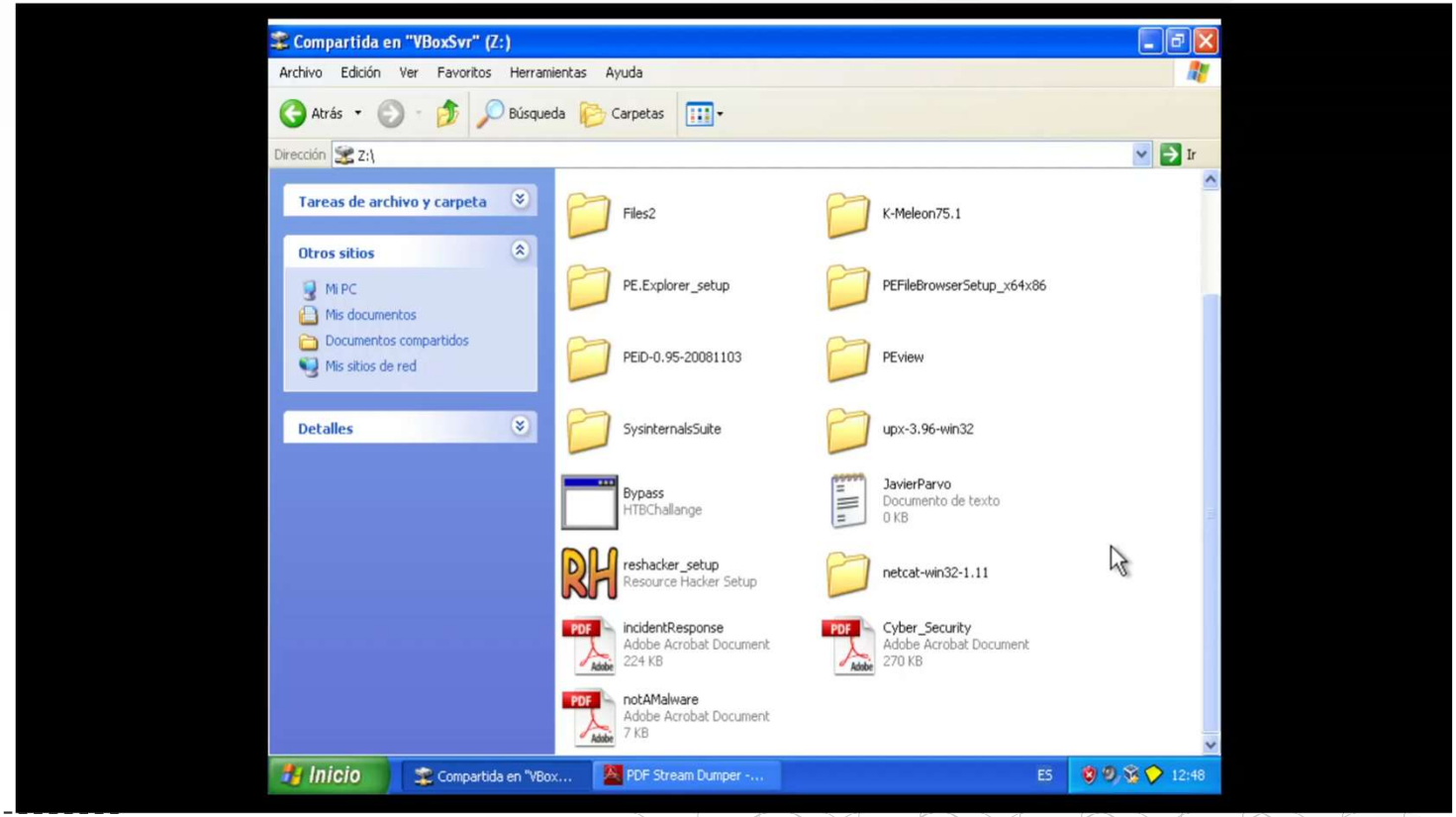
Jsunpack-n

Extract embedded files

MalObjClass

Json pdf representation

PDF analysis demo



PDF Security mechanisms

Application-level security: alert message boxes

Two configuration files *RdLang.32FRA* and *AcroRd32.dll have weakness:*

- *No integrity cheking*
- *Those files are not in read-only acces*

PDF Security mechanisms

Operating system-level security: configuration file and registry keys

- *Internet access:*

  HKU|S-1-5-21-1202660629-706699826-854245398-1003|Software|Adobe| Acrobat Reader|8.0|TrustManager|cDefaultLaunchURLPerms

- *Full screen display:*

  HKU\S-1-5-21-1202660629-706699826-854245398-1003\Software\Adobe\ Acrobat Reader\8.0\FullScreen\iShowDocumentWarning

- JavaScript

  HKU\S-1-5-21-1202660629-706699826-854245398-1003\Software\Adobe\ Acrobat Reader\8.0\JSPrefs

- Opening of appended (embedded or attached) documents

  HKU\S-1-5-21-1202660629-706699826-854245398-1003\Software\Adobe\ Acrobat Reader\8.0\Attachments\cUserLaunchAttachmentPerms\

CREDITS

Javier Díaz Santiso

- Pdf analysis demo
- Analysis mechanims
- Pdf analysis tools
- Introduction and PDF basics

Juan González Iglesias

- Pdf structure
- Pdf backdoor demo
- Pdf vectors
- Introduction and PDF basics

Questions?