這份快速參考概要的給出了逆向分析惡意文檔這些文檔(包括：微軟 office(DOC,XLS,PPT) 和 Adobe Acrobat 的 PDF 文件)的一些技巧和工具。

譯者： wpulog

■ 基本途徑/方法：

1. 找到潛在的被嵌入的惡意代碼，如 shellcode，VBA 宏，或 JavaScript。
2. 從文件中提取可疑的代碼片斷。
3. 如果是 shellcode 則反彙編或調試它。
4. 如果是腳本類代碼如 JavaScript, ActionScript, or VB 宏則對它們解碼剖析。
5. Understand next steps in the infection chain

■ 微軟二進制 office 文件格式說明：

1. 在微軟 office 二進制文件中定義了一個結構化(OLE SS)的文件系統。
2. 數據以目錄「存儲」和文件「流」形式存放(Data can be 「storage」 (folder) and 「stream」 (file))。
3. Excel 將數據存儲在 「workbook」 流中。
4. PowerPoint 將數據存放在「PowerPoint Document」流中。
5. Word 將數據存放在多個流中。

■ 微軟 Office 文件的分析工具：

1. OfficeMalScanner - 定位微軟 office(DOC, XLS, and PPT) 文件中的 shellcode 和 VBA 宏
2. DisView - 在微軟 office 文件指定偏移處反彙編字節碼。(OfficeMalScanner 的一部分)

3. MalHost-Setup - 從微軟 office 文件給定偏移處提取 shellcode，並且能夠將 shellcode 嵌入到 exe 文件中，方便更加深入的分析。 (OfficeMalScanner 的一部分)
4. Offvis - 顯示微軟 office 文件的原始內容和結構，並能鑑別一些常見的 exploit。
5. Office Binary Translator - 轉換 DOC, PPT,和 XLS 文件為 Open XML 文件(包括 BiffView 工具)。
6. OfficeCat 根據一些已知的漏洞在微軟 office 文件中掃瞄嵌入的利用程序(exploit)。
7. FileHex (不免費) and FileInsight 十六進制編輯器,能夠解析和編輯 OLE 結構。

■ 有用的微軟 Office 文件分析命令：

- OfficeMalScanner *file.doc* scan brute ➢ 定位 shellcode, OLE 數據, PE 文件
- OfficeMalScanner *file.doc* info ➢ 定位 VB 宏代碼
- OfficeMalScanner *file.docx* inflate ➢ 解壓縮 *file.docx* ，定位 VB 宏代碼 (XML files)
- DisView *file.doc* 0x4500 ➢ 在文件的 0x4500 處反彙編 shellcode
- MalHost-Setup *file.doc* *out.exe* 0x4500 ➢ 在文件 0x4500 處提取 shellcode *並保存為 ut.exe*

Adobe PDF 文件格式說明：

一個 PDF 文件由頭，對象，交叉引用表(定位對象)和尾組成(trailer)。

1. 「/OpenAction」 和 「/AA」 (Additional Action) 指定能夠自動運行的腳本或動作。
2. 「/Names」, 「/AcroForm」, 「/Action」 也能夠指定和執行腳本或動作。
3. 「/JavaScript」 指定可運行的 JavaScript。

4. 「/GoTo*」 在當前文件中或其它 PDF 文件中更改指定的瀏覽位置。
5. 「/Launch」 啟動一個程序或打開一個文檔。
6. 「/URI」 通過網址訪問資源。
7. 「/SubmitForm」 和 「/GoToR」 給指定的 URL 發送數據。
8. 「/RichMedia」 在 PDF 文件中嵌入 Flash。
9. 「/ObjStm」 在對象流中隱藏對象。
10. 要注意用 16 進制混淆的代碼,例如「/JavaScript」對應於「/J#61vaScript」。

(See examples)

Adobe PDF 文件分析工具：

1. PDFiD - 鑑別 PDF 文件中是否包含與腳本和動作相關的字符串。(PDF Tools 的一部分)
2. PDF-parser - 鑑別 PDF 文件的關鍵元素。 (Part of PDF Tools)
3. Origami - Walker 查看 PDF 文件結構。
4. Origami pdfscan - 識別PDF 文件是否包含腳本和動作 identifies PDFs that contain strings associated with scripts and actions.
5. Origami extractjs 和 Jsunpack-n's pdf.py - 從 PDF 文件中提取 javascript。
6. Sumatra PDF 和 MuPDF - 輕量級的 PDF 文件查看工具。
7. Malzilla - 能夠在 PDF 文件中提取和解壓經 Zlib 壓縮的數據流，還能幫助分析被混淆的 JavaScript。
8. Jsunpack-n - 能夠提取和解碼 pcap 中的 JavaScript，還能解碼 PDF 文件。
9. CWSandbox, Wepawet, and Jsunpack - 能分析一些惡意 PDF 文件。

有用的 PDF 分析命令：

pdfid.py *file.pdf* 定位相關的腳本和動作字符串

pdf-parser.py *file.pdf*顯示文件結構辨別可以部分

pdfscan.rb *file.pdf* 查看和顯示文件結構

extractjs.rb *file.pdf* 提取 PDF 中的 javascript 腳本

pdf.py *file.pdf* 提取 PDF 中的 javascript 腳本

其它的惡意文件分析工具：

1. ExeFilter 從 Office 和 PDF 文件過濾腳本。
2. ViCheck.ca 自動檢測惡意 Office 和 PDF 文件。
3. VirusTotal 多引擎病毒掃瞄工具能夠鑑別一些惡意文檔。

參考文獻：

1. Adobe Portable Document Format (PDF) Reference
2. Physical and Logical Structure of PDF Files
3. Methods for Understanding and Analyzing Targeted Attacks with Office Documents (video)
4. Analyzing MSOffice Malware with OfficeMalScanner (follow-up presentation)
5. PDF Security Analysis and Malware Threats
6. Malicious Origami in PDF (follow-up presentation)
7. OffVis 1.0 Beta: Office Visualization Tool article
8. Reverse-Engineering Malware cheat sheet