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](http://www.pulog.org/" \t "_blank)

* 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](http://www.reconstructer.org/code/OfficeMalScanner.zip) - 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](http://go.microsoft.com/fwlink/?LinkId=158791) - 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](http://b2xtranslator.sourceforge.net/) -  轉換DOC, PPT,和XLS文件為Open XML文件(包括[BiffView](http://b2xtranslator.sourceforge.net/snapshots/BiffView.zip)工具)。
6. [OfficeCat](http://www.snort.org/vrt/vrt-resources/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](http://www.heaventools.com/) (不免費) and [FileInsight](http://vil.nai.com/vil/averttools.aspx) 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](http://blog.didierstevens.com/2008/04/29/pdf-let-me-count-the-ways/))  
**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](http://blog.didierstevens.com/programs/pdf-tools/)的一部分)
2. PDF-parser - 鑑別PDF文件的關鍵元素。 (Part of [PDF Tools](http://blog.didierstevens.com/programs/pdf-tools/))
3. [Origami](http://security-labs.org/origami/) - Walker 查看PDF文件結構。
4. [Origami](http://security-labs.org/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](http://security-labs.org/origami/) extractjs 和 [Jsunpack-n's](http://jsunpack.blogspot.com/2009/06/jsunpack-n-updates-for-pdf-decoding.html) pdf.py - 從PDF文件中提取javascript。
6. [Sumatra PDF](http://blog.kowalczyk.info/software/sumatrapdf/) 和 [MuPDF](http://ccxvii.net/mupdf/) - 輕量級的PDF文件查看工具。
7. [Malzilla](http://www.malzilla.org/) - 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](http://jsunpack.blogspot.com/2009/06/jsunpack-n-updates-for-pdf-decoding.html) - 能夠提取和解碼pcap中的JavaScript，還能解碼PDF文件。
9. [CWSandbox](https://cwsandbox.org/?site=1&page=submit), [Wepawet](http://wepawet.cs.ucsb.edu/), and [Jsunpack](http://jsunpack.jeek.org/dec/go) - 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](http://www.decalage.info/exefilter) 從Office和PDF文件過濾腳本。  
    2. [ViCheck.ca](https://www.vicheck.ca/) 自動檢測惡意Office和PDF文件。  
    3. [VirusTotal](http://www.virustotal.com/) 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](http://www.adobe.com/devnet/pdf/pdf_reference.html)  
    2.[Physical and Logical Structure of PDF Files](http://blog.didierstevens.com/2008/04/09/quickpost-about-the-physical-and-logical-structure-of-pdf-files/)  
    3. [Methods for Understanding and Analyzing Targeted Attacks with Office Documents](http://recon.cx/2008/a/bruce_dang/recon08_final.zip) ([video](http://bork.informatik.uni-erlangen.de/pub/ccc/25c3/video_h264_720x576/25c3-2938-en-methods_for_understanding_targeted_attacks_with_office_documents.mp4))  
    4. [Analyzing MSOffice Malware with OfficeMalScanner](http://www.reconstructer.org/papers/Analyzing%20MSOffice%20malware%20with%20OfficeMalScanner.zip) ([follow-up presentation](http://2009.hack.lu/archive/2009/New%20advances%20in%20Ms%20Office%20malware%20analysis.pdf))  
    5. [PDF Security Analysis and Malware Threats](http://www.blackhat.com/presentations/bh-europe-08/Filiol/Presentation/bh-eu-08-filiol.pdf)  
    6. [Malicious Origami in PDF](http://security-labs.org/fred/docs/pacsec08/pacsec08-fr-gd-full.pdf) ([follow-up presentation](http://security-labs.org/fred/docs/hack.lu09-origamis-strike-back.pdf))  
    7. [OffVis 1.0 Beta: Office Visualization Tool article](http://holisticinfosec.org/toolsmith/docs/september2009.html)  
    8. [Reverse-Engineering Malware cheat sheet](http://www.pulog.org/Resources/1299/MD-Cheat-Sheet/reverse-malware-cheat-sheet.html)