MALICIOUS DOCUMENT DETECTION & ANALYSIS

Tim Hsu

CHROOT.ORG

徐千洋(timhsu)

- HIT(Hacks In Taiwan)主辦人
- 網駭科技創辦人
- 資訊安全技術研究者，同時也是 Linux 愛好者。
- 專長於網路程式設計、網路滲透測試、駭客攻擊手法研究、惡意程式分析及嵌入式系統等。
- 著作
  - Linux C 函式庫詳解辭典 (旗標)
  - The Wargame 駭客訓練基地 - 決戰台灣版(旗標) (與曾信田、莊明躍、何弈甫等人合著)

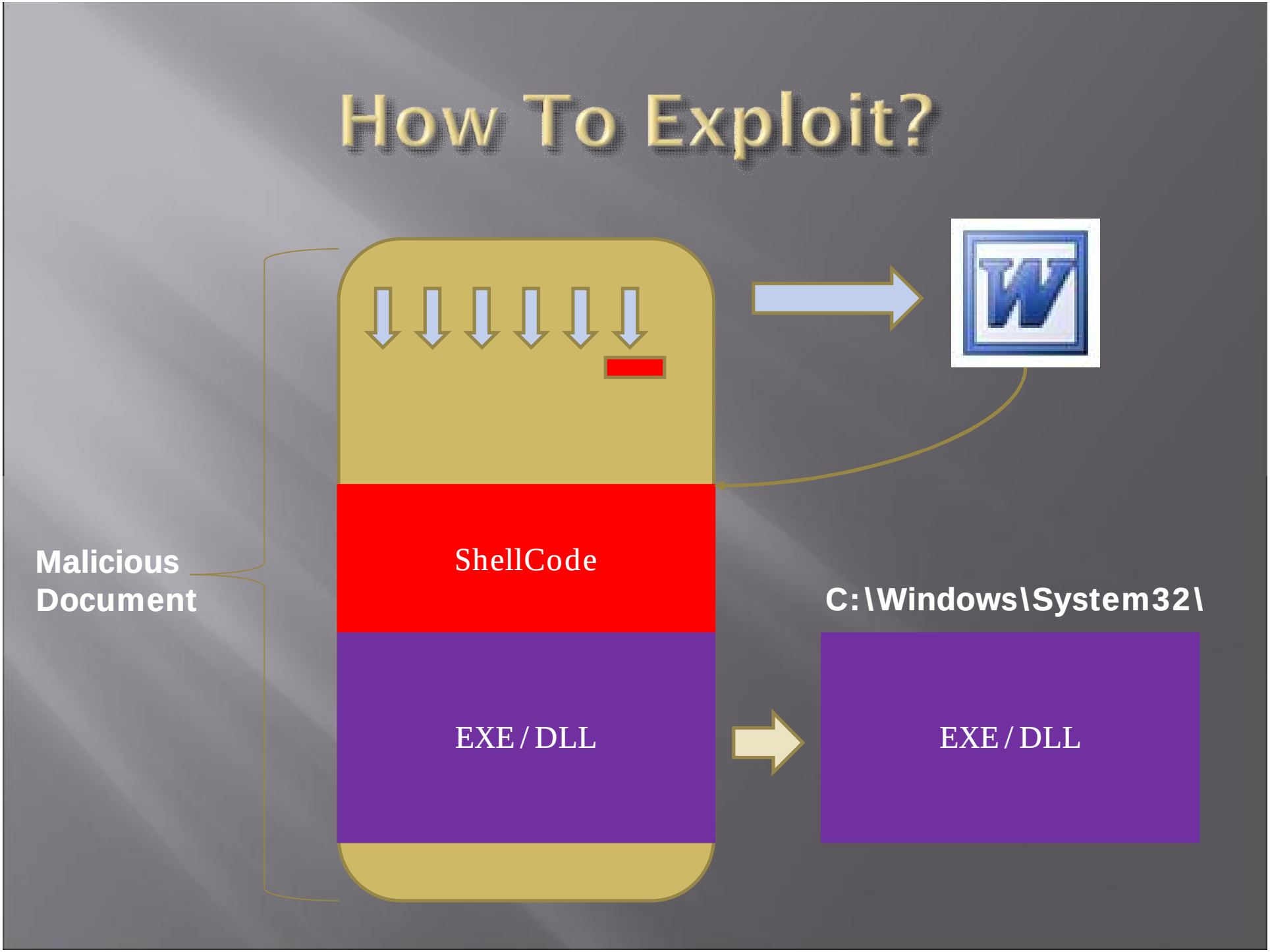
Outline

- Malicious Document Introduction
- How To Identify The Malicious Code In The Document
- Static And Dynamic Analysis
- To Identify Which Vulnerable
- Demo

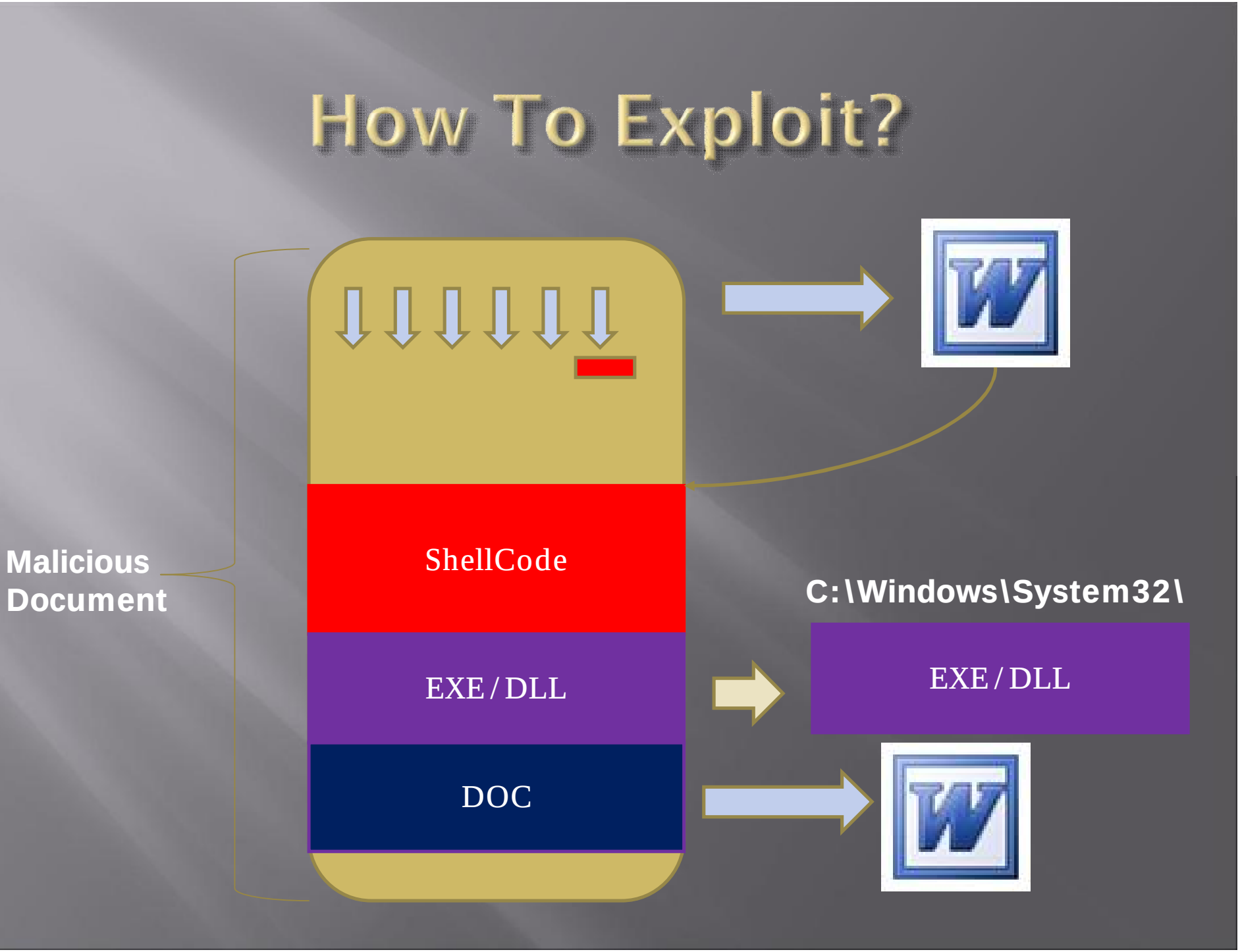
Malicious Document

- What Is Malicious Document?
- Office Document / CHM / PDF
- Other
  - RAR / ZIP / MDB / …

How To Exploit?



How To Exploit?



Windows Shellcode

- Kernel32.DLL
- Different Windows Version
- Find kernel32.dll
  - PEB(Process Environment Block)
  - SEH(Structured Exception Handling)
  - TOPSTACK

Process Environment Block

- PEB (Process Environment Block)
- Every running process that can always be found at fs:[0x30] from within the process.
- Works on: 95 / 98 / ME / NT / 2K / XP / 2K3
- mov eax, fs:[eax+0x30]

SEH

- SEH(Structured Exception Handling)
- Starts at local process address fs:0
- Works on: 95 / 98 / ME / NT / 2K / XP / 2K3
- xor ecx, ecx
- mov esi, fs:[ecx]

TOPSTACK

- Uses Thread Environment Block (TEB) located at fs:0x18h as starting point
- works on: NT / 2K / XP / 2K3
- xor esi, esi
- mov esi, fs:[esi + 0x18]

WHY - Shellcode Encoded

- Anti-Virus Software
- Signature
  - Typical opcode
  - PE/MZ Format
    - MZ

Shellcode Encoded

Shellcode Decoder

- Decode
  - XOR
- Get EIP
  - CALL/POP

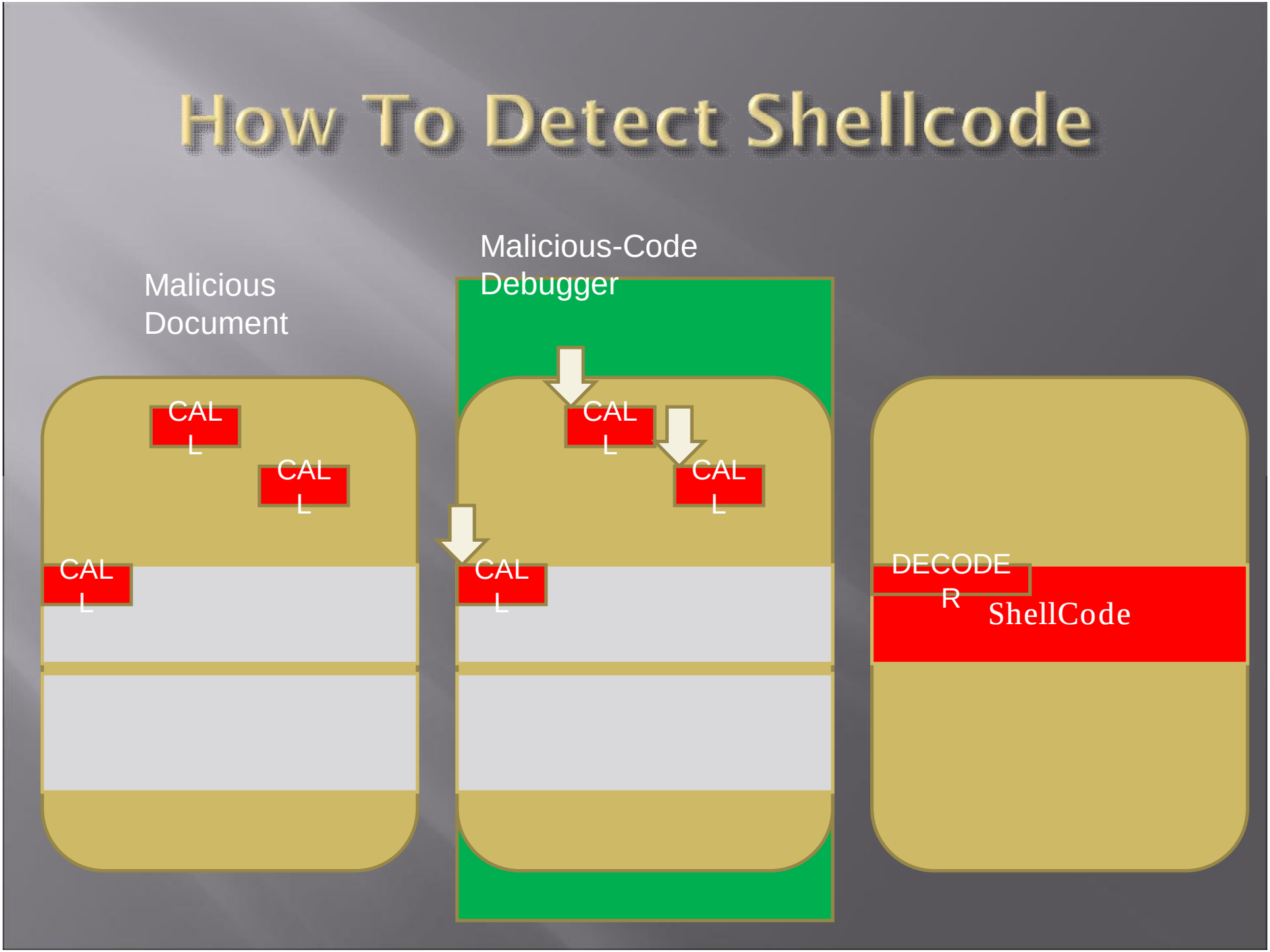
Detect Shellcode Decoder

- Dynamic Analysis
  - Debugger
    - Trace Step
    - Environment
  - Emulator
    - Instruction
    - Environment

MDScan - Tracer

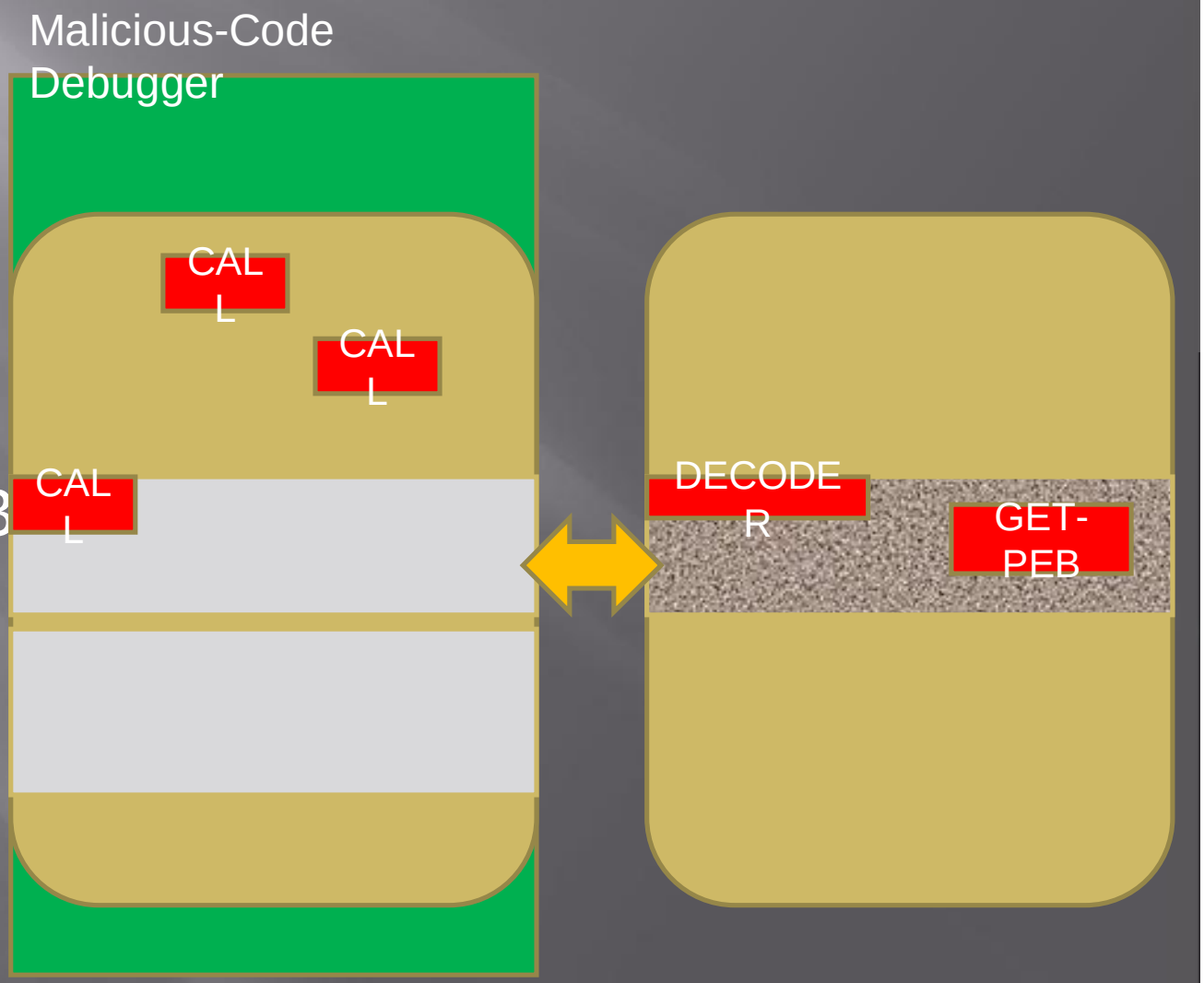
- Search [CALL / JNE / LOOP*] opcode
  - Named "DECODER"
- Initialization
  - Registers
  - Copy it to stack
- Jump to "DECODER"
- Debugger:
  - Call&Pop found?
  - GetPEB found?
  - BadOpcode?
    - Interrupt / sysenter / halt / …
  - Loop ?
  - Save EIP
  - Next Step

How To Detect Shellcode

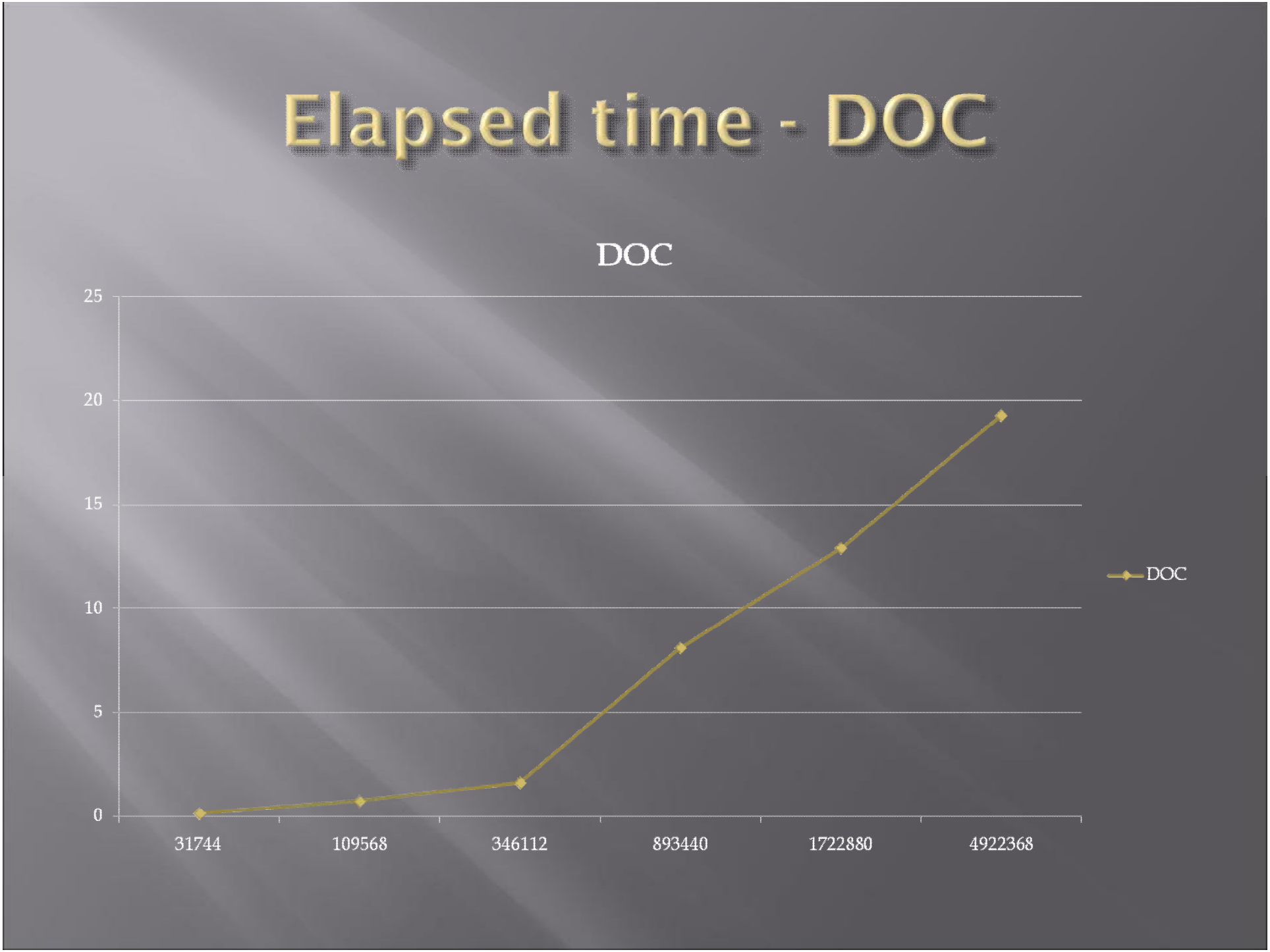


Memory Diff

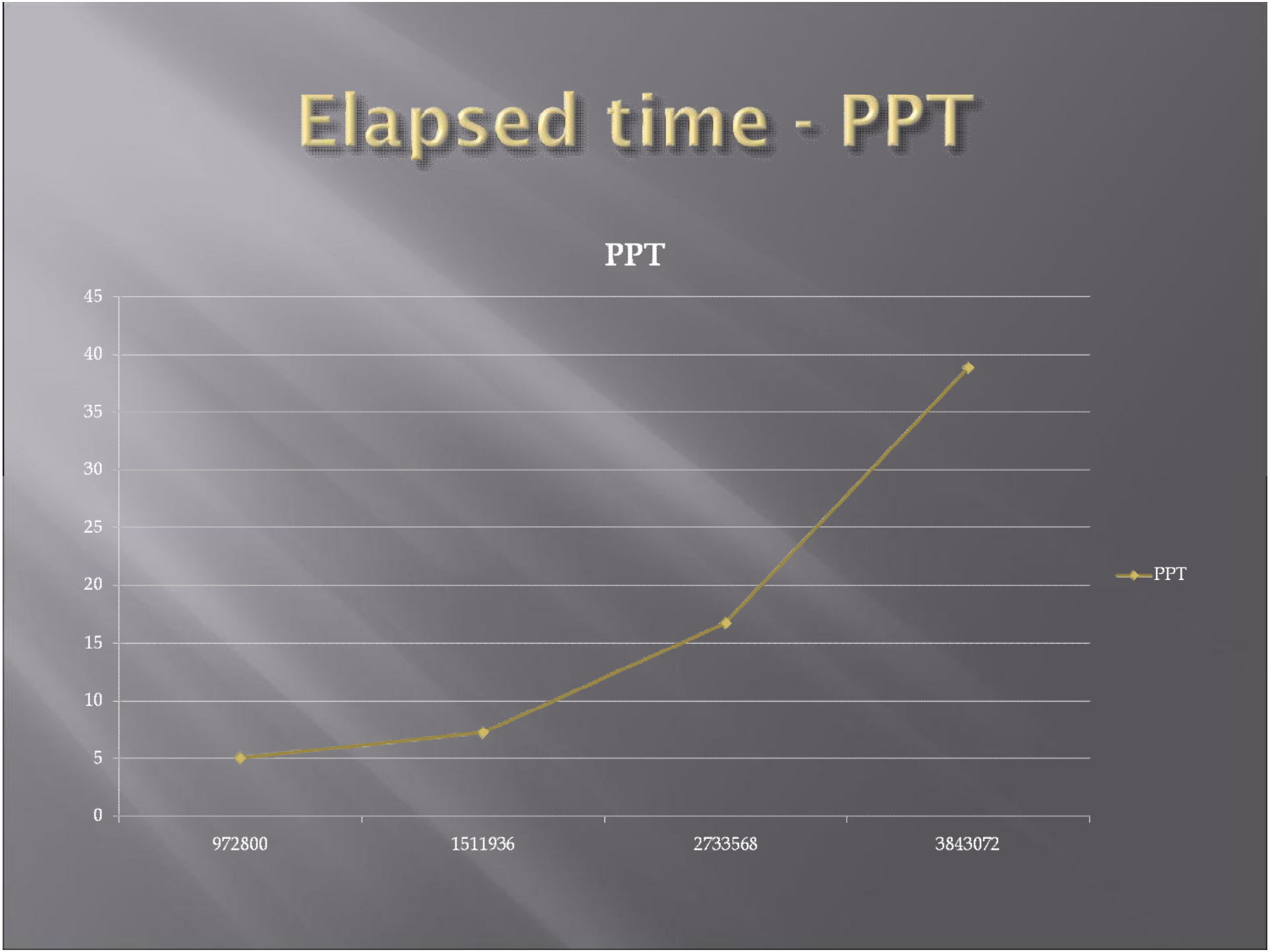
- How many bytes be modified?
- Scan Get-PEB opcode



Elapsed time - DOC



Elapsed time - PPT



MDScan Demo

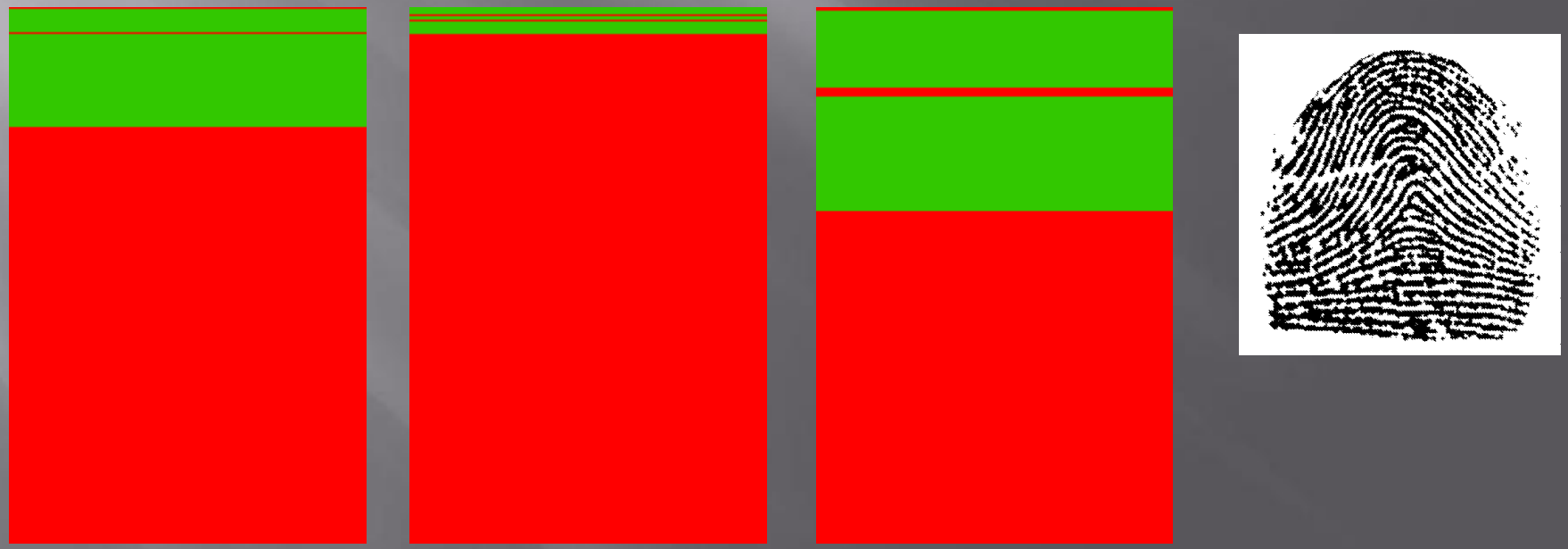
```
Malicious Document Scan Tool    Version 0.4
------------------------------------------------
Copyright (c) 2008 CHROOT.ORG. All rights reserved.

Scanning sample/4e87a852b2afe5aa8fe5cbc6219ca794.doc
* Total overwrite 262 bytes!
* Found Get-PEB shellcode after decode memory.
Found: Short CALL and POP:
At file offset: 0x2de9
084A0809  E8 E4 FF FF FF          lcall   0xFFFFFFE9
084A07F2  8B 34 24                movl    (%esp), %esi
084A07F5  33 DB                   xor     %ebx, %ebx
084A07F7  33 C0                   xor     %eax, %eax
084A07F9  56                      push    %esi
084A07FA  5F                      pop     %edi
084A07FB  33 C9                   xor     %ecx, %ecx
084A07FD  66 B9 82 02             mov     $0x0282, %cx
084A0801  AC                      lodsbb  %ds:(%esi), %al
084A0802  34 EF                   xor     $0xEF, %al
084A0804  AA                      stosbb  %al, %es:(%edi)
084A0805  E2 FA                   lloop   0xFFFFFFFC
```

OLE Storage Access

- Libgsf
  - 'libgsf' is a simple i/o library that can read and write common file types and handle structured formats that provide file-system-in-a-file semantics. There are some additional utilities for document centric applications
- Source Archive: http://ftp.acc.umu.se/pub/GNOME/sources/libgsf/1.14/libgsf-1.14.4.tar.gz
- Licenses: LGPLv2.1

OLE Storage Fingerprint

- OLE Storage offset and size
- Draw the diagram with MDScan
- Green: OLE Red: Data or Unknow



OfficeCat

- http://www.snort.org/vrt/tools/officecat.html

C:\>officecat.exe ATest.doc

Sourcefire OFFICE CAT v2

Microsoft Office File Checker *

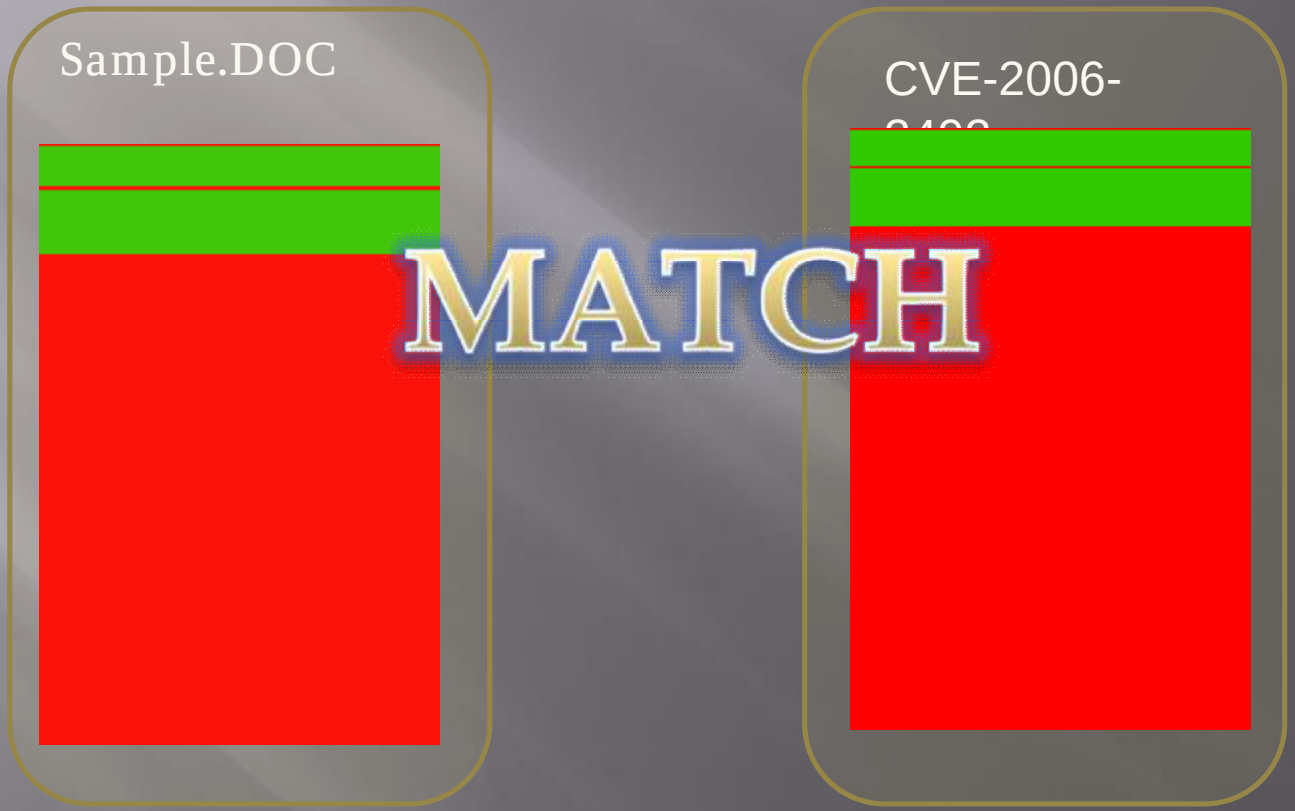
Processing ATest.doc VULNERABLE

OCID: 5

CVE-2006-6456

Type: Word

Identify CVE



MDScan – OLE Scan Demo

CHM/RAR/PDF

- CHM(compiled html help)
  - Extract it
  - Any PE/MZ Execuated file?
- RAR
  - WinRAR "lzh.fmt" LHA Archive Processing Client-Side Buffer Overflow Vulnerability
  - CVE-2006-3845
  - .RAR but LHA magic?
- PDF
  - Adobe Products JavaScript Method Code Execution Vulnerability
  - Embed Javacript?

Extract CHM

```
-rw-r--r-- 1 timhsu timhsu    262 2008-06-22 23:49 chmin-boop.html
-rw-r--r-- 1 timhsu timhsu 103936 2008-06-22 23:49 exe.exe
drwxr-xr-x 2 timhsu timhsu   4096 2008-06-22 23:49 $FIftiMain/
-rw-r--r-- 1 timhsu timhsu   4096 2008-06-22 23:49 #IDXHDR
-rw-r--r-- 1 timhsu timhsu    521 2008-06-22 23:49 index.html
drwxr-xr-x 2 timhsu timhsu   4096 2008-06-22 23:49 #ITBITS/
-rw-r--r-- 1 timhsu timhsu   2751 2008-06-22 23:49 $OBJINST
-rw-r--r-- 1 timhsu timhsu      5 2008-06-22 23:49 #STRINGS
-rw-r--r-- 1 timhsu timhsu   4217 2008-06-22 23:49 #SYSTEM
-rw-r--r-- 1 timhsu timhsu     32 2008-06-22 23:49 #TOPICS
-rw-r--r-- 1 timhsu timhsu     52 2008-06-22 23:49 #URLSTR
-rw-r--r-- 1 timhsu timhsu     24 2008-06-22 23:49 #URLTBL
drwxr-xr-x 2 timhsu timhsu   4096 2008-06-22 23:49 $WWAssociativeLinks/
drwxr-xr-x 2 timhsu timhsu   4096 2008-06-22 23:49 $WWKeywordLinks/
```

PDFCheck.sh

```
#!/bin/sh
if [ "$1" == "" ]; then
        echo Usage: $0 [pdf]
        exit 0;
fi
file $1 | grep "PDF document" > /dev/null 2>&1
if [ "$?" == "1" ]; then
echo Sorry! Not PDF file.
exit 0;
fi
pdfsize=`pdfinfo $1 | grep "File size" | awk '{print $3}'`
pdftotext -raw $1 /tmp/pdf_test.txt
pdftxtsize=`stat -c %s /tmp/pdf_test.txt`
if [ $pdftxtsize == 1 ];then
        echo Warnning: $1 maybe not be safe!
        exit 1;
else
        echo $1 is safe.
        exit 0;
fi
```

Reference

- Exploit Modify Tips & 0day – Nanika
  - HIT 2006 (http://www.hitcon.org/oldweb/sch.htm)
- Understanding Windows Shellcode
  - http://www.hick.org/code/skape/papers/win32-shellcode.pdf
- Windows Memory Layout, User-Kernel Address Spaces
  - http://www.openrce.org/reference_library/files/reference/Windows%20Memory%20Layout,%20User-Kernel%20Address%20Spaces.pdf
- Dynamic analysis of malicious code
  - http://www.cs.ucsb.edu/~chris/research/doc/virology06_dynamic.pdf
- OfficeCat
  - http://www.snort.org/vrt/tools/officecat.html

Question?