

Exploiting Software Vulnerabilities

Program Binary Analysis

© All wrongs reversed – under CC-BY-NC-SA 4.0 license



1542

Universidad
Zaragoza

Dept. of Computer Science and Systems Engineering
University of Zaragoza, Spain

Course 2022/2023

Master's Degree in Informatics Engineering

UNIVERSITY OF ZARAGOZA

Seminar A.22, Ada Byron building



Outline

- 1** Introduction to Program Binary Analysis
- 2** Static Analysis Techniques
- 3** Dynamic Analysis Techniques

Outline

- 1** Introduction to Program Binary Analysis
- 2 Static Analysis Techniques
- 3 Dynamic Analysis Techniques

Introduction

```
#include <stdio.h>
```

```
int main(int argc, char *argv[])  
{  
    printf("hello world!\n");  
    return 0;  
}
```

```
push    ebp  
mov     ebp, esp  
and     esp, -16  
sub     esp, 16  
call   ___main  
mov     DWORD PTR [esp], OFFSET FLAT:LC0  
call   _puts  
mov     eax, 0  
leave    
ret
```

■ Programs are written in text

- Both source code and assembly!
- **Character sequences (bytes)**
- Difficult to work with (for humans, not for machines)
- We need some **structured representation**

Introduction

Program Analysis

Automatically reason and derive properties about the behavior of computer programs

Approaches

■ Static Program Analysis

- Without running the program
- The abstract model of the program is obtained and (symbolically) executed
- Analysis performed through the abstract model
- **Examples:** CFA, DFA, symbolic execution, . . .

■ Dynamic Program Analysis

- Running the program on some chosen inputs
- Traces are collected and then analyzed
- Analysis performed through these concrete executions
- **Examples:** software testing, taint analysis, concolic execution. . .

Introduction

Input program formats for analysis

- **Abstract model:** all unnecessary information for analysis have been removed. Only the necessary information remains
- **Source code:** Keep track of high-level, human-readable information about the program (variables, types, functions, etc.)
- **Bytecode:** may vary depending on the bytecode considered, but keep a record of little high-level information about the program, such as types and functions. The programs are unstructured
- **Binary file:** just keep track of statements in an unstructured way (no for-loop, no clear argument passing in procedures, etc). No type, no names. The binary file can include meta-data that can be useful for analysis (symbols, debug, etc.)
- **Memory dump:** Pure assembler instructions with a full memory state of the current execution. We no longer have the meta-data of the executable file

Binary code is the closest format of what will be executed!

Introduction

Binary code vs. source code

What you code is not what you execute!

We want to analyze binary code. It can come as:

- an executable file,
- an object file,
- a dynamic library,
- a firmware,
- a memory dump,
- ...

We do not trust to obtain the corresponding high-level source code

Introduction

Motivations

Why should we analyze binary programs?

- Lack of high-level source code
- Low-level assembly code embedded in source code
- Legacy code
- Commercial Off-The-Shelf software (COTS)
- App stores (for mobile phones and tablets)
- Malware (or other “hostile” programs)
- Technology forecast
- Mistrust in the compilation chain
- C compiler possibly buggy
- Checking for low-level bugs (e.g., exploiting a stack buffer overflow)
- Errors with strong hardware interconnection

Introduction

Understanding papers on Program Analysis

For those who keep track of such things, checkers in the research system typically traverse program paths (flow-sensitive) in a forward direction, going across function calls (inter-procedural) while keeping track of call-site-specific information (context-sensitive) and toward the end of the effort had some of the support needed to detect when a path was infeasible (path-sensitive).

Note these terms

■ Flow-(in)sensitive

■ Context-(in)sensitive

■ Inter-(intro)procedural

■ Path-(in)sensitive

Further reading: *A few billion lines of code later: using static analysis to find bugs in the real world.* Al Bessey, Ken Block, Ben Chelf, Andy Chou, Bryan Fulton, Seth Hallem, Charles Henri-Gros, Asya Kamsky, Scott McPeak, Dawson Engler. *Communications of the ACM*, vol. 53, iss. 2, pp. 66-75 (February 2010). doi: 10.1145/1646353.1646374

Outline

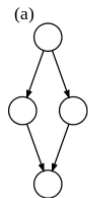
1 Introduction to Program Binary Analysis

2 Static Analysis Techniques

3 Dynamic Analysis Techniques

Static Analysis Techniques

Control-Flow Graphs

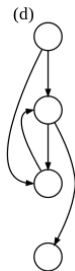
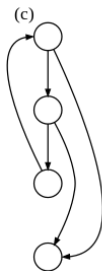


- **Control flow within a function**

- **Nodes: basic blocks**

- Sequence of consecutive program instructions that have an entry point (first instruction executed) and an exit point (last instruction executed)
- Entry and exit blocks

- **Edge:** control flows from A to B



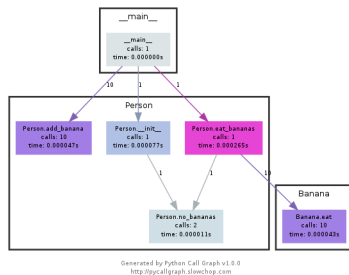
Applications

- Compiler optimizations
- Data-flow analysis (taint analysis)
- Behavioral-based monitors

Credits: https://en.wikipedia.org/wiki/Control_flow_graph

Static Analysis Techniques

Call Graphs



- **Interprocedural CFG. Information flow between functions**
- **Nodes:** functions
- **Edge:** A could call B
- Types: static, dynamic (record of a program execution)
- Application: find procedures never called
- **Available tools for automatic generation of call-graphs**

Credits: https://en.wikipedia.org/wiki/Call_graph

Static Analysis Techniques

Disassembling

0040166B	..0F85 24010000	JNZ xconv.00401795	
00401671	. 6A 0E	PUSH 0E	
00401673	. 68 2D544000	PUSH xconv.0040542D	Count = E (14.)
00401678	. 68 80000000	PUSH 80	Buffer = xconv.0040542D
0040167D	. FF75 08	PUSH DWORD PTR SS:[EBP+8]	ControlID = 80 (128.)
00401680	. E8 C1040000	CALL <JMP.&user32.GetDlgItemTextA>	hWnd
00401685	. 83F8 0C	CMP EAX,0C	GetDlgItemTextA
00401688	..7F 4B	JG SHORT xconv.004016D5	
0040168A	. 33F3 94	CMPSX 4	
0040168D	..7C 46	JL SHORT xconv.004016D5	
0040168F	. 68 2D544000	PUSH xconv.0040542D	String2 = "DeAtH"
00401694	. 68 06334000	PUSH xconv.00403306	String1 = xconv.00403306
00401699	. E8 26050000	CALL <JMP.&kernel32.lstrcpyA>	lstrcpyA
0040169E	. 6A 1B	PUSH 1B	Count = 1B (27.)
004016A0	. 68 62544000	PUSH xconv.00405462	Buffer = xconv.00405462
004016A5	. 68 81000000	PUSH 81	ControlID = 81 (129.)
004016AA	. FF75 08	PUSH DWORD PTR SS:[EBP+8]	hWnd
004016AD	. E8 94040000	CALL <JMP.&user32.GetDlgItemTextA>	GetDlgItemTextA
004016B2	. E8 48020000	CALL xconv.00401501	
004016B7	. 83F8 01	CMP EAX,1	
004016BA	..74 32	JE SHORT xconv.004016EE	
004016BC	. 8005 B8534000	ADD BYTE PTR DS:[4053B8],1	
004016C3	. 805D B8534000	CMPSI PTR DS:[4053B8],3	
004016C9	..0F84 9C000000	JE xconv.00401795	
004016D0	..E9 81000000	JMP xconv.00401755	
004016D5	> 6A 10	PUSH 10	Style = MB_OK MB_ICONHAND MB_APPLMODAL
004016D7	. 68 79304000	PUSH xconv.00403079	Title = "Sorry"
004016DC	. 68 53324000	PUSH xconv.00403253	Text = "Sorry username must be at least 4 characters long and not n
004016E1	. FF75 08	PUSH DWORD PTR SS:[EBP+8]	hOwner
004016E4	. E8 69040000	CALL <JMP.&user32.MessageBoxA>	MessageBoxA
004016E9	..E9 D7000000	JMP xconv.004017C5	
004016EE	> E8 A0030000	CALL xconv.00401A98	
004016F3	. C605 72434000	MOVB BYTE PTR DS:[404372],1	
004016F7	. 6A 40	PUSH 40	Style = MB_OK MB_ICONASTERISK MB_APPLMODAL
004016FC	. 68 33304000	PUSH xconv.00403033	Title = "Thank you!"
00401701	. 68 3E304000	PUSH xconv.0040303E	Text = "Registration done. Thank you for registering this program!"
00401706	. FF75 08	PUSH DWORD PTR SS:[EBP+8]	hOwner
00401709	. E8 44040000	CALL <JMP.&user32.MessageBoxA>	MessageBoxA
0040170E	. 6A 00	PUSH 0	Result = 0

- Generally speaking, read PUSH EAX instead of 0x50
- **Lots of tools** see https://en.wikibooks.org/wiki/X86_Disassembly/Disassemblers_and_Decompliers
 - Win32Dasm
 - OllyDBG (also debugger)
 - IDA Pro (also debugger)
 - r2 (also debugger)

Static Analysis Techniques

Disassembling

Main challenges

- **Variable-length instruction sets:** overlapping instructions
- **Mixed data and code:** misclassifying data as instructions
- **Indirect jumps:** Any location can be the start of an instruction!
- **Start of functions:** when the calls are indirect
- **End of functions:** when there is no dedicated return instruction exists
 - Handwritten assembly code may not conform to standard call conventions
- **Code compression:** the code of two functions overlaps
- **Self-modifying code**

Static Analysis Techniques

Decompilation – example

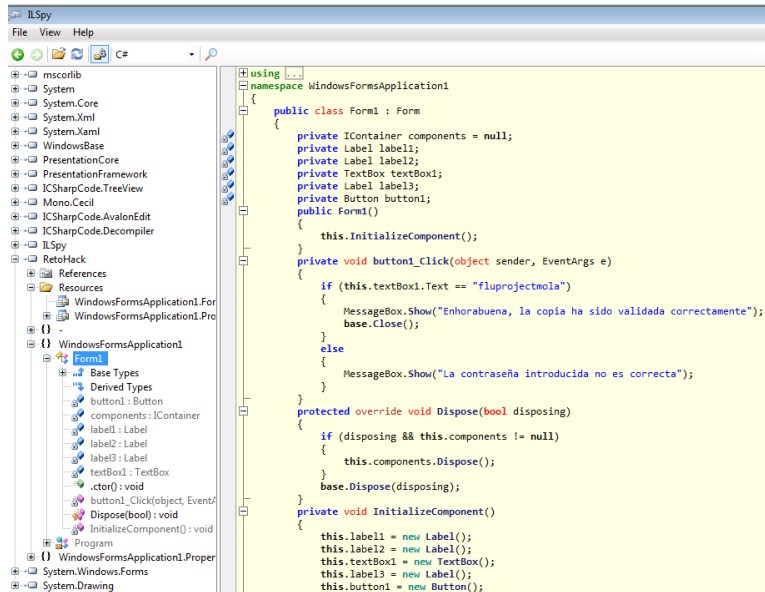
```
int __stdcall sub_40162C(HWND hDlg, int a2, int a3, int a4){
    HICON v4; // eax@2
    UINT v5; // eax@5

    switch ( a2 ) {
    case 272:
        v4 = LoadIconA(hInstance, (LPCSTR)0x64);
        SendMessageA(hDlg, 0x80u, 1u, (LPARAM)v4);
        break;
    case 273:
        if ( a3 == 126 ) {
            v5 = GetDlgItemTextA(hDlg, 128, dword_40542D, 14);
            if ( (signed int)v5 > 12 || (signed int)v5 < 4 ) {
                MessageBoxA(hDlg, "Sorry username must be at least 4
characters\r\nlong
and not more than 12 characters.", "Sorry", 0x10u);
            } else {
                lstrcpyA(dword_403306, dword_40542D);
                GetDlgItemTextA(hDlg, 129, byte_405462, 27);
                if ( sub_401901() == 1 ) {
                    sub_401A9B();
                    byte_404372 = 1;
                    MessageBoxA(hDlg, "Registration done. Thank you for registering
this
program!", "Thank you!", 0x40u);
                    EndDialog(hDlg, 0);
                    EnableWindow(dword_403363, 0);
                    SetWindowTextA(
                        dword_4054A7,
                            "X-Convertor v1.0 2005 by TDC and BoR0\r\n\r\n
Coded by\t: TDC and BoR0\r\n\r\nVersion\t\t: 1.0\r\n\r\nRelease
date\t: 18-08-2005\r\n\r\n\r\nX-Convertor converts up to 4KB
each convert.\r\n\r\n\r\nRegistered version. Thank you.\r\n\r\n");
                    lstrcatA(byte_403330, dword_403306);
                    SetWindowTextA(dword_4054AB, byte_403330);
                }
            }
        }
    }
}

else {
    ++byte_4053B8;
    if ( byte_4053B8 == 3 ) {
        MessageBoxA(hDlg, "Your serial is not correct",
            "Sorry", 0x10u);
        byte_4053B8 = 0;
        EndDialog(hDlg, 0);
    } else {
        MessageBoxA(hDlg, "Your serial is not correct",
            "Sorry", 0x10u);
    }
}
}
} else {
    if ( a3 == 127 ) {
        byte_4053B8 = 0;
        EndDialog(hDlg, 0);
    }
}
break;
case 16:
    byte_4053B8 = 0;
    EndDialog(hDlg, 0);
    break;
}
return 0;
}
```

Static Analysis Techniques

Decompilation



The screenshot shows the ILSpy decompiler interface. On the left, a tree view displays the project structure, including references to various .NET assemblies like System, System.Core, and System.Windows.Forms. The 'Form1' class is selected, and its decompiled C# code is shown in the main pane. The code defines a class 'Form1' that inherits from 'Form'. It contains several private fields for UI controls (Label, TextBox, Button) and an 'IContainer' for components. The 'Form1()' constructor calls 'InitializeComponent()'. A 'button1_Click' event handler shows a message box when the text box contains 'fluprojectmola' and another message box when the text is empty. The 'Dispose' method is overridden to dispose of the components and the base class. The 'InitializeComponent' method initializes the UI controls with 'new' statements.

```
using ...
namespace WindowsFormsApplication1
{
    public class Form1 : Form
    {
        private IContainer components = null;
        private Label label1;
        private Label label2;
        private TextBox textBox1;
        private Label label3;
        private Button button1;
        public Form1()
        {
            this.InitializeComponent();
        }
        private void button1_Click(object sender, EventArgs e)
        {
            if (this.textBox1.Text == "fluprojectmola")
            {
                MessageBox.Show("Enhorabuena, la copia ha sido validada correctamente");
                base.Close();
            }
            else
            {
                MessageBox.Show("La contraseña introducida no es correcta");
            }
        }
        protected override void Dispose(bool disposing)
        {
            if (disposing && this.components != null)
            {
                this.components.Dispose();
            }
            base.Dispose(disposing);
        }
        private void InitializeComponent()
        {
            this.label1 = new Label();
            this.label2 = new Label();
            this.textBox1 = new TextBox();
            this.label3 = new Label();
            this.button1 = new Button();
        }
    }
}
```


Static Analysis Techniques

Decompilation

Main challenges

- **Disassembly:** **first step of any decompiler!**
- **Target language:** assembly code may not correspond to any (correct) source code
- **Library functions**
- **Compiler-dependent instruction equivalences**
 - `int a = 0` \rightarrow `mov eax, [a]; xor eax, eax`
- **Target architecture artifacts:** unnecessary jumps-to-jumps
- **Structured control flow**
- **Compiler optimizations:** loop unrolling, shifts, adds, ...
- **Loads/stores:** operations on arrays, records, pointers, and objects
- **Self-modifying code:** normally, the segment code will be unchanged, although there are programs that modify themselves!

Static Analysis Techniques

Data Flow Analysis

- Analyze the **effect of each instruction**
- Compose instruction effects to derive information at the basic block boundaries
- Framework for **providing facts about programs**. Based on all paths through program (including also infeasible paths)
- **Derive information about the dynamic behavior of a program by examining the code statically**

Useful for...

- **Program debugging**: what definitions (of variables) can reach a program point?
- **Program optimizations**: constant folding, copy propagation, elimination of common sub-expressions, etc.

Static Analysis Techniques

Data Flow Analysis

Consider the statement $a = b + c$

Statement effects

- **Uses variables** (b, c)
 - **“Kills” a previous definition** (old value of a)
 - **New definition** (a)
-
- **Compose effect of statements** → **effect of a basic block**
 - *Locally exposed usage*: usage of a data item that is not preceded in the basic block by a data item definition
 - Any definition of a data item *kills* all definitions of the same data item that reach the basic block
 - *Locally available definition*: last definition of the data item in the basic block

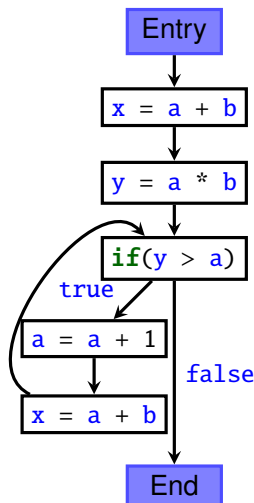
Static Analysis Techniques

Data Flow Analysis

■ Facts

- $a + b$ is available
- $a * b$ is available
- $a + 1$ is available

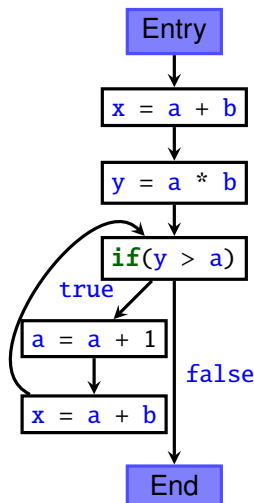
- Let's calculate the facts that hold for each program point!



Static Analysis Techniques

Data Flow Analysis

Statement	Gen	Kill
$x = a + b$	$a + b$	
$y = a * b$	$a * b$	
$y > a$		
$a = a + 1$		$a + b$ $a * b$ $a + 1$



Static Analysis Techniques

Data Flow Analysis

- *Forward vs. backward*: data flow from in to out (vs. from out to in)
- *Must vs. may*: at joint points, just keep facts that hold on all paths (vs. any path) that are joined

	Must	May
Forward	Available expressions	Reaching definitions
Backward	Very busy expressions	Live variables

Limitations

- **Data-Flow Analysis is good for analyzing local variables**
 - What happens to values stored in the heap?
 - Not modeled on traditional data flow
- In general, **it is difficult to analyze pointers**. Suppose $*x = p$
 - Assume all data flow facts are killed
 - Or assume writing via x can affect any variable whose address has been taken

Static Analysis Techniques

Symbolic Execution

- **Allows us to scale and model all possible executions of a program**
- Concrete vs. symbolic execution
 - **Tests work, but each test only explores one possible execution path**
- **Symbolic execution generalizes testing**
 - Allows unknown *symbolic* variables in evaluation
 - **Checks the feasibility of the program paths**

Challenges

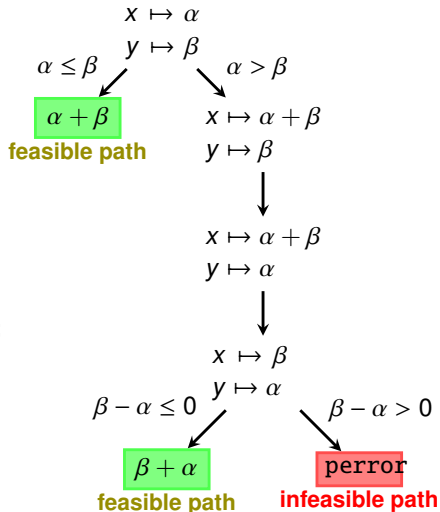
- **Path explosion**
- **Modeling statements and environments**
- **Constraint resolution**

Further reading: Roberto Baldoni, Emilio Coppa, Daniele Cono D'elia, Camil Demetrescu, and Irene Finocchi. *A Survey of Symbolic Execution Techniques*. ACM Comput. Surv. 51, 3, Article 50 (July 2018), 39 pages. doi: 10.1145/3182657

Static Analysis Techniques

Symbolic Execution

```
1  int f(int x, int y)
2  {
3      if(x > y)
4      {
5          x = x + y;
6          y = x - y;
7          x = x - y;
8          if(x - y > 0)
9              perror("Error!");
10     }
11
12     return x + y;
13 }
```



How to decide which branches are feasible?

Combine the path condition with the branch condition and ask an SMT solver!

Static Analysis Techniques

Symbolic Execution – example: bug finding

Catch the error! What value (or values) triggers it?

```
1  int bar(int i)
2  {
3      int j = 2*i;
4      i++;
5      i = i*j;
6      if (i < 1)
7          i = -i;
8
9      i = j/i;
10     return i;
11 }
```

False branch condition

$$i = (i_{in} + 1)2i_{in}$$

$$(i_{in} + 1)2i_{in} \geq 1$$

True branch condition

$$i = -(i_{in} + 1)2i_{in}$$

$$(i_{in} + 1)2i_{in} < 1$$

Static Analysis Techniques

Symbolic Execution – example: bug finding

Catch the error! What value (or values) triggers it?

```
1  int bar(int i)
2  {
3      int j = 2*i;
4      i++;
5      i = i*j;
6      if (i < 1)
7          i = -i;
8
9      i = j/i;
10     return i;
11 }
```

False branch condition

$$i = (i_{in} + 1)2i_{in}$$

$$(i_{in} + 1)2i_{in} \geq 1$$

True branch condition

$$i = -(i_{in} + 1)2i_{in}$$

$$(i_{in} + 1)2i_{in} < 1$$

Division by zero creates problems...

Static Analysis Techniques

Symbolic Execution – example: bug finding

Catch the error! What value (or values) triggers it?

```
1  int bar(int i)
2  {
3      int j = 2*i;
4      i++;
5      i = i*j;
6      if (i < 1)
7          i = -i;
8
9      i = j/i;
10     return i;
11 }
```

False branch condition

$$i = (i_{in} + 1)2i_{in}$$

$$(i_{in} + 1)2i_{in} \geq 1$$

True branch condition

$$i = -(i_{in} + 1)2i_{in}$$

$$(i_{in} + 1)2i_{in} < 1$$

Division by zero creates problems...

False branch is always safe

$$(i > 0, \forall i_{in} | (i_{in} + 1)2i_{in} \geq 1)$$

What about the true branch?

Static Analysis Techniques

Symbolic Execution – example: bug finding

Catch the error! What value (or values) triggers it?

```
1  int bar(int i)
2  {
3      int j = 2*i;
4      i++;
5      i = i*j;
6      if (i < 1)
7          i = -i;
8
9      i = j/i;
10     return i;
11 }
```

False branch condition

$$i = (i_{in} + 1)2i_{in}$$

$$(i_{in} + 1)2i_{in} \geq 1$$

True branch condition

$$i = -(i_{in} + 1)2i_{in}$$

$$(i_{in} + 1)2i_{in} < 1$$

Division by zero creates problems...

False branch is always safe

$$(i > 0, \forall i_{in} | (i_{in} + 1)2i_{in} \geq 1)$$

What about the true branch?

$$-(i_{in} + 1)2i_{in} = 0$$

Static Analysis Techniques

Symbolic Execution – example: bug finding

Catch the error! What value (or values) triggers it?

```
1  int bar(int i)
2  {
3      int j = 2*i;
4      i++;
5      i = i*j;
6      if (i < 1)
7          i = -i;
8
9      i = j/i;
10     return i;
11 }
```

False branch condition

$$i = (i_{in} + 1)2i_{in}$$

$$(i_{in} + 1)2i_{in} \geq 1$$

True branch condition

$$i = -(i_{in} + 1)2i_{in}$$

$$(i_{in} + 1)2i_{in} < 1$$

Division by zero creates problems...

False branch is always safe

$$(i > 0, \forall i_{in} | (i_{in} + 1)2i_{in} \geq 1)$$

What about the true branch?

$$-(i_{in} + 1)2i_{in} = 0 \rightarrow i_{in} = -1, i_{in} = 0$$

Outline

- 1 Introduction to Program Binary Analysis
- 2 Static Analysis Techniques
- 3 Dynamic Analysis Techniques**

Dynamic Analysis Techniques

Debugging

- **Execute program instructions with special software:** *debuggers*
 - We can see the values of each register of the CPU, the stack, the memory, etc.
- Source code vs. binary debugging
- **Breakpoints:** stops execution when reached
 - Software breakpoints (memory)
 - Hardware breakpoints
 - On execute, read, or write operations
- Step into / step onto

Dynamic Analysis Techniques

Debugging (example: OLLyDBG)

OLLyDbg - OLLYDBG.EXE [CPU - main thread, module OLLYDBG]

File View Debug Plugins Options Window Help

LEMTW H C / K B R ... S

00401000 \$->EB 10 JMP SHORT OLLYDBG.00401012

00401002	62	DB 62	CHAR 'f'
00401003	62	DB 62	CHAR 'b'
00401004	3A	DB 3A	CHAR 'r'
00401005	43	DB 43	CHAR 'c'
00401006	2B	DB 2B	CHAR '+'
00401007	2B	DB 2B	CHAR '+'
00401008	4B	DB 4B	CHAR 'h'
00401009	4F	DB 4F	CHAR 'o'
0040100A	4F	DB 4F	CHAR 'o'
0040100B	4B	DB 4B	CHAR 'k'
0040100C	90	NOP	
0040100D	59	POP E9	
0040100E	> 2B014B00	DD OFFSET OLLYDBG.____CPPdebugHook	
00401012	> A1 1B014B00	MOV EAX, DWORD PTR DS:[4B011B]	
00401017	> C1E0 02	SHL EAX, 2	
0040101A	> A3 1F014B00	MOV DWORD PTR DS:[4B011F], EAX	
0040101F	> 52	PUSH EDX	
00401020	> 6A 00	PUSH 0	
00401022	> E8 4BE0A000	CALL <JMP, &KERNEL32.GetModuleHandleA>	[Module = NULL GetModuleHandleA
00401027	> 8B00	MOV EDX, EAX	
00401029	> E8 DA240A00	CALL OLLYDBG.004A3508	
0040102E	> 5A	POP EDX	
0040102F	> E8 70100A00	CALL OLLYDBG.004A2804	
00401034	> E8 D3240A00	CALL OLLYDBG.004A350C	
00401039	> 6A 00	PUSH 0	
0040103B	> E8 F8390A00	CALL OLLYDBG.004A4438	[Arg1 = 00000000 OLLYDBG.004A4438
00401040	> 59	POP EAX	
00401041	> C8 4004B00	PUSH OLLYDBG.004B00C4	
00401044	> 00	PUSH 0	
00401048	> E8 25E00A00	CALL <JMP, &KERNEL32.GetModuleHandleA>	[Module = NULL GetModuleHandleA

00401012=OLLYDBG.00401012

Registers (FPU)

EAX	00000000
ECX	0012FFB0
EDX	7C91E4F4 ntdll.KiFastSystemC
EBX	7FFDE000
ESP	0012FFC4
EBP	0012FFFB
ESI	FFFFFFFF
EDI	7C920208 ntdll.7C920208
EIP	00401000 OLLYDBG.<ModuleEntry:

C 0 ES 0023 32bit 0 (FFFFFFFF)
P 1 CS 001B 32bit 0 (FFFFFFFF)
A 0 SS 0023 32bit 0 (FFFFFFFF)
Z 1 DS 0023 32bit 0 (FFFFFFFF)
S 0 GS 003B 32bit 7FFD0000 (FFF)
T 0 0 GS 0000 NULL
D 0 0 LastErr ERROR_FILE_NOT_FOUND
EFL 00000246 (NO, NB, E, BE, NS, PE, GI)

ST0 empty -UNORM BDEC 01050104 0
ST1 empty +UNORM 006F 005C0030 0
ST2 empty +UNORM 0069 002E0067 0
ST3 empty 0,0
ST4 empty 0,0
ST5 empty 0,0
ST6 empty 1.00000000000000000000
ST7 empty 1.00000000000000000000

FST 4020 Cond 3 2 1 0 E S I
FCW 027F Prec 1 0 0 0 Err 0 0 0

Address	Hex dump	ASCII
004B0000	00 00 60 43 4A 00 00 00	..*CJ..
004B0008	F4 45 4A 00 00 05 F0 5C	MEJ..&-l
004B0010	0A 03 00 04 03 53 00 00	..*SJ..
004B0018	00 04 93 83 4A 00 00 0A	..*S..
004B0020	AC 02 4D 00 00 0A 08 B1	*SJ..ll
004B0028	4A 00 00 0A F0 AC 4A 00	..J..0%J.
004B0030	00 0A CC 56 4A 00 00 01	..*S..J..0
004B0038	53 FE 4D 00 00 01 5C C2	*S..J..0+*
004B0040	4A 00 00 03 53 C6 4A 00	..*S..
004B0048	00 02 04 CB 4A 00 00 03	..06S..*
004B0050	9C CD 4A 00 00 00 FC D3	*-J..*%E
004B0058	4A 00 00 00 88 28 4A 00	..J...e(J.
004B0060	00 01 C3 3C 4A 00 00 00	..0+J..
004B0068	0C 46 4A 00 00 00 5D FJ..	..*S..*m
004B0070	4A 00 00 04 9C 83 4A 00	..*S..
004B0078	00 03 E0 C6 4A 00 00 02	..*0S..

0012FFC4 7C817067 RETURN to kernel32.7C817067
7C920208 ntdll.7C920208

0012FFC8 FFFFFFFF
0012FFD0 7FFDE000
0012FFD4 805448FD
0012FFD8 0012FFC8
0012FFDC 82D0B1D8
0012FFE0 FFFFFFFF
0012FFE4 7C8394C0 End of SEH chain
0012FFE8 7C817070 SE handler
00000000 kernel32.7C817070
00000000
00000000
0012FFF4 00401000 OLLYDBG.<ModuleEntryPc
0012FFF8 00000000
0012FFFC 00000000

Analysing OLLYDBG: 1017 heuristical procedures, 4602 calls to known, 7988 calls to guessed functions

Paused

Dynamic Analysis Techniques

Fuzzing

Roughly speaking, “fuzzing means...” (quoting Iñaki Rodríguez-Gastón)

Dynamic Analysis Techniques

Fuzzing

Roughly speaking, “fuzzing means...” (quoting Iñaki Rodríguez-Gastón)

- **Black-box approach** (at the beginning): no prior knowledge of the internal aspects of the program
 - **Evolved to a white-box approach**: state-of-the-art fuzzers “learn” from program behavior
- **Many abnormal inputs (unexpected, invalid, or random data) are given to the application**
- **The application is monitored for any signs of error**
 - Unexpected behavior
 - Crashes
 - Buffer overflow
 - Integer overflow
 - Memory corruption errors
 - Format string bugs

Dynamic Analysis Techniques

Fuzzing

Charlie Miller's "five lines of Python" dumb fuzzer

■ Found vulnerabilities in PDF readers and MS Powerpoint

```
numwrites = random.randrange(math.ceil((float(len(buf)) / FuzzFactor))) + 1
for j in range(numwrites):
    rbyte = random.randrange(256)
    rn = random.randrange(len(buf))
    buf[rn] = "%c"%(rbyte);
```

Dynamic Analysis Techniques

Fuzz Testing

An example: HTTP GET requests

- Standard HTTP GET request: `GET /index.html HTTP/1.1`
- **Anomalous requests**
 - `AAAAAA...AAAA /index.html HTTP/1.1`
 - `GET //////////index.html HTTP/1.1`
 - `GET %n%n%n%n%n%n.html HTTP/1.1`
 - `GET /AAAAAAAAAAAAAAAA.html HTTP/1.1`
 - `GET /index.html HTTTTTTTTTTTTTTP/1.1`
 - `GET /index.html HTTP/1.1.1.1.1.1.1.1`
 - etc.

Types of fuzzers

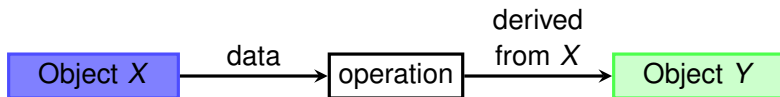
- **Mutation-based fuzzing**
- **Generational-based fuzzing**

Dynamic Analysis Techniques

Taint analysis

Measure what is the influence of the input data in the application

- Data comes from tainted sources (any external input) and ends up in tainted sinks
- *Flow* from X to Y : **an operation that uses X to derive a value Y**
- **Tainted value**: if the source of the value X is not trusted (e.g., user-supplied string)



Taint Propagation

- **Object X tainted object Y**
- **Taint operator** $t: X \mapsto t(Y)$
- **A taint operator is transitive**: $X \mapsto t(Y)$ and $Y \mapsto t(Z)$, then $X \mapsto t(Z)$

Dynamic Analysis Techniques

Taint analysis

Main challenges

■ Tainted addresses

- Distinguishing between memory addresses and cells is not always appropriate
- Taint granularity is important (bit, byte, word, etc.)

■ Undertainting

- Dynamic taint analysis does not handle some types of information flow correctly

■ Overtainting

- Deciding when to introduce taint is often easier than deciding when to remove it

■ Time of detection vs. time of attack

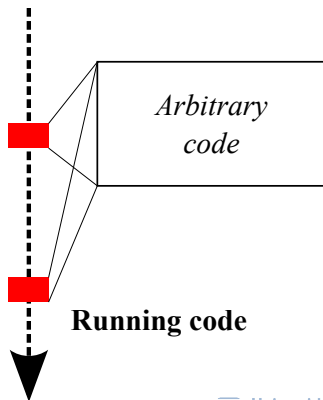
- When used for attack detection, dynamic taint analysis may generate an alert too late

Dynamic Analysis Techniques

Dynamic Binary Instrumentation

Adding arbitrary code during binary execution

- What insert? → **instrumentation function**
- Where? → **addition places**



Dynamic Analysis Techniques

Dynamic Binary Instrumentation

Adding arbitrary code during binary execution

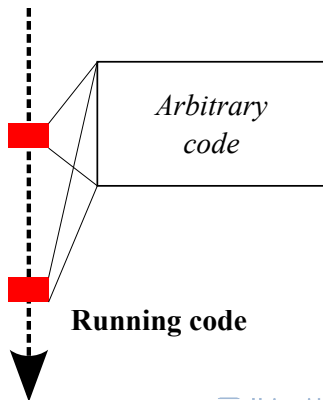
- What insert? → **instrumentation function**
- Where? → **addition places**

Advantages

- Programming language independent
- We can instrument proprietary software
- No need to recompile/relink each time
- Allows you to instrument a process already running

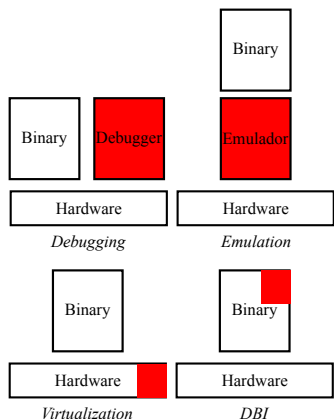
Main disadvantage

- **Overhead** ⇒ ↓ performance



Dynamic Analysis Techniques

Placing DBI in the context of dynamic analysis



- No transformation of the program file
- Full control over execution
- No need for architectural support

Credits: J-Y. Marion, D. Reynaud *Dynamic Binary Instrumentation for Deobfuscation and Unpacking*. DeepSec, 2009

Exploiting Software Vulnerabilities

Program Binary Analysis

© All wrongs reversed – under CC-BY-NC-SA 4.0 license



Universidad
Zaragoza

Dept. of Computer Science and Systems Engineering
University of Zaragoza, Spain

Course 2022/2023

Master's Degree in Informatics Engineering

UNIVERSITY OF ZARAGOZA

Seminar A.22, Ada Byron building

