CYBERSECURITY BOOTCAMP

GDG Algiers

Women Techmakers Algiers

Reverse Engineering 101

Table of Contents

What is reverse engineering

Reverse Engineering is the process of analysing a product to understand how it works

Why reverse engineering

Modifying software/hardware

Finding bugs

Finding security issues

Analyzing malware

Finding hidden features

How to perform reverse engineering

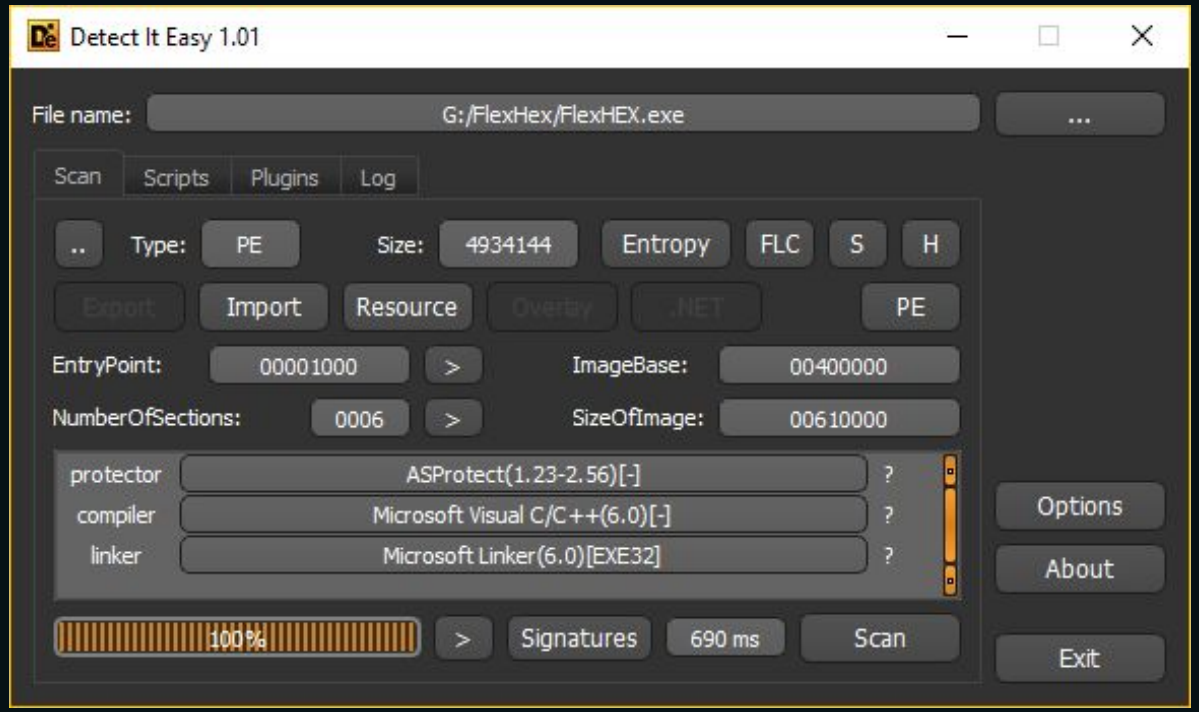
	Static analysis	Dynamic analysis
Identify	✓	✗
Execute	✗	✓

Static Analysis

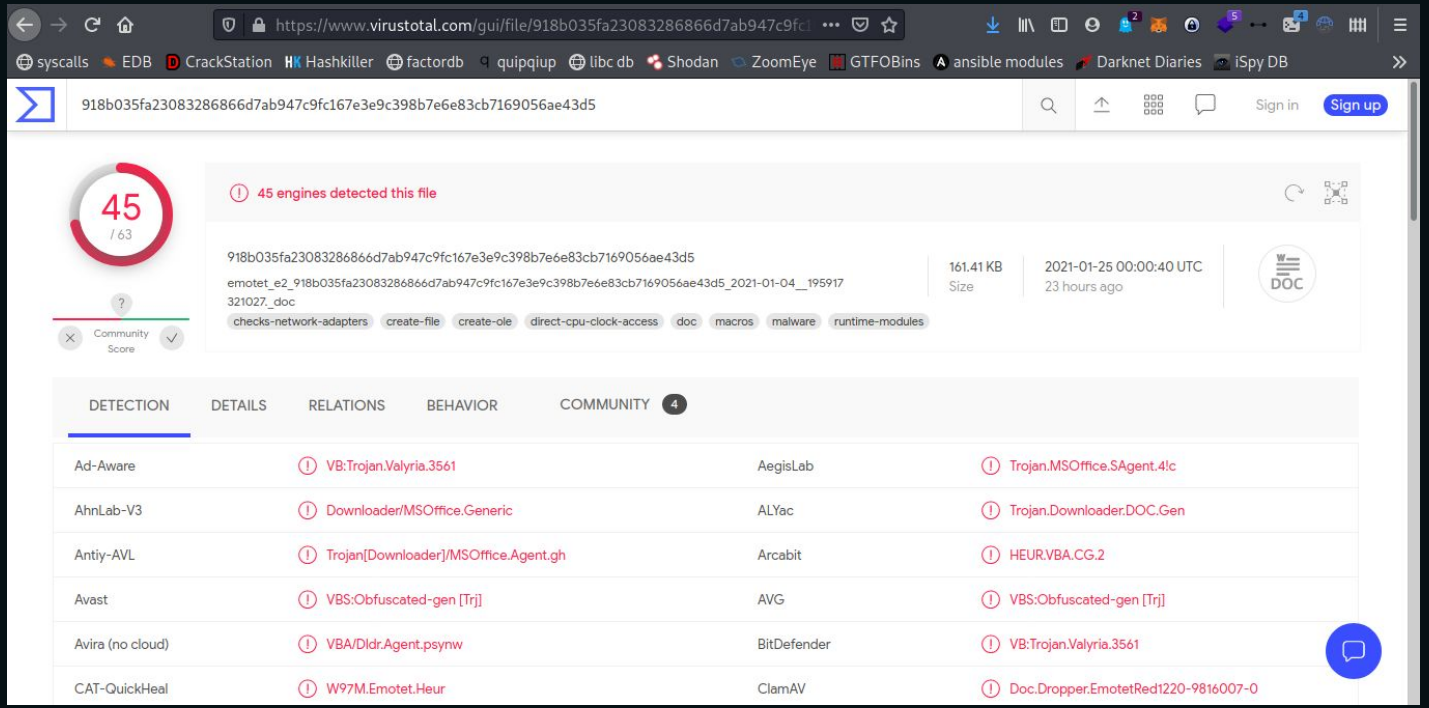
The different techniques for analyzing the sample program without executing it

- Looking up VirusTotal to check if it’s a known malware
- Looking at the strings inside the file
- Trying to know the runtime, compiler used (as well as any other packers)
- Trying to look at the code disassembly / decompilation

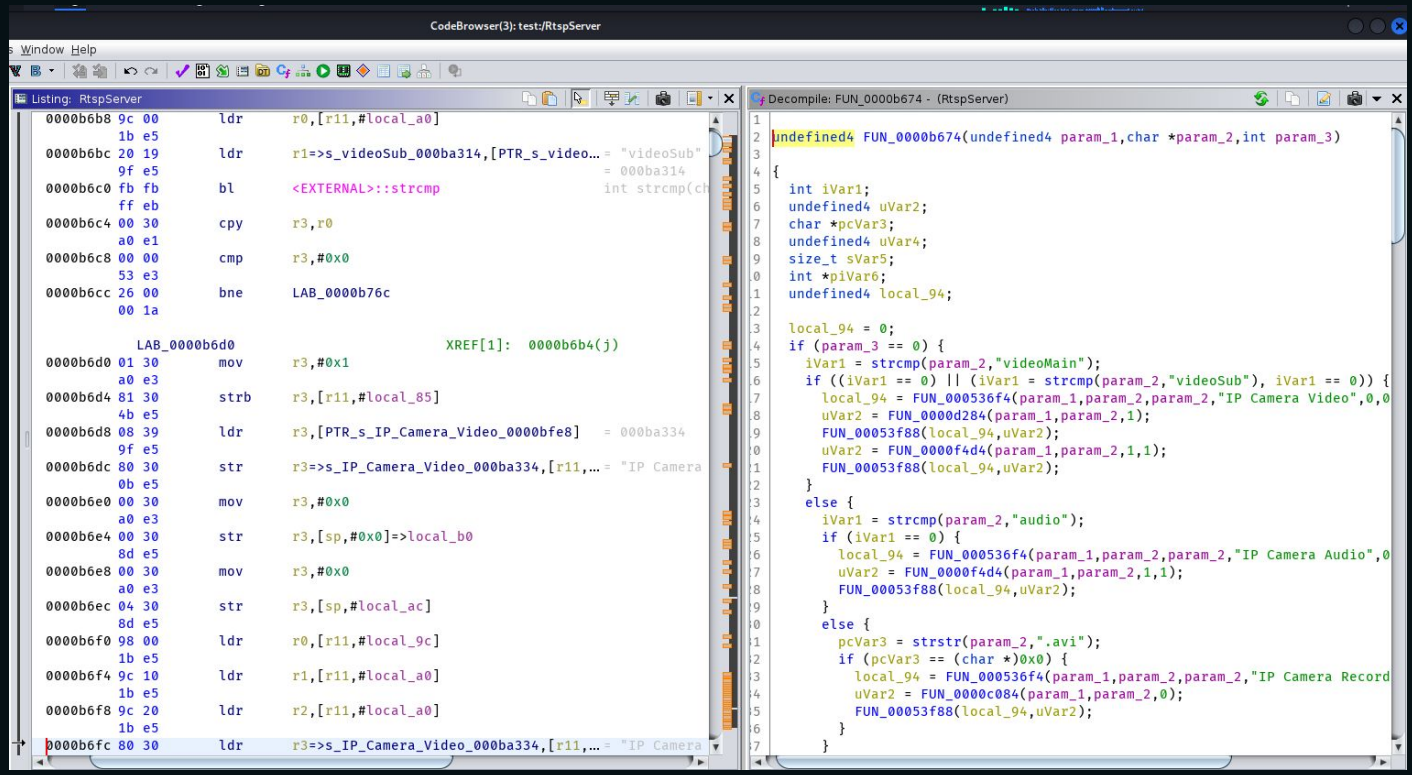
Detect it Easy on Windows



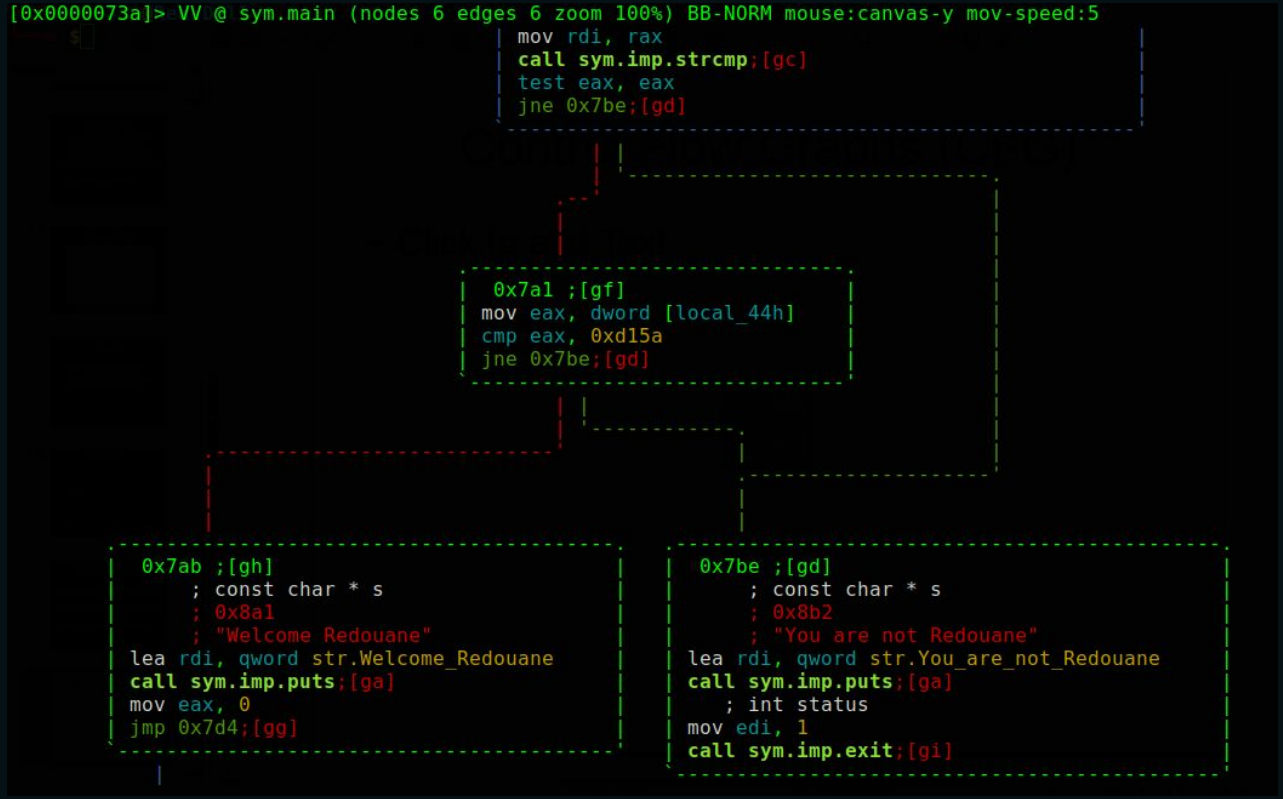
VirusTotal reporting malware



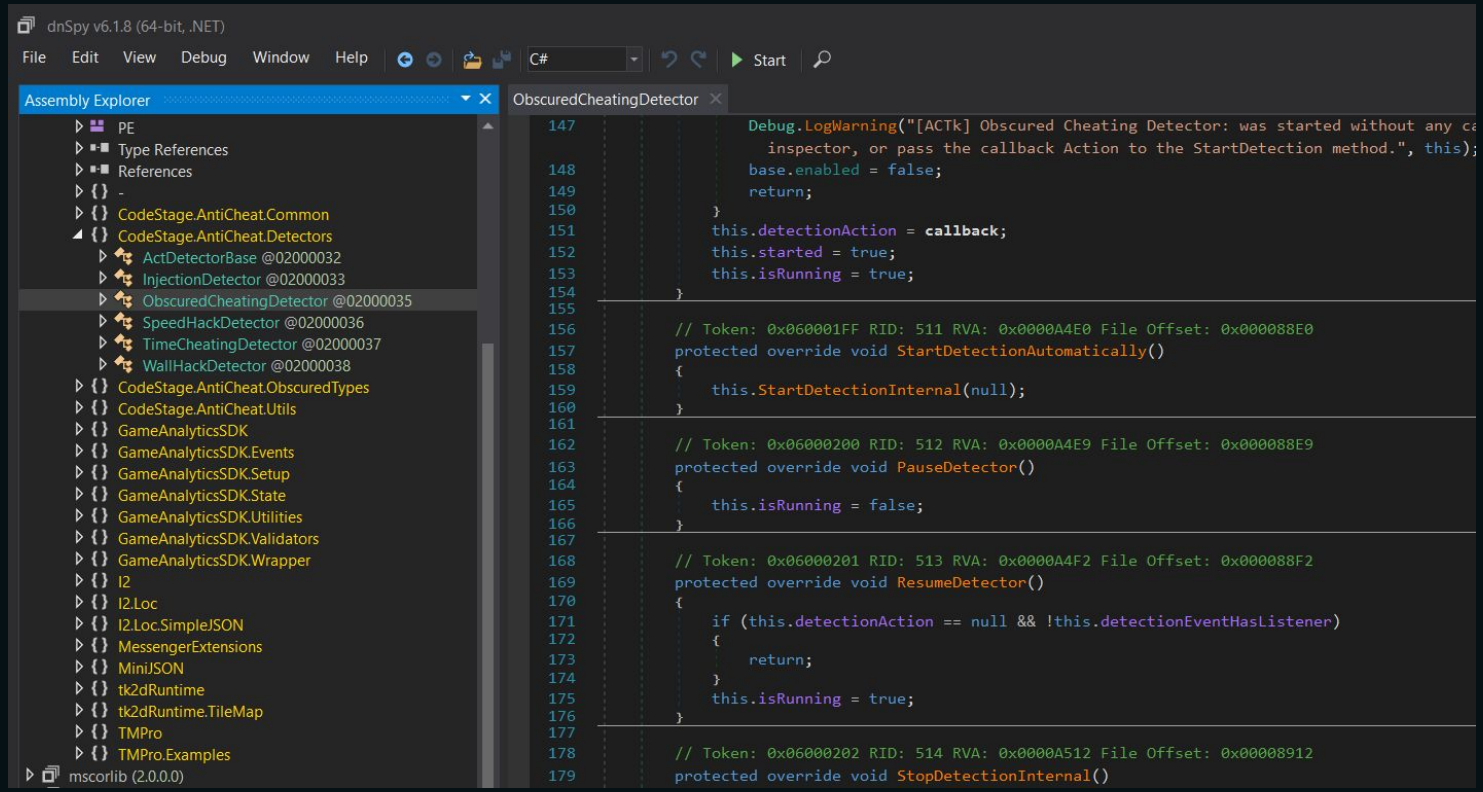
Ghidra CodeBrowser view



Control flow graph on Radare2



DnSpy decompiling a game anticheat



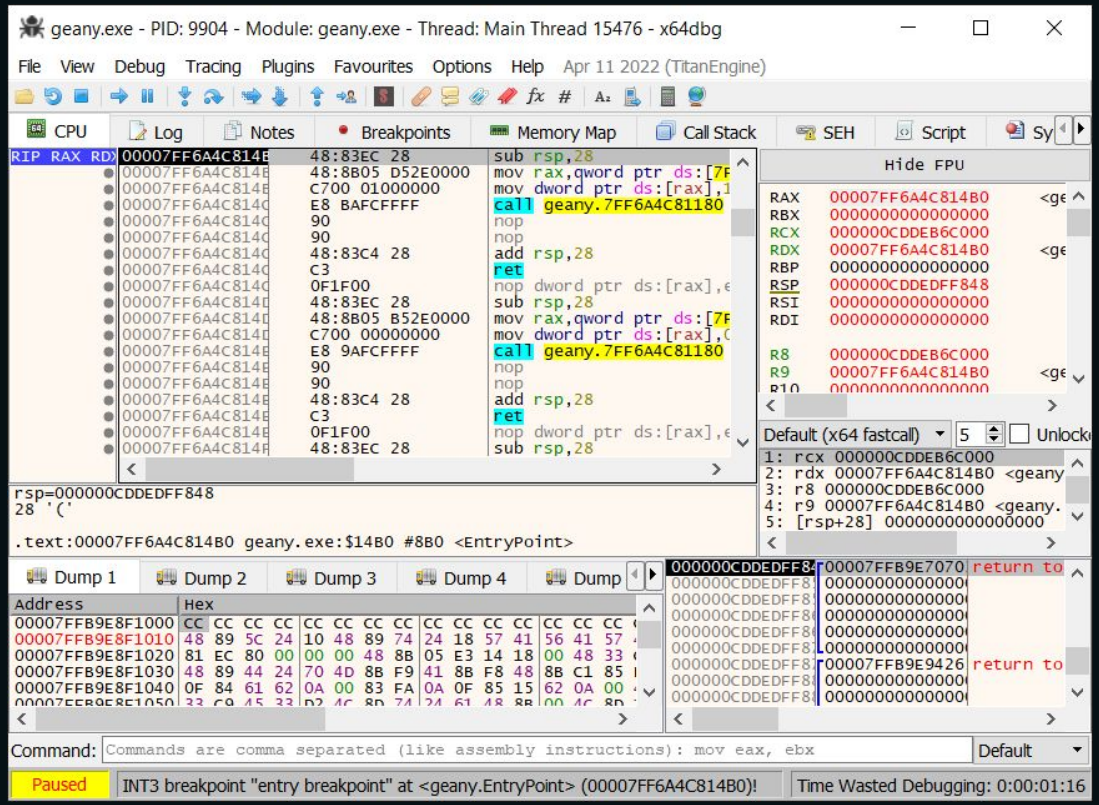
Dynamic Analysis

The techniques for analyzing the sample while it runs

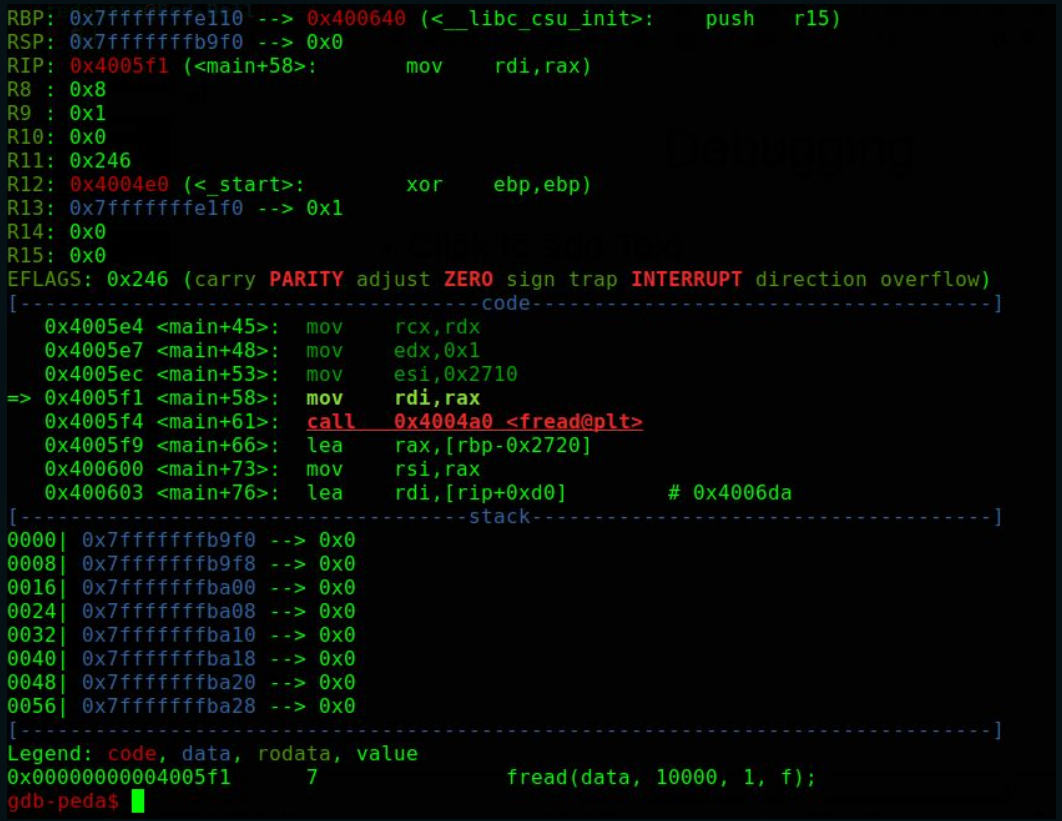
NOTE: HAS TO BE DONE IN AN ISOLATED ENVIRONMENT (VM)

- Intercepting and analyzing network traffic generated by the application
- Analyzing the API calls made by the application
- Debugging the sample
- Intercepting the decrypted data/code at runtime
- Performing instrumentation (injecting code into the application)
- Faking network traffic by setting up a fake dns / server (dnschef, inetsim)

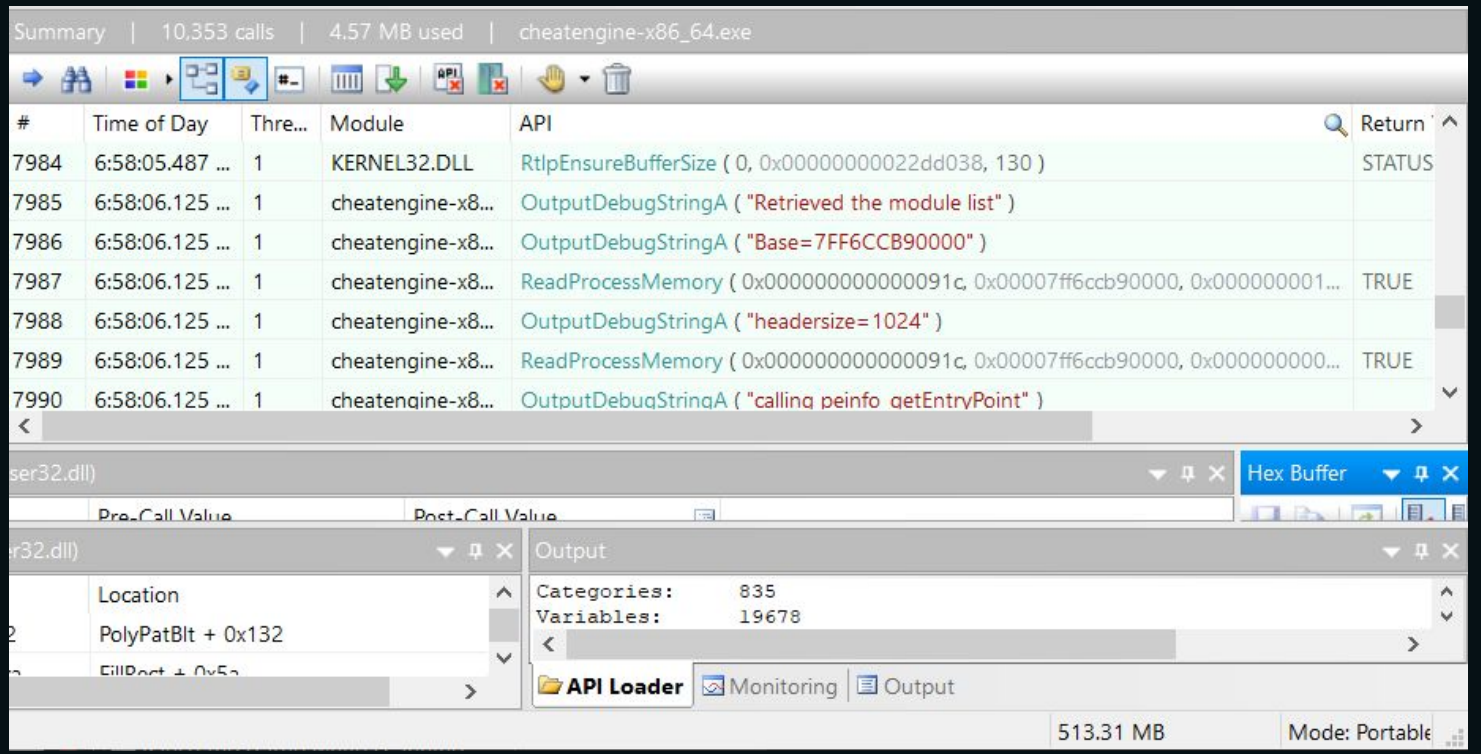
A debugging session in x64dbg



A debugging session in gdb



API montior on Windows

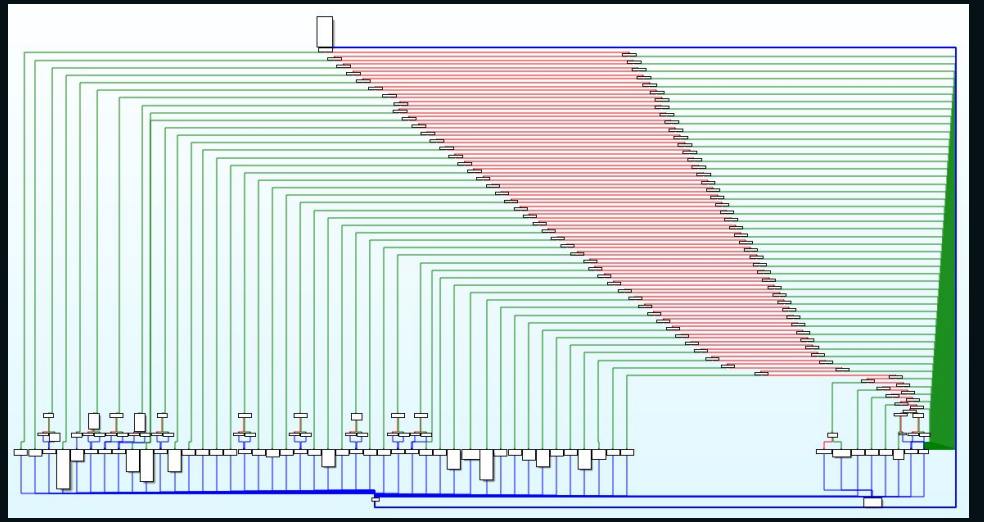


ltrace log on Linux

```
[pid 315113] memset(0×7ffeaeb1d850, '\252', 144) = 0×7ffeaeb1d850
[pid 315113] memcpy(0×311c00238b70, "\005\0\0\0\004\0\0\0", 8) = 0×311c00238b70
[pid 315113] strlen("no-zygote-sandbox")         = 17
[pid 315113] memcmp(0×311c0029a9c1, 0×55f4d423fd50, 4, 16) = 1
[pid 315113] memcmp(0×311c0029a971, 0×55f4d423fd50, 17, 0) = 0
[pid 315113] memcmp(0×55f4d423fd50, 0×311c0029a971, 17, 0) = 0
[pid 315113] sigfillset(~<31-32>)                = 0
[pid 315113] pthread_sigmask(2, 0×7ffeaeb1d520, 0×7ffeaeb1d5a0, 0×311c0020c01a) = 0
[pid 315113] clock_gettime(1, 0×7ffeaeb1d500, 0, 0) = 0
[pid 315113] fork()                              = 315116
[pid 315113] clock_gettime(1, 0×7ffeaeb1d500, 1, 1) = 0
[pid 315113] pthread_sigmask(2, 0×7ffeaeb1d5a0, 0×7ffeaeb1d6f0, 0×b3b633e00 <unfinished ...>
[pid 315116] <... fork resumed> )                = 0
[pid 315113] <... pthread_sigmask resumed> )     = 0
[pid 315116] open64("/dev/null", 0, 00 <unfinished ...>
[pid 315113] strlen("MPArch.ForkTime")           = 15
[pid 315113] memcpy(0×7ffeaeb1d4c9, "MPArch.ForkTime", 15 <unfinished ...>
[pid 315116] <... open64 resumed> )              = 11
[pid 315116] dup2(11, 0 <unfinished ...>
[pid 315113] <... memcpy resumed> )              = 0×7ffeaeb1d4c9
[pid 315113] pthread_mutex_trylock(0×55f4db266ea0, 15, 0, 50 <unfinished ...>
```

Mitigations and obstacles

Static analysis	Dynamic analysis
obfuscation	anti-debugging
encryption	ssl pinning
API hashing	VM detection
…	…



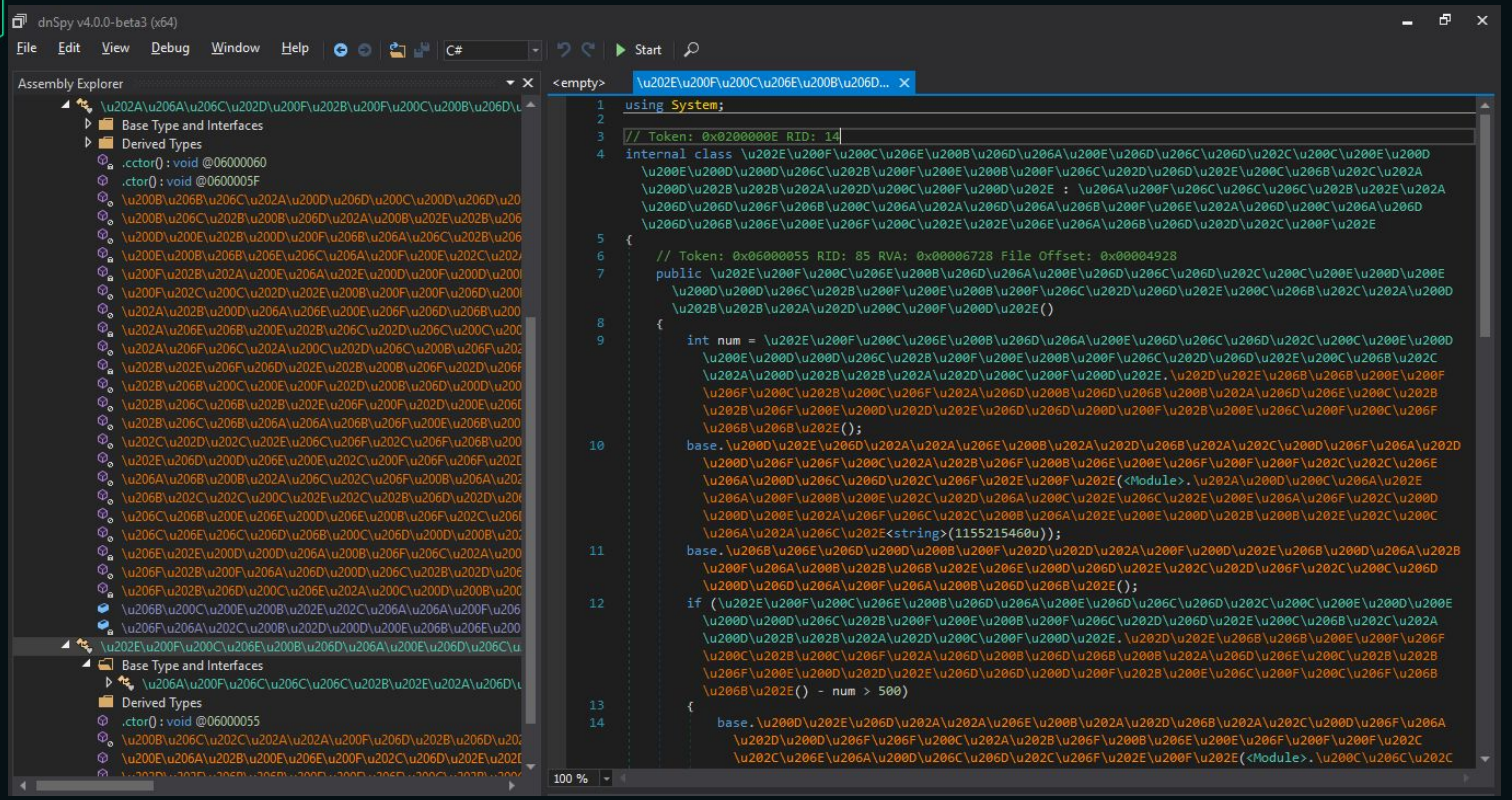
Obfuscated control flow as seen on IDA pro

Obfuscation

Techniques that aim to make it more difficult to read the code

⇒ just requires more effort and understanding from the reverser

Obfuscated C# program loaded in dnSpy



```
17   local_c = 'W';
18   local_b = 0x53;
19   local_a = 0x41;
20   local_9 = 0x52;
21   local_8 = 0x65;
22   local_7 = 99;
23   local_6 = 0x76;
24   local_5 = 0;
25   if (DAT_1002955c == (FARPROC)0x0) {
26     lpProcName = &local_c;
27     hModule = FUN_10009790();
28     DAT_1002955c = GetProcAddress(hModule,lpProcName);
29   }
30   (*DAT_1002955c)(param_1,param_2,param_3,param_4,param_5,param_6,param_7);
31   return;
32 }
33
```

Example dynamic API lokup from PlugX malware

API Hashing

Instead of calling API functions directly, the sample uses a function that looks up API functions by hash

From static analysis, it’s hard to figure out which function is being called.

```
$ cat /proc/self/status | grep TracerPid
TracerPid:      0
$
```

TracerPid on Linux, if a debugger is attached, it will be non-zero

Debugger detection

- Checking the process-list
- Trying to attach as a debugger
- Trying to trigger and handle exceptions
- Calculating timing to detect breakpoints
- Checking for integrity to detect software breakpoints

VM detection

- Checking the process-list / services
- Checking the MAC address
- CPUID instruction
- Timing measurement (RDTSC instruction)

Conclusion

- Mitigations don’t “prevent” reverse engineering.
- They just make it require more time and effort.

Their goal: Make sure that:

The effort and resources required ⩾ Outcome from reverse engineering.

Practice time

Any questions?