



Evolution of Binary Code Analysis

Jason Garman
Chief Technology Officer, Kyrus



Why am I here?

- I've been analyzing all sorts of binary code for about 8 years
 - For commercial and government entities
 - Mostly malicious code, but also ...
 - I've lost my source code! How do I read my data?
 - Did he steal my code?
 - ... you get the idea
- Now I get to step back and think of new ways to approach the problem
- **So... how have I seen the tools & techniques used in code analysis evolve over the years?**

In the beginning...



Tools, then and now

(Insert obligatory IDA Pro Screenshots here)

```
0100762C      push    eax                ; dwLanguageId
0100762D      push    eax                ; dwMessageId
0100762E      lea    eax, [ebp+Buffer]
01007634      push    eax                ; lpSource
01007635      push    FORMAT_MESSAGE_ALLOCATE_BUFFER or FORMAT_MESSAGE_FR
0100763A      call   ds:FormatMessageW
01007640      mov    eax, [ebp+var_4]
01007643      leave
01007644      retn
01007644  sub_10075F7  endp
01007644
01007645  ; ----- S U B R O U T I N E -----
01007645  ; Attributes: bp-based frame
01007645  sub_1007645  proc |near                ; CODE XREF: sub_10064BB+15C↑p
01007645                                     ; sub_1007645+AA↓p
01007645
01007645  Type          = dword ptr -14h
01007645  cbValueName   = dword ptr -10h
01007645  phkResult     = dword ptr -0Ch
01007645  var_8         = dword ptr -8
01007645  cbData        = dword ptr -4
```

```
IDA View-A
; Attributes: bp-based frame
; DWORD __stdcall ThreadProc(LPVOID)
ThreadProc proc near
arg_0= dword ptr 8
push    ebp
mov     ebp, esp
sub     esp, 8
push    esi
push    edi
push    2710h                ; dwMilliseconds
call   ds:__imp_Sleep@4    ; Sleep(x)
mov     eax, [ebp+arg_0]
test   eax, eax
jnz    short loc_401727
loc_401727:
mov     eax, 0FFFFFFFFh
pop     edi
pop     esi
mov     esp, ebp
pop     ebp
retn   4
ThreadProc endp ; sp-analysis failed
```

New techniques?

- Volatile memory acquisition & analysis
 - Enables analysis of “live” binaries as running on system
- Focus on automation, less manual static analysis
 - Automated clustering techniques
 - Automated *dynamic sandbox analysis*
- Don't forget:
 - It takes 10 years for a product to reach “usable” state
 - We are still innovating!
 - Unavoidable subjectivity in analysis



What's next?

