# All Software is Open Source

AN INTRODUCTION TO

REVERSE ENGINEERING

**DMITRIY BERYOZA** 



#### About me

Dmitriy Beryoza

@0xd13a on Discord/Twitter

https://www.linkedin.com/in/beryozad





- Senior Security Researcher at Vectra AI
- Prior to that Pentester/Secure Software Development Advocate with X-Force Ethical Hacking Team at IBM Security
- 25+ years in software design and development
- Ph.D. in Computer Science, CEH, OSCP, CISSP, CCSP
- Interests: reverse engineering, secure software development, CTFs
- Live in Ottawa, Canada; originally from Russia

# <sup>b</sup> Agenda

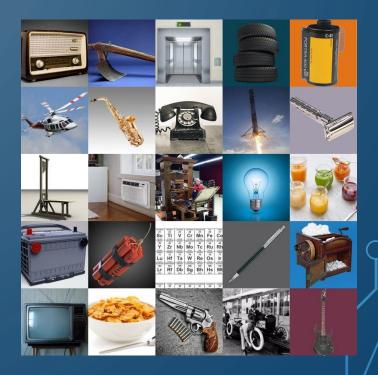
- What is Reverse Engineering
- Applications of RE
- What Can You Reverse and Why?
- Legal Issues
- Categories of RE
- Code Examples
- Tools
- Common Obstacles
- How to Practice
- General Strategies
- Resources



# What is Reverse Engineering

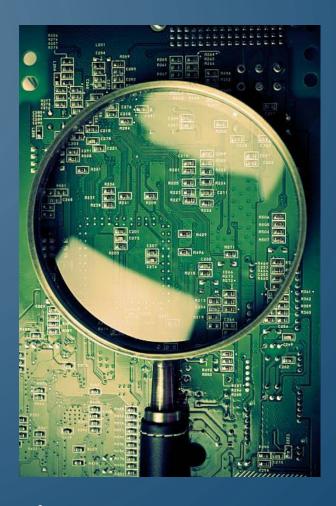
**Reverse Engineering** (or *reversing*) - process of understanding how a device, system, piece of software works by examining it (without having access to original design specs, source code, etc.)

- Becomes necessary when the original design information is not available or intentionally withheld
- Humans have been doing RE forever
  - Many inventions in human history have been analyzed and copied
- All natural sciences are essentially RE
  - Nature doesn't come with a manual
- You have probably done RE in your life more than once
  - The toy you pulled apart as a kid to see what's inside
  - The clock you tried to fix (unsuccessfully)



#### Modern Uses for RE

- Security/vulnerability research
- Pentesting
- Malware analysis
- Military/intelligence work
- Scientific research
- Commercial research/product compatibility
- Independent quality control
- Patent infringement detection
- ...and more
- We will talk about software RE
  - But that includes most hardware too much of modern hardware is driven by software



# Why Should You Do RE?

- A lot of software and hardware around us are "black boxes"
  - Documentation and architecture blueprints are scarce or withheld
- We want to be able to:
  - Analyze product for safety/quality
  - Analyze product for security/hidden backdoors
  - Recover legacy knowledge
  - Build compatible products
  - Patch in behavior changes or bug fixes
- RE gives you that power
  - "When you know assembly all software is open source" -- Unknown



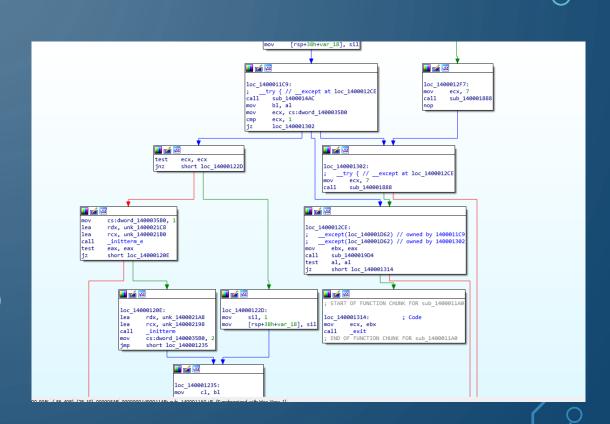
# Why Should You Do RE? (cont.)

- Required skill if you are in Infosec
  - Vulnerability analysis (most CVEs require some RE)
  - Pentesting (analyzing customer software for holes)
  - Incident response (samples of ransomware, malware, phishing, C2, remote shells)
- RE can be lots of fun!
  - Gives you sense of discovery and accomplishment
  - You are overcoming challenges and discovering secrets
  - You will probably like it if you like:
    - Putting together puzzles
    - Solving crosswords
    - Geocaching
    - •



#### What Can You Reverse?

- Pretty much anything!
- Windows, macOS, Linux natively compiled executables
- Compiled Java, .Net, Python portable code
- iOS/Android apps
- Minified/obfuscated JavaScript and WASM
- Obfuscated PowerShell
- Obfuscated Office VBA
- Compiled automation scripts (AutoIt, AutoHotKey, etc...)
- BIOS/bootloaders
- Shellcode
- Hardware firmware
- ...and the list goes on



# <sup>b</sup>Legality of RE

#### Big caveat - I'm not a lawyer, and this is not legal advice

- Is RE Legal? It depends...
- Most license agreements contain anti-RE clauses
  - Vendors want to protect their intellectual property
  - Vendors want flaws to be harder to find
- Existing laws restrict RE
  - Digital Millennium Copyright Act (DMCA)
  - Computer Fraud and Abuse Act (CFAA)
  - Copyright law
  - EU Directive 2009/24
  - ...and others

(d) You may not, and you agree not to or enable others to, copy (except as expressly permitted by this License), decompile, reverse engineer, disassemble, attempt to derive the source code of, decrypt, modify, or create derivative works of the Apple Software or any services provided by the Apple Software

#### Apple iOS 14 License Agreement

(vi) reverse engineer, decompile, or disassemble the software, or attempt to do so, except and only to the extent that the foregoing restriction is (a) permitted by applicable law; (b) permitted by licensing terms governing the use of open-source components that may be included with the software; or (c) required to debug changes to any libraries licensed under the GNU Lesser General Public License which are included with and linked to by the software; and

Windows 10 Retail License

# <sup>b</sup>Legality of RE (cont.)

- You have more protection if you own the device or software you are reversing
- The Fair Use defense in Copyright Law
  - Using copyrighted works for good-faith security research is likely fair use
- Under DMCA legal owner of the program may RE it and circumvent its protection to achieve "interoperability"
- Intent and what you do with information discovered through RE is key

• For deeper analysis see this guide by Harvard Law School & EFF: <a href="https://clinic.cyber.harvard.edu/files/2020/10/Security\_Researchers\_Guide-2.pdf">https://clinic.cyber.harvard.edu/files/2020/10/Security\_Researchers\_Guide-2.pdf</a>

# <sup>b</sup>Legality of RE: 3 Categories

#### Safe to do

- RE of non-commercial code or anything you yourself produce
- RE of commercial products done in private, without publishing or taking advantage of results
- Malware analysis

#### Tread carefully

- RE of commercial products for security research (bounty programs offer some protection)
- Publishing of limited reversed code/data as part of responsible vulnerability disclosure
- Publishing of tools based on RE discoveries
- Famous RE legal cases
  - Jon Lech Johansen arrested in 2000 for circumventing DVD copy protection; acquitted
  - Dmitry Sklyarov arrested after DEF CON in 2001 for circumventing Adobe protections; charges eventually dropped
- Even if what you are doing is not breaking the law, legal harassment is possible

#### Lawyer-up!

- Large-scale disclosure of discovered proprietary information, or (even worse) profiting from it
- Building compatible or competing products



# <sup>o</sup>Types of RE

- Data/communications analysis
  - Analysis of proprietary data formats (graphics, storage, databases)
  - Analysis of proprietary network protocols
  - Requires careful experimentation and use binary editors and capture tools
- Disassembly
  - Conversion of compiled binary to human-readable machine instructions
  - Native code (x86, ARM, MIPS, ...) and p-code (JVM, .Net, Python, ...)
- Decompilation
  - Attempt to recover original code from compiled binary
  - For native code results are spotty due to code optimizations; a C-like pseudocode is usually recovered
  - P-code is higher level and results are much better

# <sup>b</sup>Static vs. Dynamic Analysis

- Static involves analysis without execution
  - Helps where neither hardware nor emulators are available
  - Helps when debugging is not possible or anti-debugging measures are employed
- Dynamic is analysis though execution and debugging
  - Helps in cases of obfuscated/compressed/encrypted code and data
  - Helps confirm insight gleaned from static analysis
- Disassemblers and decompilers are often integrated with debuggers
- A combination approach often works best

# Disassembly Example: C

- Consider a simple C example
- Compiled representation looks intimidating...

```
#include <stdio.h>

// Password checker
int main( int argc, char *argv[] ) {

    // Secret password string
    char* password = "Sup3rSecretP4ssw0rd!!1";

    // Check the number of arguments
    if (argc == 2) {
        // Report success if the password matches
        if (!strcmp(argv[1],password)) {
            return 1;
        }
     }
     return 0;
}
```

```
8B 05 DD 2F 00 00 48 85 C0 74 02
                                                            Hfì.Hk.Ý/..H..Àt.
00001010
                                                             ÿÐHfÄ.Ã.....
                         08 C3 00 00 00 00 00 00 00 00 00
00001020
                                                             ÿ5â/..ÿ%ä/....@.
00001030
                               00 00 00 00 E9 E0 FF FF
00001040
00001050
                                                             líI‰Ñ^H‰âHfäðPTL
00001060
00001070
00001080
                         00 00 48 8D 05 A2 2F 00
00001090
                                                             øt.H<.>/..H..Àt.ÿ
000010A0
                                                             H.=y/..H.5r/..H)
000010B0
                                                             bH%öHÁî?HÁø.H.ÆH
000010C0
                                                            Ñbt.H<../..H..Àt.
000010D0
000010E0
          FF E0 66 0F 1F 44 00 00 C3 0F 1F 80 00 00 00 00
                                                            ÿàf..D..Ã..€....
000010F0
                                                             €=9/...u/UHf=ö..
00001100
                                                             ..H%åt.H<=./..è-
                                                             ŸŸŸèhŸŸŸÆ../...]
00001110
          FF FF FF E8 68 FF FF FF C6 05 11 2F 00 00 01 5D
00001120
                                                             Ã..€...Ã..€...
00001130
                                                             é{ÿÿÿUH%åHfì %}ì
                                                             H%uàH..1...H%Eøf
00001140
00001150
                                                             }ì.u%H<EàHfÀ.H</pre>.
                                                             H< UøH%ÖH%ÇèÁþÿÿ..
00001160
00001170
00001180
00001190
                                                             AWL.=O,..AVI%ÖAU
                                                            I‰őATA‰üUH.-@,..
000011A0
                                                             SL) ýHfì.èCbÿÿHÁý
000011B0
00001100
                                                             .t.1Û...L%òL%îD%
                                                             cAÿ.BHfÃ.H9ÝuêHf
000011D0
                                                             Ä.[]A\A]A^A Ã...
                                                            Ã...Hfì.HfÄ.Ã...
```

## Disassembly Example: C (cont.)

But if we open the executable in a disassembler things look much clearer

```
    Even though you may not know

                                                             PUSH
                                                                         RBP
                                                                        RBP ,RSP
                                                             MOV
          assembly language you can get a rough
                                                                        RSP ,0x20
                                                             SUB
                                                                        dword ptr [RBP + local 1c ], EDI
                                                             MOV
          idea of how C gets translated into it
                                                                        qword ptr [RBP + local 28 ], RSI
                                                             MOV
                                                                        RAX ,[s Sup3rSecretP4ssw0rd!!1 00102004 ]
                                                             LEA
                                                                        qword ptr [RBP + local 10 ], RAX => s Sup3rSecretP
                                                             MOV
#include <stdio.h>
                                                             CMP
                                                                        dword ptr [RBP + local 1c],0x2
                                                                        LAB 0010117a
                                                             JNZ
                                                                        RAX ,qword ptr [RBP + local 28 ]
                                                             MOV
// Password checker
int main( int argc, char *argv[] ) {
                                                             ADD
                                                                        RAX ,0x8
                                                                        RAX , qword ptr [RAX ]
                                                             MOV
                                                             MOV
                                                                        RDX => s Sup3rSecretP4ssw0rd!!1 00102004,qword
   // Secret password string
                                                                        RSI => s Sup3rSecretP4ssw0rd!!1 00102004, RDX
    char* password = "Sup3rSecretP4ssw0rd!!
                                                             MOV
                                                                        RDI , RAX
                                                             MOV
                                                             CALL
    // Check the number of arguments
                                                                         strcmp
                                                             TEST
                                                                        EAX , EAX
    if (argc == 2) {
       // Report success if the password matches
                                                                        LAB 0010117a
                                                             JNZ
                                                                        EAX ,0x1
       if (!strcmp(argv[1],password)) {
                                                             MOV
                                                                        LAB 0010117f
            return 1; _
                                                             JMP
                                                      LAB 0010117a:
                                                                        EAX , 0x0
                                                      LAB 0010117f:
    return 0;
                                                             LEAVE
                                                             RET
```

# Decompilation Example: C

- With a decompiler we can get an even better picture of the code
- Quality of decompilation varies and depends on complexity of an application

```
#include <stdio.h>

// Password checker
int main( int argc, char *argv[] ) {

    // Secret password string
    char* password = "Sup3rSecretP4ssw0rd!!1";

    // Check the number of arguments
    if (argc == 2) {

        // Report success if the password matches
        if (!strcmp(argv[1],password)) {

            return 1;
        }
    }
    return 0;
```

```
indefined8 main(int param_1,long param_2)

{
  int iVarl;

  if ((param_1 == 2) &&
      (iVarl = strcmp(*(char **)(param_2 + 8), "Sup3rSecretP4ssw0rd!!l"), iVarl == 0)) {
      return 1;
    }
    return 0;
}
```

# Decompilation Example: Java

```
import java.util.Base64;

public class re2 {
    // Testing for another password
    public static void main(String[] args) {
        if (args.length != 1) {
            System.out.println("Specify password on command line");
            return;
        }

        // The expected password is AnOtherH1dd3nP4ssword?!
        String encodedPassword = Base64.getEncoder().encodeToString(args[0].getBytes());
        if ("QW4wdGhlckgxZGQzblAOc3N3b3JkPyE=".equals(encodedPassword))
            System.out.println("Correct!");
        else
            System.out.println("Try again...");
    }
}
```

 Quality of decompilation of pcode language is much better

# Decompilation Example: C#/.Net

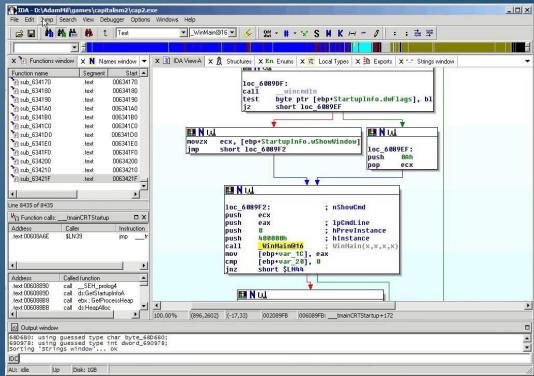
```
using System;
namespace re3
   class re3
        // This program will check if the password matches 'YetAnOth3rC00lPa55wOrd'
        static int Main(string[] args)
            if ((args.Length != 1) || (args[0].Length != 22)) {
                Console.WriteLine("Supply password on command line");
                return 0:
            // Load the expected password
            String key = args[0];
            // XOR against constant array
           byte[] xorArray = {0x82, 0xbe, 0xaf, 0x9a, 0xb5, 0xeb, 0xaf, 0xb3, 0xe8,
            0xa9, 0x98, 0xeb, 0xeb, 0xb7, 0x8b, 0xba, 0xee, 0xee, 0xac, 0xeb, 0xa9, 0xbf};
            for (int i = 0; i < xorArray.Length; i++)</pre>
                if (xorArray[i] != (Convert.ToByte(key[i]) ^ 0xDB)) {
                    Console.WriteLine("Wrong password!");
                    return 0:
            Console.WriteLine("Correct!");
            return 1;
```

```
using System;
namespace re3
   internal class re3
       private static int Main(string[] args)
           int result;
           if (args.Length != 1 || args[0].Length != 0x16)
                Console.WriteLine("Supply a password on command line");
                result = 0:
           else
               string text = args[0];
               byte[] array = new byte[]
                   0x82,0xBE,0xAF,0x9A,0xB5,0xEB,0xAF,0xB3,0xE8,0xA9,0x98,
                   0xEB,0xEB,0xB7,0x8B,0xBA,0xEE,0xEE,0xAC,0xEB,0xA9,0xBF
                for (int i = 0; i < array.Length; i++)
                   if (array[i] != (Convert.ToByte(text[i]) ^ 0xDB))
                        Console.WriteLine("Wrong password!");
                        return 0:
               Console.WriteLine("Correct!");
                result = 1:
            return result:
```

## Tools: IDA Pro

- The ultimate RE tool
- Tons of support for many processors and environments
- Disassembler + Debugger
- +Decompiler add-on (Hex-Rays)
- Plugins, scripting, ...
- Expensive US\$2K-4K
- ...but there is a Free and a Home (US\$365) version

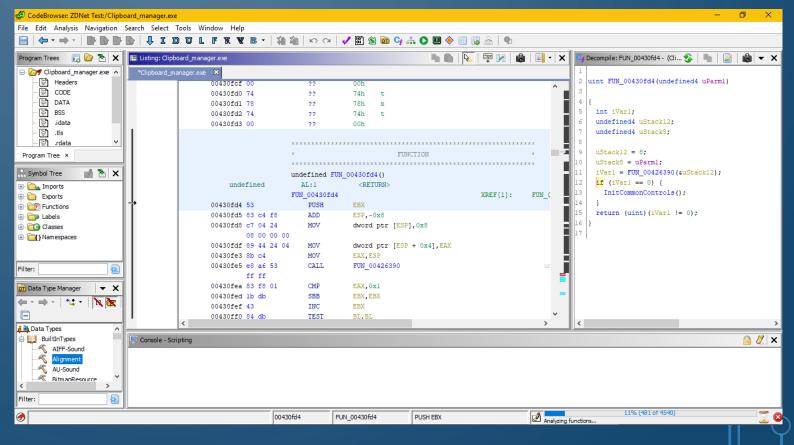




## Tools: Ghidra

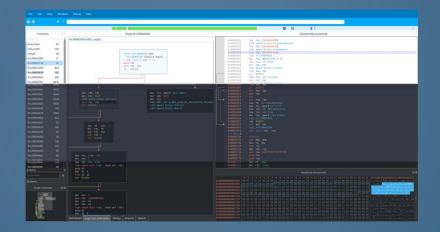
- Open-sourced by NSA (!)
- Many supported architectures
- Disassembler +Decompiler
- Now has a Debugger too
- Plugins, scripting, ...

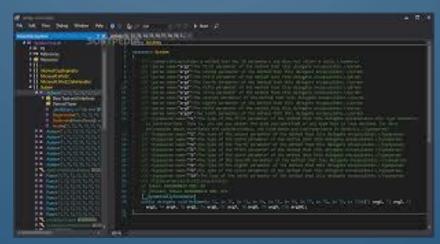




#### <sup>o</sup> More Tools

- Complete frameworks
  - Radare2 (+ Cutter GUI)
  - Binary Ninja
- Decompilers
  - retdec
  - Snowman
- Debuggers
  - OllyDbg
  - EDB
  - x64dbg





- VMs
  - FlareVM
  - REMnux
- ... and the list goes on

- .Net
  - dnSpy
  - ILSpy
- Java
  - JD-GUI
  - CFR
- Python
  - uncompyle6
- WASM
  - wasmdec

# Obstacles: Information Lost in Compilation

As part of compilation and optimizations code is made more difficult to reverse

- Lost:
  - Code comments
  - Meaningful names of functions and variables
  - Structure of data
  - Objects (e.g. C++)
- Variables move between stack and registers
- Execution flow constructs obscured (loops, conditional statements, exception handling, ...)
- Function inlining and embedding of libraries explodes body of code

# Obstacles: Counter-reversing Measures

- Software publishers know about RE and are not happy about it
- Many countermeasures are deployed to complicate analysis
- Obfuscation
  - Deliberate name obfuscation for scripting and p-code languages
  - Code minification
- Compression
  - Code and data compressed with "packers" (e.g. UPX)
- Encryption
  - Code and data decrypted at run time
  - Multiple nested levels and algorithms can be used



# Obstacles: Counter-reversing Measures (cont.)

- Anti-disassembly
  - Jumps into the middle of instruction
  - False branches
  - Self-modifying code
- Anti-decompilation
  - Program flow intentionally obscured
    - Jump to register
    - Jumps through return
    - Call as a jump
  - Useless/dead code insertion
  - Extreme optimization
  - VM implementation (e.g. OISCs Subleq, M/o/Vfuscator)



# Obstacles: Counter-reversing Measures (cont.)

- Anti-debugging
  - Debugger detection
  - VM detection
  - Sporadic insertion of breakpoints
  - Use of debug handlers (e.g. PTRACE) for execution
  - Signal-based execution
  - SEH-based execution
  - Alarms and time-dependent code



...and this is by no means an exhaustive list

# Obstacles: Counter-reversing Measures (cont.)

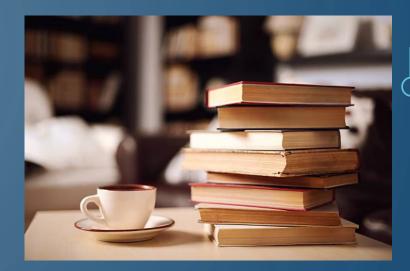
- It's an arms race
  - Good debuggers and VMs can hide their presence to a degree
  - Disassemblers/decompilers can overcome some counter-measures
  - Deobfuscation/decompression tools are available

- Expect these challenges in your RE work
- They can all be overcome with adequate techniques and tools
- There is nothing out there that can't be analyzed and reversed, it just a question of time and effort

# Practicing: LO - Starting Out

• The only way to get good at RE is to study and practice

- Required skills
  - Beginner ability to write in a programming language
    - Most languages are fine, but C/C++ an asset
  - Basic knowledge of target computer architecture
    - Processor registers, machine language instructions, memory organization
  - General idea of compiler operation
    - It helps to understand how high-level language is translated into low-level (because you will be doing the reverse)
  - Attention to detail and patience
- You can learn as you go, but need basics to enjoy it and to make good progress



# Practicing: L1 - Do-it-yourself Code Samples

- Write and compile a few primitive programs that use different aspects of the programming language
- Open compiled programs in a few decompilers and disassemblers
  - Study different tool features
  - Look at how compiled representation maps to original code
  - What code compiler adds to the executable (setup/teardown, function prologue/epilogue, library code)?
  - Try debugging
- This way you are taking one of the variables out of the equation, you know what the end result of your efforts should look like

# Practicing: L2 - Crack-me's

- These are challenges designed by someone else
- Typically small applications that are hiding a secret ("flag")
- Go at your own pace

- Great resource We Chall (<a href="http://www.wechall.net/">http://www.wechall.net/</a>)
- Directory of challenge sites (50+)
- Global scoreboard to gamify experience



## <sup>o</sup>Practicing: L3 - CTFs

- Online and in-person competitions
- All levels (from beginner to pro)
- Occur all over the world regularly
- RE challenges are present in almost every CTF
- Typically 48-hour events
- Adds a time pressure to the experience (puts learning in over-drive! :) )
- Recommended resource CTF Time (<a href="https://ctftime.org/">https://ctftime.org/</a>)
- A directory of online (and sometimes offline) CTFs
- Global scoreboard



# Practicing: L4 - Flare-On

- Annual competition created by FireEye (<a href="http://flare-on.com/">http://flare-on.com/</a>)
- The toughest RE competition in the world (as far as I know please let me know if you know of something better)
- 10-12 challenges of increasing difficulty
- 40 days to complete them
- Windows, macOS, Linux, scripting, mobile, embedded, esoteric...
- Many challenges are based on real malware samples or malware techniques
- Finalists get a prize



# General RE Strategies

- Make it an enjoyable experience
  - Requires a lot of time, patience and attention to detail
  - Make sure it's something you are motivated to do and enjoy
    - Solving a puzzle, finding a secret, trying to discover a bug...



- Keep your eyes on the prize
  - Fully understanding even an average application is unrealistic due to size
    - Simple application like Windows Notepad has ~13K machine instructions
  - Even if you had fully commented original source code it would take days and weeks
  - Focus on the information are looking for (algorithm, piece of data) and expand your understanding around that

# General RE Strategies (cont.)

- Combine static and dynamic analysis
  - Actively use debugger for analysis analyzing by hand is often too labor-intensive
  - Make the application do the heavy lifting (e.g. to bypass encryption and obfuscation algorithms)
  - Bypass problematic code, jump around
  - Make static analysis simpler by scripting some of the reversing work (e.g. decoding/decrypting data)

- Scripting for the win
  - Reversing requires a lot of processing
  - Take advantage of scripting to automate tasks
  - Use scripting available within tools

# General RE Strategies (cont.)

- Look for the signs
  - Magic numbers
  - Common algorithms (hashing, encryption, ...)
  - Systems calls and library functions
  - There are tools and plugins to help
- Document, document, document
  - A lot of info is lost in compilation
  - Body of code is usually very large, and details are easy to forget
  - As you analyze the code:
    - Add names to functions and variables
    - Restore data structures and objects
    - Comment important code blocks
  - Rinse and repeat



#### <sup>o</sup> Resources

- Lots of blogs, articles, writeups online
  - Google is your friend

- Conferences
  - REcon (held in Montreal and Brussels) <a href="https://recon.cx/">https://recon.cx/</a>



- Getting help online
  - Reverse Engineering Discord <a href="https://discord.com/invite/weKN5wb">https://discord.com/invite/weKN5wb</a> (invite may expire)
  - https://www.reddit.com/r/ReverseEngineering/
  - https://reverseengineering.stackexchange.com/

## Books

• Eilam, E. "Reversing: Secrets of Reverse Engineering"

Yurichev, D. "Reverse Engineering for Beginners"

• Sikorski, M. "<u>Practical Malware Analysis: The Hands-On Guide to Dissecting Malicious Software</u>"

Dang, B. "<u>Practical Reverse Engineering: x86, x64, ARM, Windows Kernel, Reversing Tools, and Obfuscation</u>"

• Eagle, C. "The IDA Pro Book, 2nd Edition: The Unofficial Guide to the World's Most Popular <u>Disassembler</u>"

## Wrap-up

- A lot of software and hardware around us are "black boxes"
- RE allows you to look inside and understand how they work
- You can help make them more secure, build compatible solutions, and modify them in new ways based on RE knowledge
- Plenty of tools are available to help reverse any product out there
- As you learn more about it, I'm sure you will enjoy RE as a fun and intellectually-rewarding activity!



# Q&A

@0xd13a on Discord/Twitter

https://www.linkedin.com/in/beryozad

Slides can be found at <a href="https://github.com/0xd13a/presentations">https://github.com/0xd13a/presentations</a>

Thank you!