# Python & Reverse Engineering Software

*by Alexander Hanel*

# Disclaimer

- Reverse engineering is a very very broad topic.
- A lot of overhead knowledge is required.
- There are no click and it’s reverse engineered tools.
- Please ask questions. Some of these topics are esoteric and will need explanations.

# Who Am I.

- Reverse engineer malware at my $dayjob
- Been programming in Python since 2010.
- Wrote the *The Beginner's Guide to IDAPython*, xxxswf.py and author of the blog *Hooked on Mnemonics*.
- Not 1337.
- @nullandnull

# Outline

- Why Python?
  - History from an RE perspective.
  - Why it was adopted.
- What kind of tasks?
- What modules?
- Awesome links!

# Why Python? - History

- Python has been in use by the reverse engineering community for over a decade.
- In 2005 on OpenRCE a number of influential tools were released in Python.
  - IDAPython and Pefile are the most popular.
- In 2005 and 2006 security companies started posting Python as a desired skill.

# Why Python? - Because...

- Python is friendly.
- Prototyping in Python is quick.
  - Continuous recompiling isn't.
- Adding modules is painless.
- Python is everywhere.
- Batteries included.
  - "import zlib" vs Visual Studio's Project properties -> Linker -> Input -> Additional dependencies.. etc

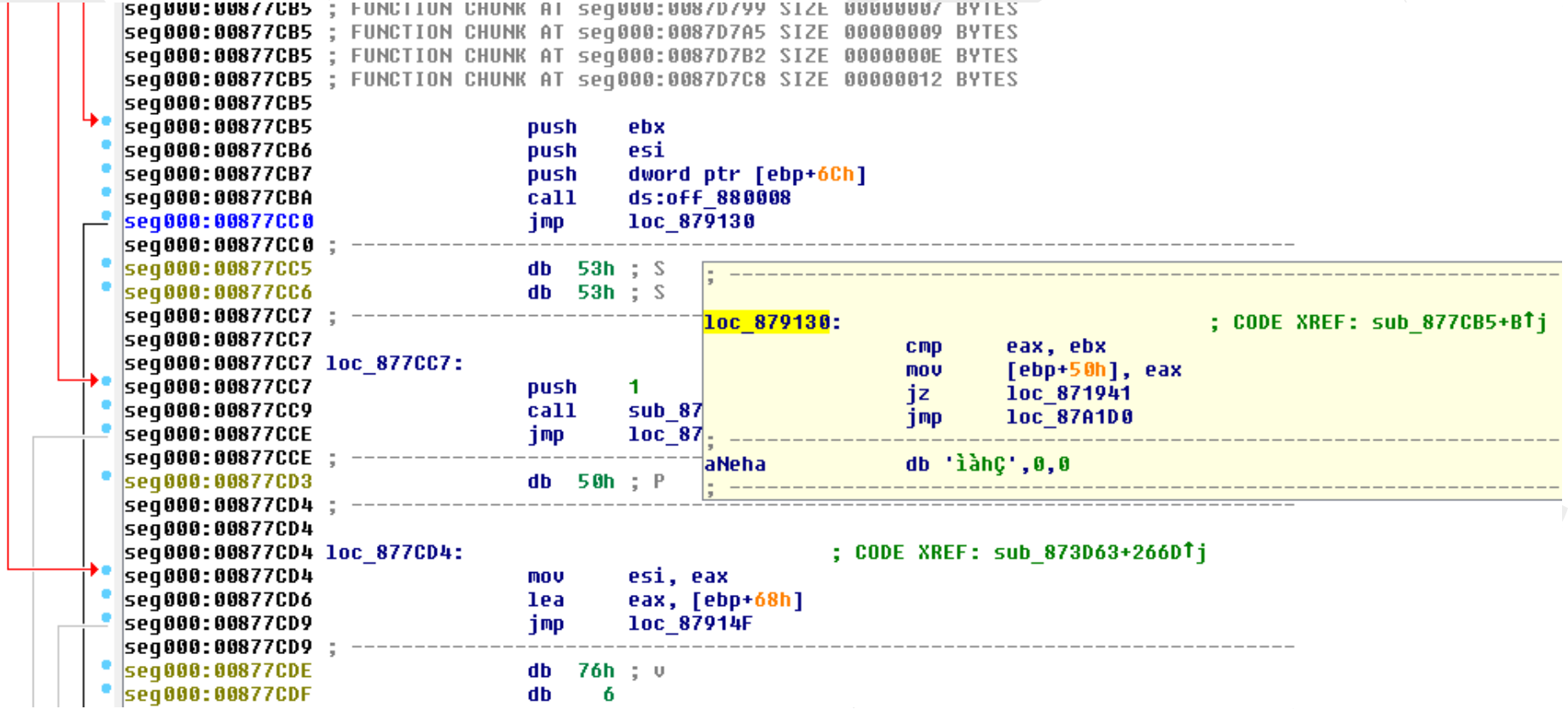
# What kind of tasks?

- It depends of the type of research.
- Python is commonly used for.
  - Binary analysis
  - Forensics
  - Malware analysis
  - Network analysis
  - Exploring file formats
  - Vulnerability and exploit analysis

# Binary Analysis - Tasks

- Disassembling code
  - Binary to Assembly
- Automating the analysis of code
- Deobfuscating code
- Solving Cracks Me
- Full binary analysis frameworks have been written in Python

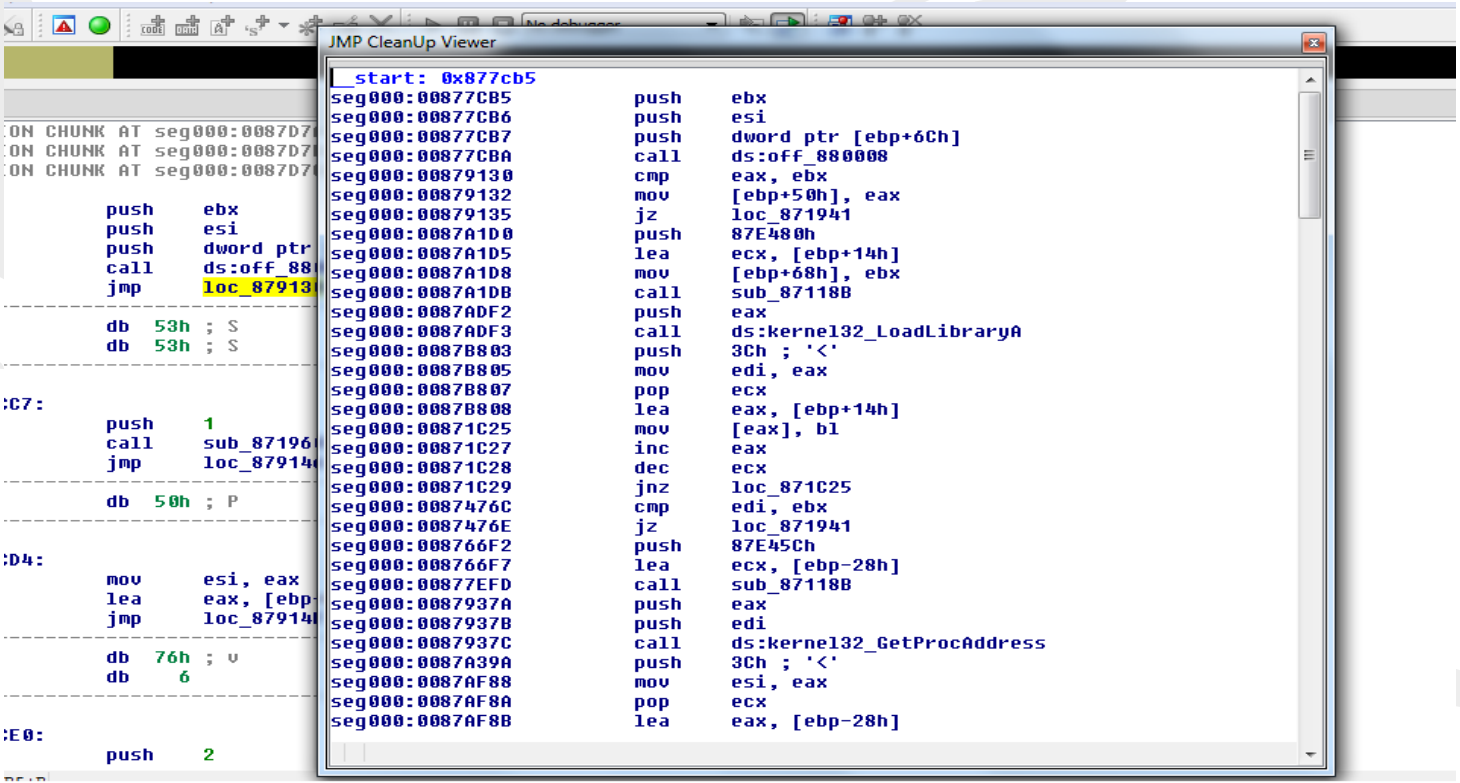
# Binary Analysis - Example



# Binary Analysis - Example

```
# Python code to remove JMPs from obfuscated code in IDA
# created by alexander dot hanel at gmail dot come
# # Note you will need to have your cursor at what is the start
# of the function or at least in the path.

from idaapi import *
import idautils
import idc
import sys

class JMPJMP:
    def __init__(self):
        self.ea = ScreenEA()
        self.errorStatus = 'Good'
        self.funcStartAddr = GetFunctionAttr(self.ea, FUNCATTR_START)
        self.checkFunctionStart()
        self.buffer = []
        self.count = 0
        self.condJmps = ['jo', 'jno', 'jb', 'jnae', 'jc', 'jnb', 'jae', 'jnc', 'jz', \
                         'je', 'jnz', 'jne', 'jbe', 'jna', 'jnbe', 'ja', 'js', 'jns', \
                         'jp', 'jpe', 'jnp', 'jpo', 'jl', 'jnge', 'jnl', 'jge', 'jle', \
                         'jng', 'jnle', 'jg']
        self.condJmpsAddr = set([])
        self.retn = ['retn', 'ret', 'retf']
        self.callAddr = set([])
        self.call = 'call'
        self.callByte = 0xe8
        self.jmp = 'jmp'
        self.visitedAddr = set([])
        self.target = set([])

    def getJmpAddress(self, addr):
        "returns the address the JMP instruction jumps to"
        return GetOperandValue(addr, 0)

    def checkFunctionStart(self):
        'checks if the address is valid'
        if self.funcStartAddr is BADADDR:
            print "Could not find find function start address"
            self.errorStatus = 'Bad!'
```

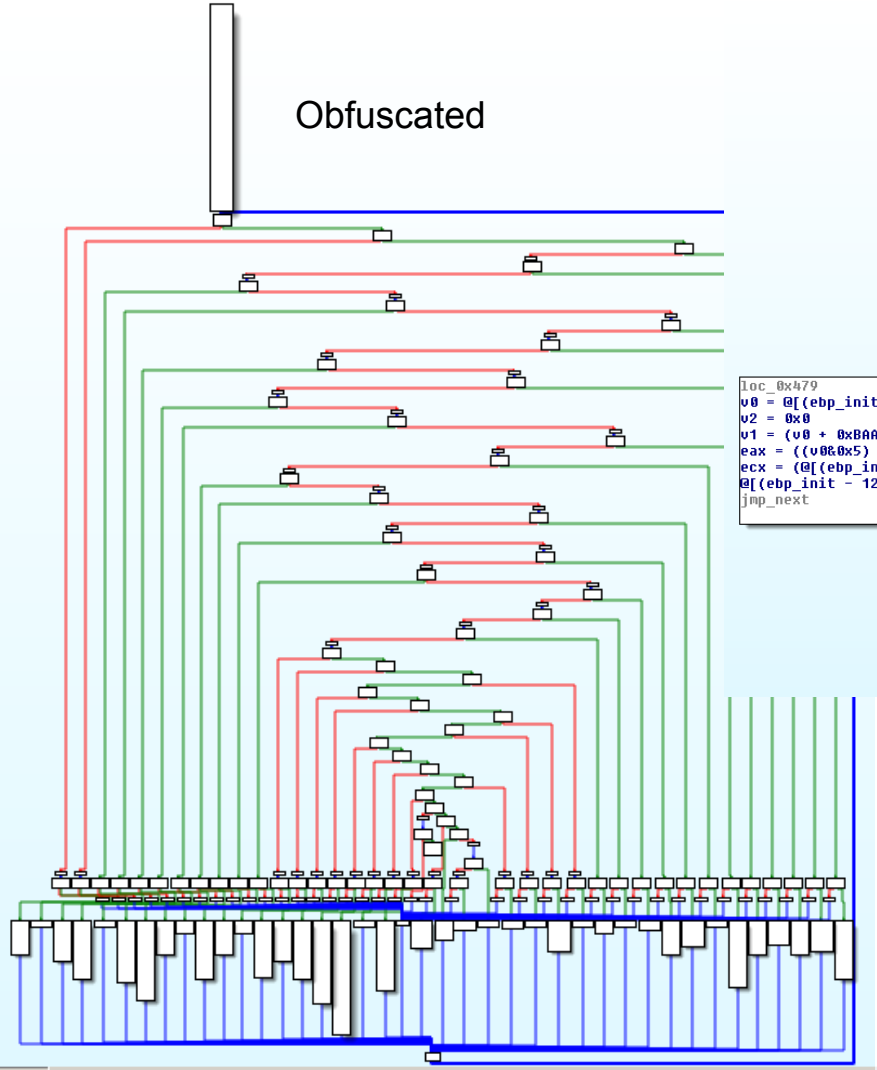
# Binary Analysis - Example



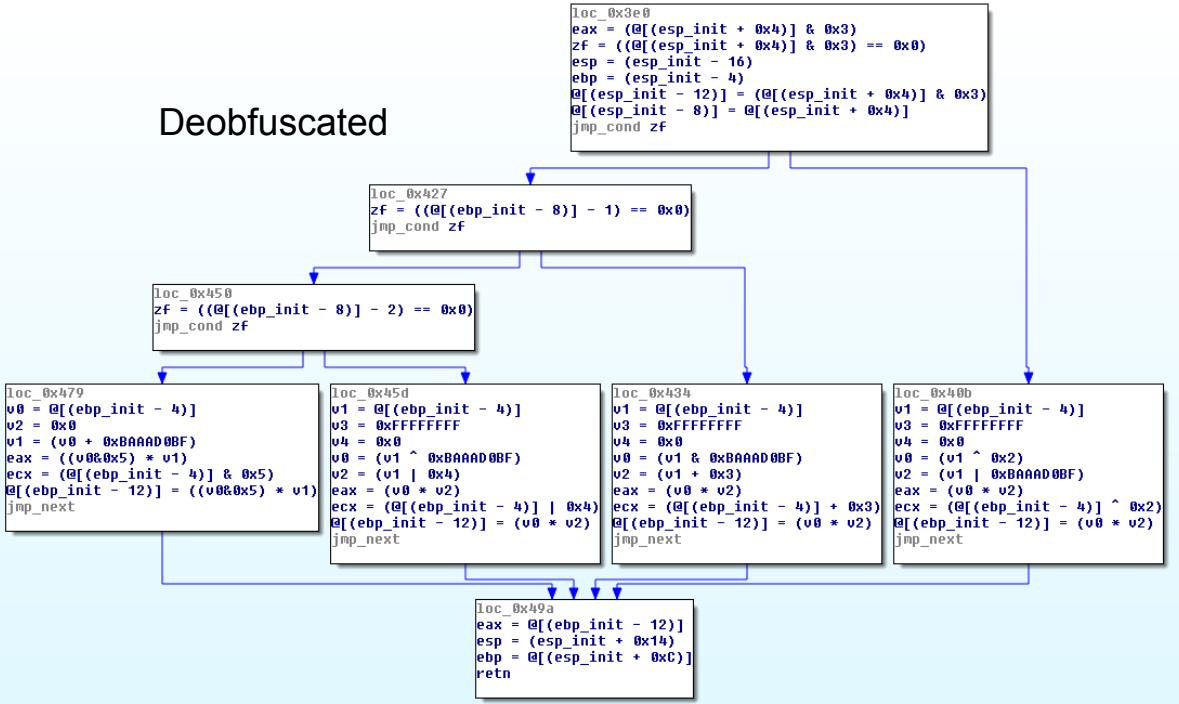
# Binary Analysis - Cool Examples!

- Using Microsoft's Z3 Theorem Prover to solve a CrackMe **.
- Deobfuscation: recovering an OLLVM-protected program **.
- Breaking Kryptonite's Obfuscation: A Static Analysis Approach Relying on Symbolic Execution **.

Obfuscated



Deobfuscated




Via Quarkslab blog

# Forensics - Why & Tasks

- Most tools are *not* platform dependent.
  - Example: Analyzing a Windows memory dump in Linux
- File and disk analysis.
  - Timelines
- Parsing the registry.
  - Extracting shellbag data to see folders accesses.
- Memory analysis.
  - Analyzing memory dumps

# Forensics - Example

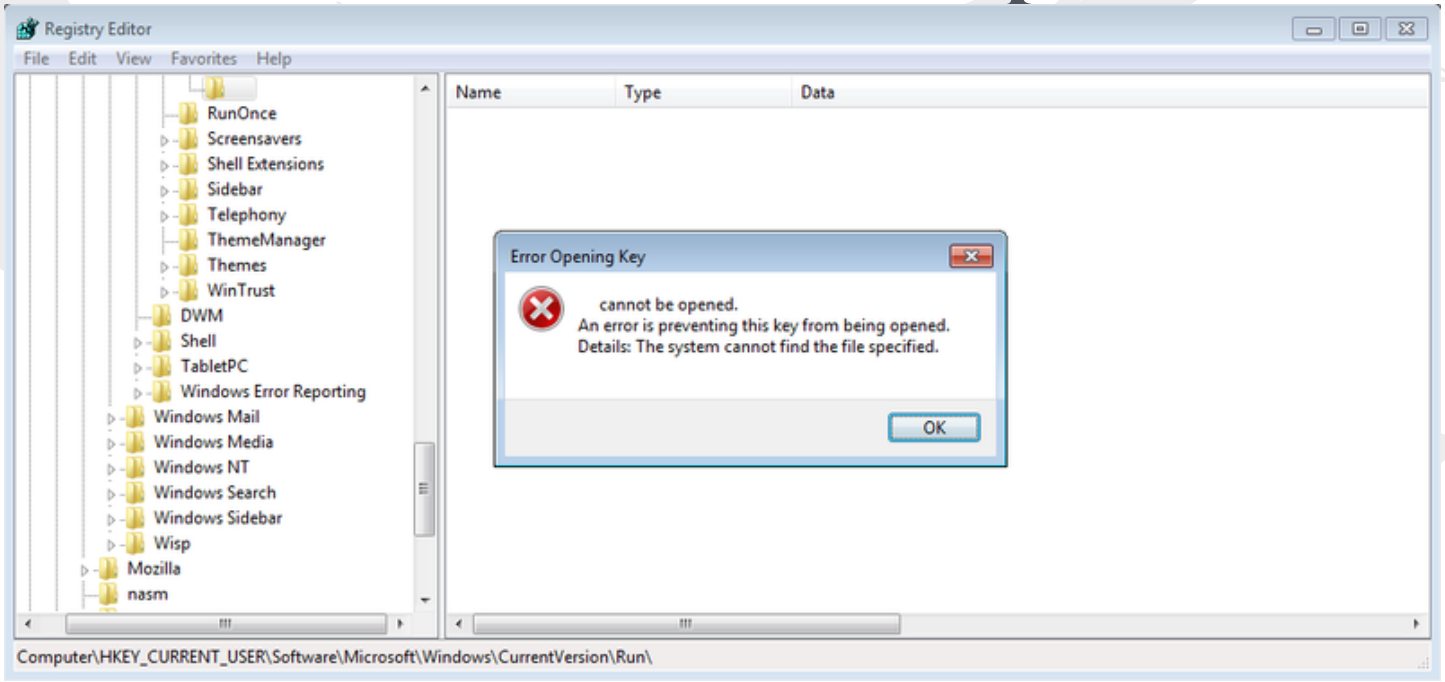



Image via G-Data blog

- Error is caused by regedit unable to display invalid characters written by poweliks. An analyst could copy %userprofile%/ntuser.dat to a separate machine and parse the hive using python-registry.

# Forensics - Cool Examples!

- python-registry Introduction by Willi Ballenthin **.
- Automating DFIR Series by David Cowen **.
- Stuxnet’s Footprint in Memory with Volatility 2.0 **.
- Extracting the Powelik's DLL from the Registry **.

# Malware Analysis- Tasks

- Automating the analysis of samples in a sandbox environment.
- Deobfuscation, decompressing and decoding data.
- Debugging and disassembling.
- Scanning files.
- Extracting data.
- Hard to describe so many amazing projects.

# Malware Analysis - Sample Automation

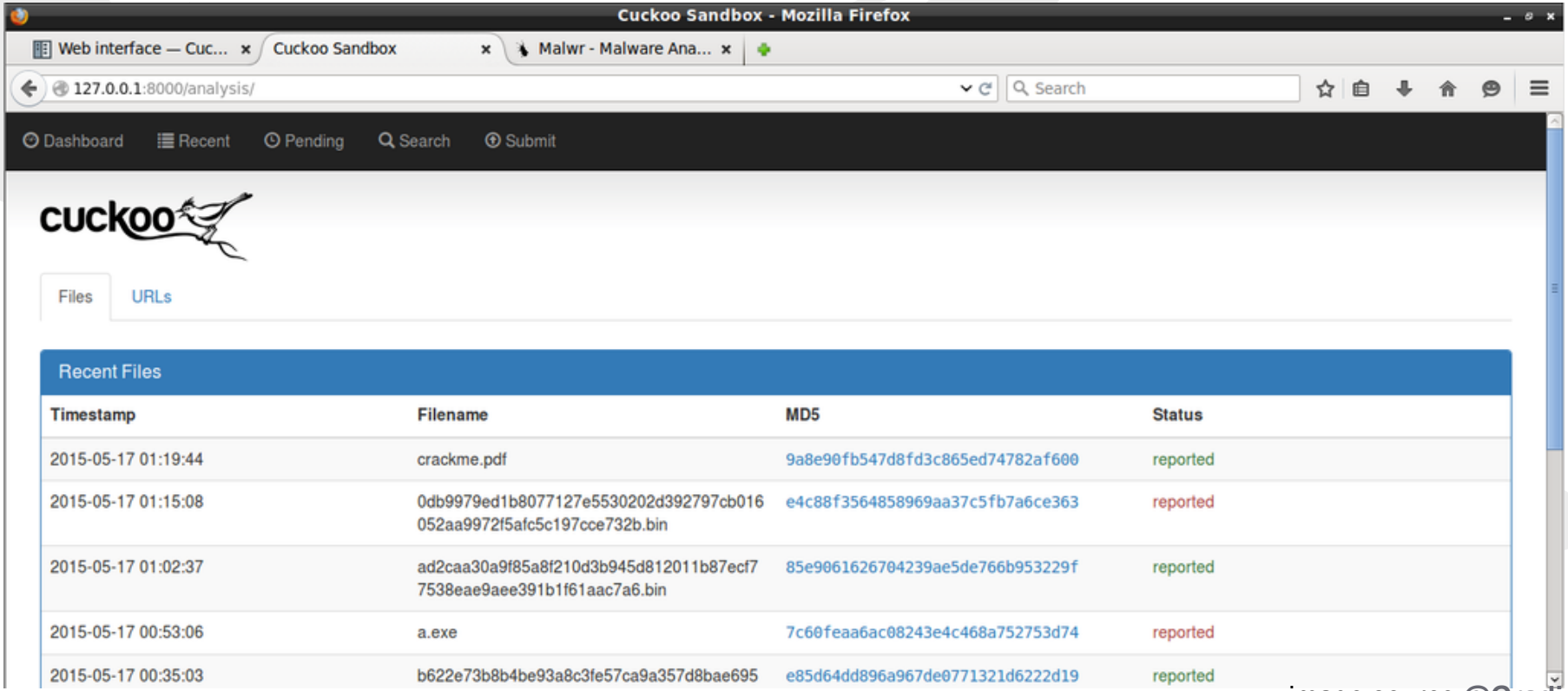



image source @GradiusX

# Malware Analysis - Deobfuscation Example

```actionscript
// obfuscated
package
{
    import flash.utils.*;

    public class II111IIIlllIl1 extends ByteArray
    {

        public function II111IIIlllIl1()
        {
            return;
        }// end function

        public function +lllIII111I11() : void
        {
            return;
        }// end function

        public function 11ll1III111I11() : int
        {
            return 0;
        }// end function

    }
}
```

```actionscript
//de-obfuscated
package
{
    import flash.utils.*;

    public class _tare extends ByteArray
    {

        public function _tare()
        {
            return;
        }// end function

        public function _secant() : void
        {
            return;
        }// end function

        public function _carat() : int
        {
            return 0;
        }// end function

    }
}
```

# Malware Analysis - Example

```
class ObfStrReplacer():
    """
    A module that can be used to de-obfuscate code by searching
    for strings that match a regular express pattern and replace
    them with more readable characters.
    """
    def __init__(self):
        self.regex_pattern = None
        self.compiled_regex = None
        self.file_glob_pattern = None
        self.test_regex = False
        self.script_name = None
        self.globbed_files = None
        self.word_list = [
                "abacus",  "iota",  "nu", "baryon", "ceres", "dean", "zipf",
                "mu", "epsilon", "lune", "fermat", "gamma", "carat", "gaudi",
                "ides", "alpha", "iris", "julia", "tare", "omicron", "pascal",
                "kappa", "aeon", "umbra", "secant", "lambda", "beta", "lemma",
                "eta", "mars", "nocebo", "occam", "chaos", "arc", "omega",
                "xenon", "pareto", "locus", "psi", "rho", "delta", "sigma",
                "pi", "simson", "tau", "gnomen", "theta", "atlas", "upsilon",
                "phi", "venus", "ogive", "surd", "xi", "zeta", "sabot", "chi",
                "kite"]
        self.match_set = set([])
        self.names = []
        self.name_mapping = {}

    def get_args(self):
        """
        gets the command line arguments.
        """
        parser = argparse.ArgumentParser(
            description='Replaces strings matched by a regular expression with more \
                         distinguishable text/strings.')
```

“Pretty” replacement words

```
    def run(self):
        self.get_args()
        self.get_files()
        self.compile_regex()
        if self.test_regex:
            self.print_regex_matches()
            return
        self.get_matches()
        self.create_str()
        self.string_to_name()
        self.replace_str()

if __name__ == "__main__":
    xx = ObfStrReplacer()
    xx.run()
```

# Malware Analysis- Projects!

- My Favorites or at least should be mentioned
  - Cuckoo Sandbox
  - Yara
  - winappdbg, pydbg, pykd and vivisect
  - Capstone Project
  - IdaPython
  - pefile
  - IDAScope

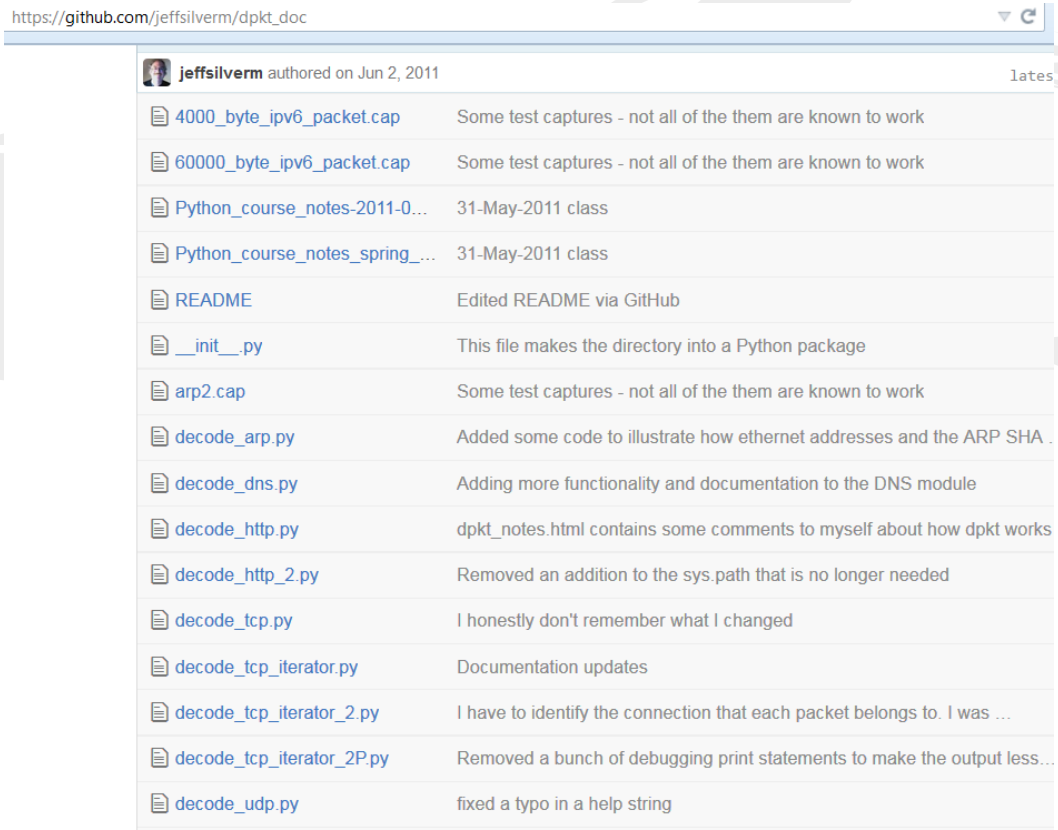
# Network Analysis- Tasks

- Protocol and decoding analysis
- Network and browser emulation.
- PCAP parsing.
- Packet creation, sniffing and manipulation.
- Custom passive DNS tool.
- Automating URL lookups

# Network Analysis- Projects!

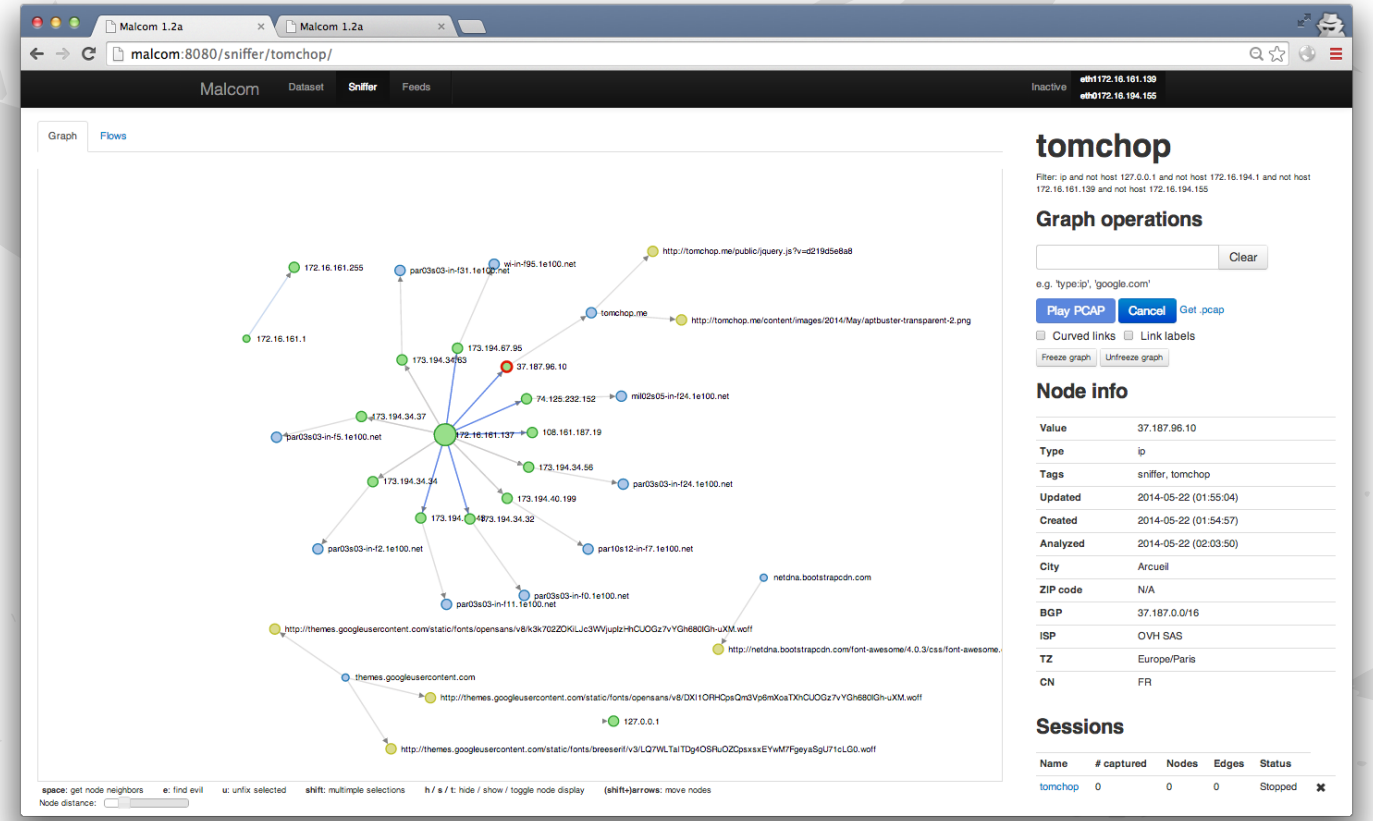
- Scapy or Dpkt
- Chopshop built on top of Pynids
- fakedns.py or Fakenet (python bindings)
- jsunpack
- Malcom

# Network Analysis- Example




dpkt - example code and documentation.

# Network Analysis- Example




Malcom - graphical view

# Exploring file formats- Tasks

- Carving out embedded files in a data streams.
- Exploring structured data.
- Decompressing files.
  - SWF files are compressed zlib.
- Writing binary parsers
- Analyzing and extracting firmware.

# Exploring file formats- Projects

- pdf-parser.py or peepdf
- oletools or oledump.py
- xxxswf.py
- pe-carv.py
- hachoir
  - hachoir-urwid
  - hachoir-subfile
- construct & vstruct

# Exploring file formats- Example

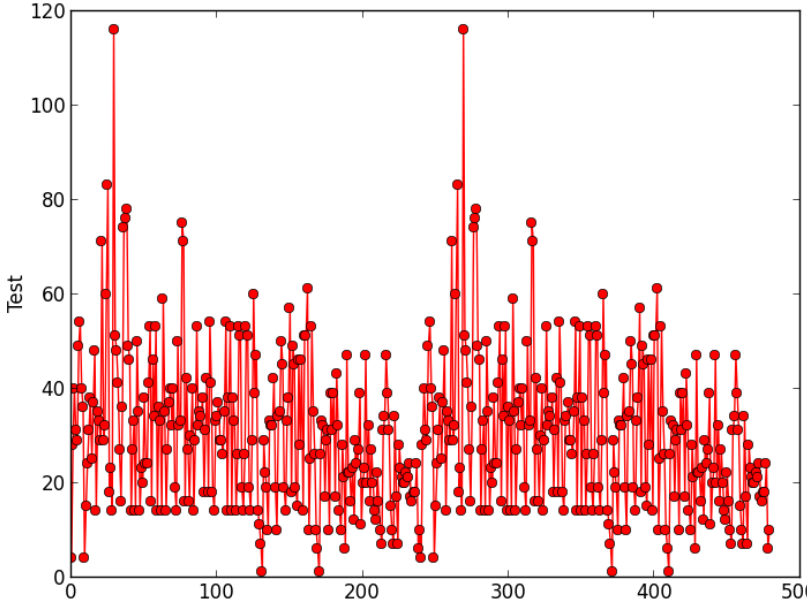
| Offset(h) | 00 | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 0A | 0B | 0C | 0D | 0E | 0F | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 00000000 | 64 | 6A | 61 | 6B | 6C | 66 | 66 | 66 | 66 | 73 | 64 | 61 | 46 | 64 | 0D | 0A | djaklffffsdaFd.. |
| 00000010 | 61 | 66 | 61 | 66 | 6C | 6B | 61 | 6A | 66 | 6C | 6B | 61 | 6E | 66 | 76 | 6D | afaflkajflkanfvm |
| 00000020 | 2E | 2C | 6D | 76 | 6B | 6C | 61 | 6A | 6B | 6C | 61 | 64 | 6A | 66 | 61 | 73 | .,mvklajkladjfas |
| 00000030 | 0D | 0A | 66 | 09 | 32 | 34 | 75 | 74 | 6F | 69 | 67 | 6A | 72 | 6A | 6C | 6B | ..f.24utoigjrjlk |
| 00000040 | 64 | 73 | 61 | 76 | 6B | 6C | 76 | 6C | 6B | 64 | 76 | 6E | 61 | 6B | 6E | 20 | dsavklvlkdvnakn |
| 00000050 | 6E | 66 | 64 | 0D | 0A | 64 | 73 | 63 | 6C | 61 | 6B | 6A | 76 | 6B | 6C | 6A | nfd..dsclakjvklj |
| 00000060 | 76 | 6B | 61 | 68 | 6B | 6A | 4D | 5A | 90 | 00 | 03 | 00 | 00 | 00 | 04 | 00 | vkahkjMZ........ |
| 00000070 | 00 | 00 | FF | FF | 00 | 00 | B8 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 40 | 00 | ..ÿÿ..¸.......@. |
| 00000080 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | ................ |
| 00000090 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | ................ |
| 000000A0 | 00 | 00 | 80 | 00 | 00 | 00 | 0E | 1F | BA | 0E | 00 | B4 | 09 | CD | 21 | B8 | ..€.....º..´.Í!¸ |
| 000000B0 | 01 | 4C | CD | 21 | 54 | 68 | 69 | 73 | 20 | 70 | 72 | 6F | 67 | 72 | 61 | 6D | .LÍ!This program |
| 000000C0 | 20 | 63 | 61 | 6E | 6E | 6F | 74 | 20 | 62 | 65 | 20 | 72 | 75 | 6E | 20 | 69 | cannot be run i |
| 000000D0 | 6E | 20 | 44 | 4F | 53 | 20 | 6D | 6F | 64 | 65 | 2E | 0D | 0D | 0A | 24 | 00 | n DOS mode....$. |
| 000000E0 | 00 | 00 | 00 | 00 | 00 | 00 | 50 | 45 | 00 | 00 | 4C | 01 | 03 | 00 | 5B | 4E | ......PE..L...[N |
| 000000F0 | D3 | 49 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | E0 | 00 | 0F | 03 | 0B | 01 | ÓI........à..... |

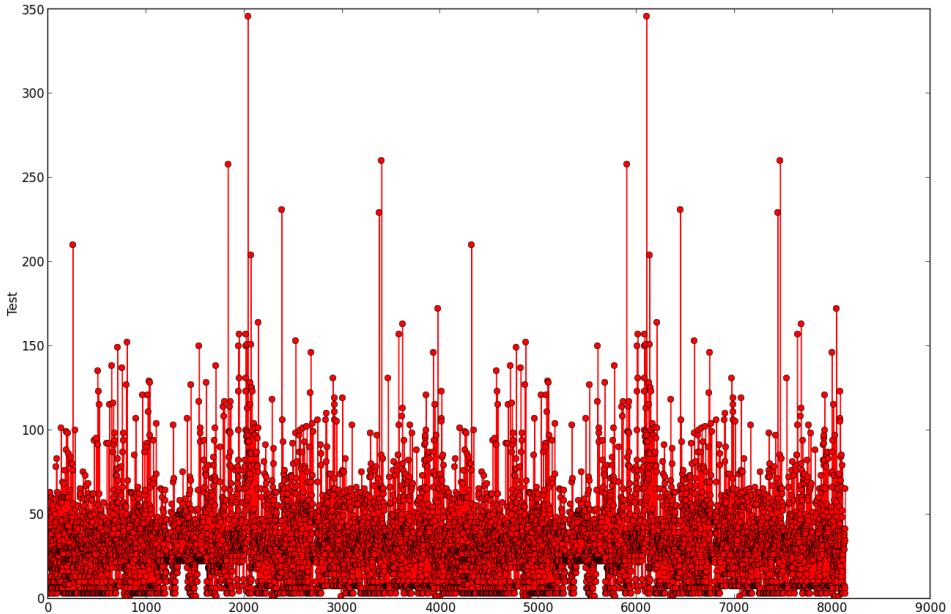
Data Stream with Embedded Portable Executable

File carving with pe-carv.py

```
>dir
01/31/2013  04:42 PM    <DIR>          .
01/31/2013  04:42 PM    <DIR>          ..
01/31/2013  04:08 PM             2,379 pe-carv.py
01/31/2013  10:01 AM           147,558 xxx.bin
               2 File(s)        149,937 bytes
>pe-carv.py xxx.bin
       * exe found at offset 0x66
>dir
01/31/2013  04:42 PM    <DIR>
01/31/2013  04:42 PM    <DIR>          ..
01/31/2013  04:42 PM           147,456 1.exe
01/31/2013  04:08 PM             2,379 pe-carv.py
01/31/2013  10:01 AM           147,558 xxx.bin
```

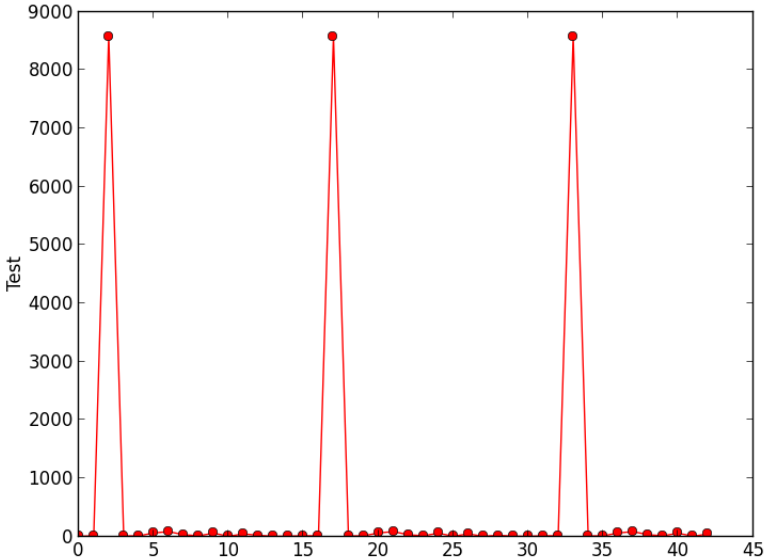
# Exploring file formats- Example

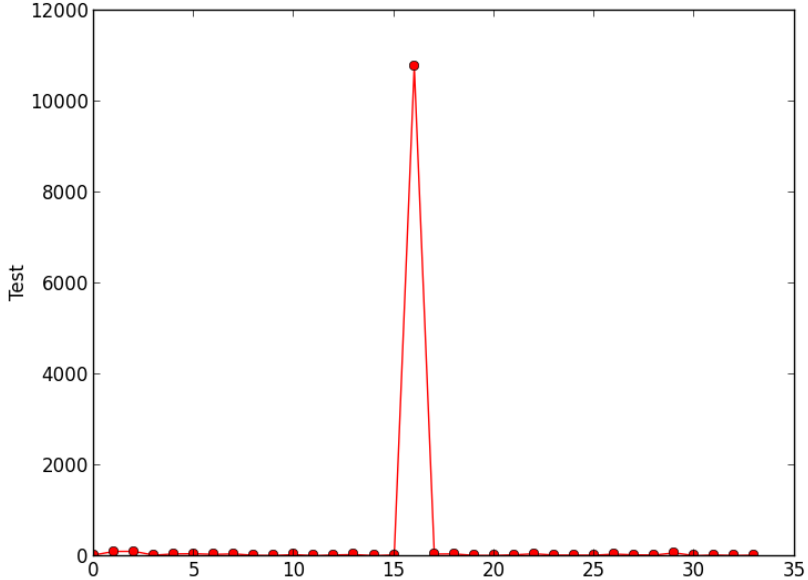



Non-obfuscated JavaScript

# Exploring file formats- Example




Obfuscated JavaScript

# Task -Vulnerability & Exploit Analysis

- Small subset of a very complex area.
- Fuzzing
  - Providing invalid, unexpected or random data as input to see if the data invokes exceptions or crashes
  - Projects
    - Sulley
    - Peach 2
    - python-afl (for fuzzing python code)

# Task -Vulnerability & Exploit Analysis

- Auditing Binaries
  - Scripts in IDAPython commonly used.
  - Simple search for commonly buggy functions
    - Enumerating suspicious function calls (strcpy, fgets, etc)
    - Enumerating file and network input and output
  - Diffing the assembly of patched and vulnerable executables
  - Analyzing data flow and allocation of variables

Links

- Simple Deobfuscation of Code Transformation
- - http://hooked-on-mnemonics.blogspot.com/2012/10/simple-deobfuscation-of-code.html
- Using Z3 to solve a crack me - http://wiremask.eu/hackingweek-2015-reverse-4/
- Deobfuscation: recovering an OLLVM-protected program
- - http://blog.quarkslab.com/deobfuscation-recovering-an-ollvm-protected-program.html
- Breaking Kryptonite's Obfuscation: A Static Analysis Approach Relying on Symbolic Execution - https://doar-e.github.io/blog/2013/09/16/breaking-kryptonites-obfuscation-with-symbolic-execution/
- python-registry Introduction by Willi Ballenthin - https://github.com/williballenthin/python-registry
- Automating DFIR - http://www.hecfblog.com/2015/02/automating-dfir-how-to-series-on.html
- Stuxnet's Footprint in Memory with Volatility 2.0 - http://mnin.blogspot.com/2011/06/examining-stuxnets-footprint-in-memory.html
- Extracting the Powelik's DLL from the Registry - http://sketchymoose.blogspot.com/2015/08/extracting-poweliks-dll-from-registry.html
- ObfStrReplacer - http://hooked-on-mnemonics.blogspot.com/2015/06/obfstrreplacer-extractsubfile-snippets.html
- Cuckoo Sandbox - https://github.com/cuckoobox/cuckoo
- dpkt - https://github.com/jeffsilverm/dpkt_doc
- malcom - https://github.com/tomchop/malcom
- The Very Unofficial Dummies Guide To Scapy - https://theitgeekchronicles.files.wordpress.com/2012/05/scapyguide1.pdf
- Using Python to Fight Cybercrime - https://speakerdeck.com/krmaxwell/using-python-to-fight-cybercrime

Books!

- Hacking Secret Ciphers with Python
- Gray Hat Python: Python Programming for Hackers and Reverse Engineers
- The Art of Memory Forensics: Detecting Malware and Threats in Windows, Linux, and Mac Memory
- Black Hat Python: Python Programming for Hackers and Pentesters
- Python Forensics: A workbench for inventing and sharing digital forensic technology
- The Beginner’s guide to IDAPython :)

# Questions?

*Thanks!*