

Building a Budget-Friendly Malware Analysis Pipeline

Maximizing Efficiency without Breaking the Bank

Joshua Reynolds

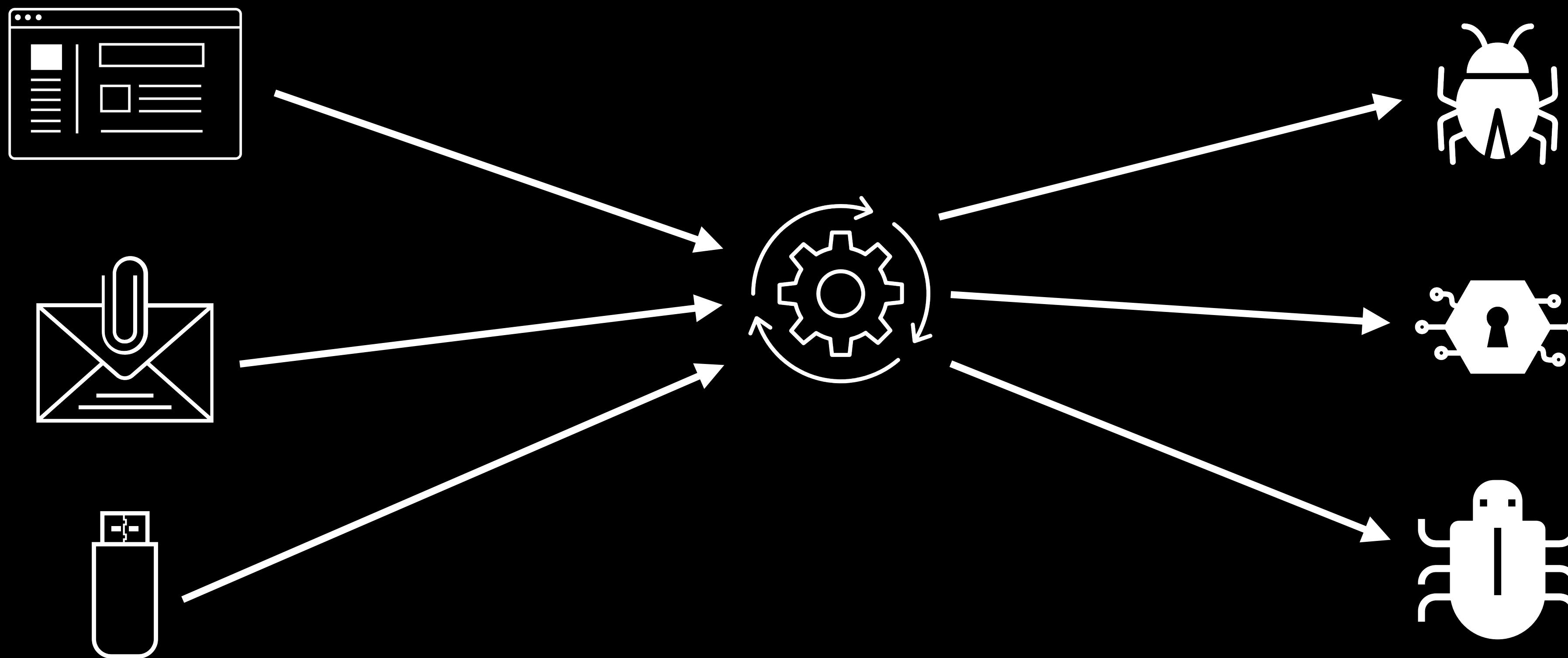
Founder, Invoke RE

- Over ten years of security-related experience working for industry leading companies
- Spoken at RSA, DEF CON and Virus Bulletin on ransomware and malicious document analysis
- Co-developed malware analysis course taught at Southern Alberta Institute of Technology
- @jershmagersh / @InvokeReversing
- info@invokere.com

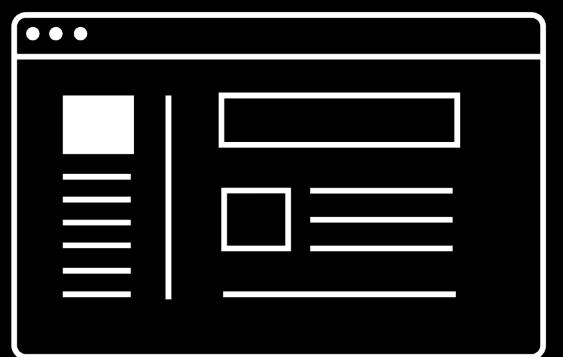


Malware Analysis

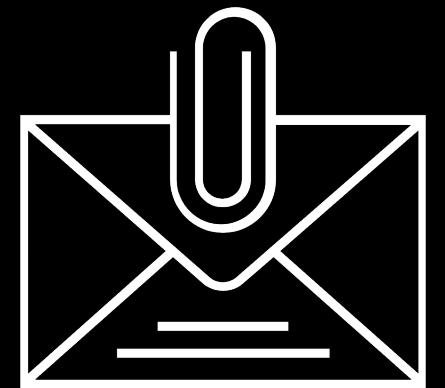
What is Malware Analysis?



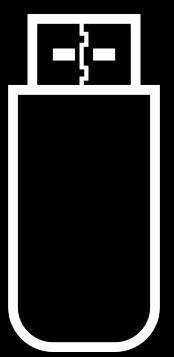
Example Infection Vector Scenarios



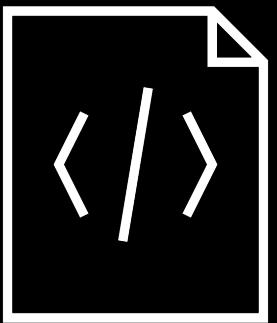
Malvertising leads to download of fake update



Email document to user containing malicious macros

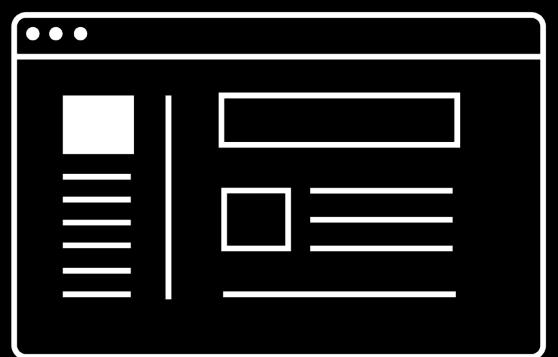


USB emulates keystrokes to execute PS payload

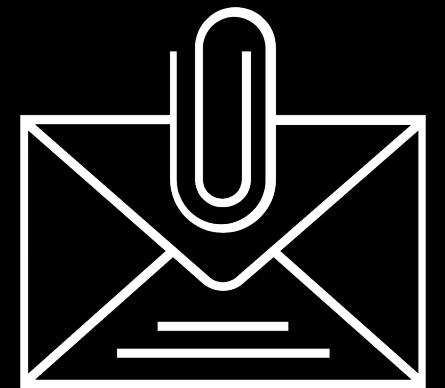


User downloads a ZIP containing a malicious script

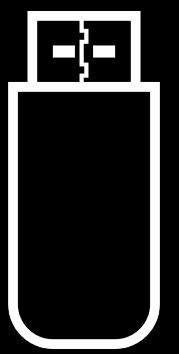
Example Collection Scenarios



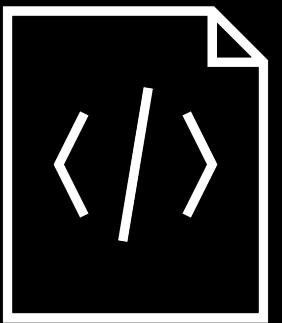
Fake update binary is collected using HTTP proxy



Malicious document download blocked at IPS

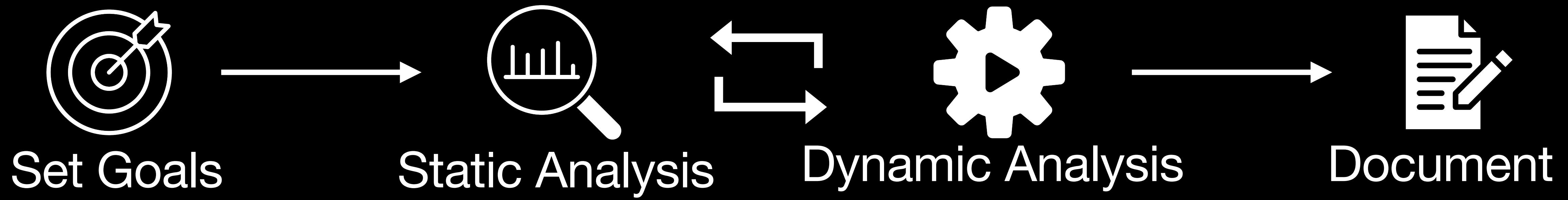


PowerShell launched by USB blocked by EDR

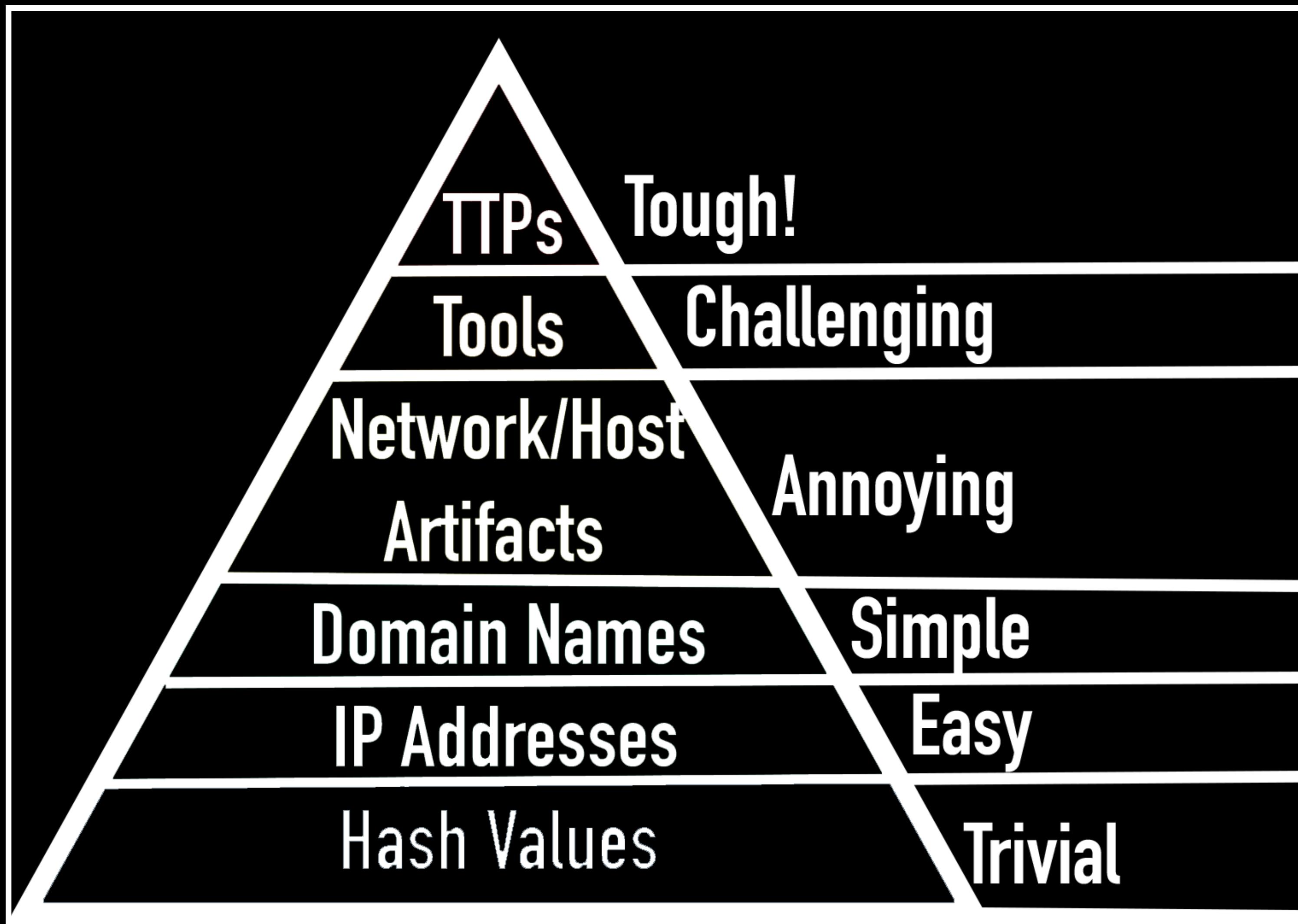


Malicious script download detected by IDS

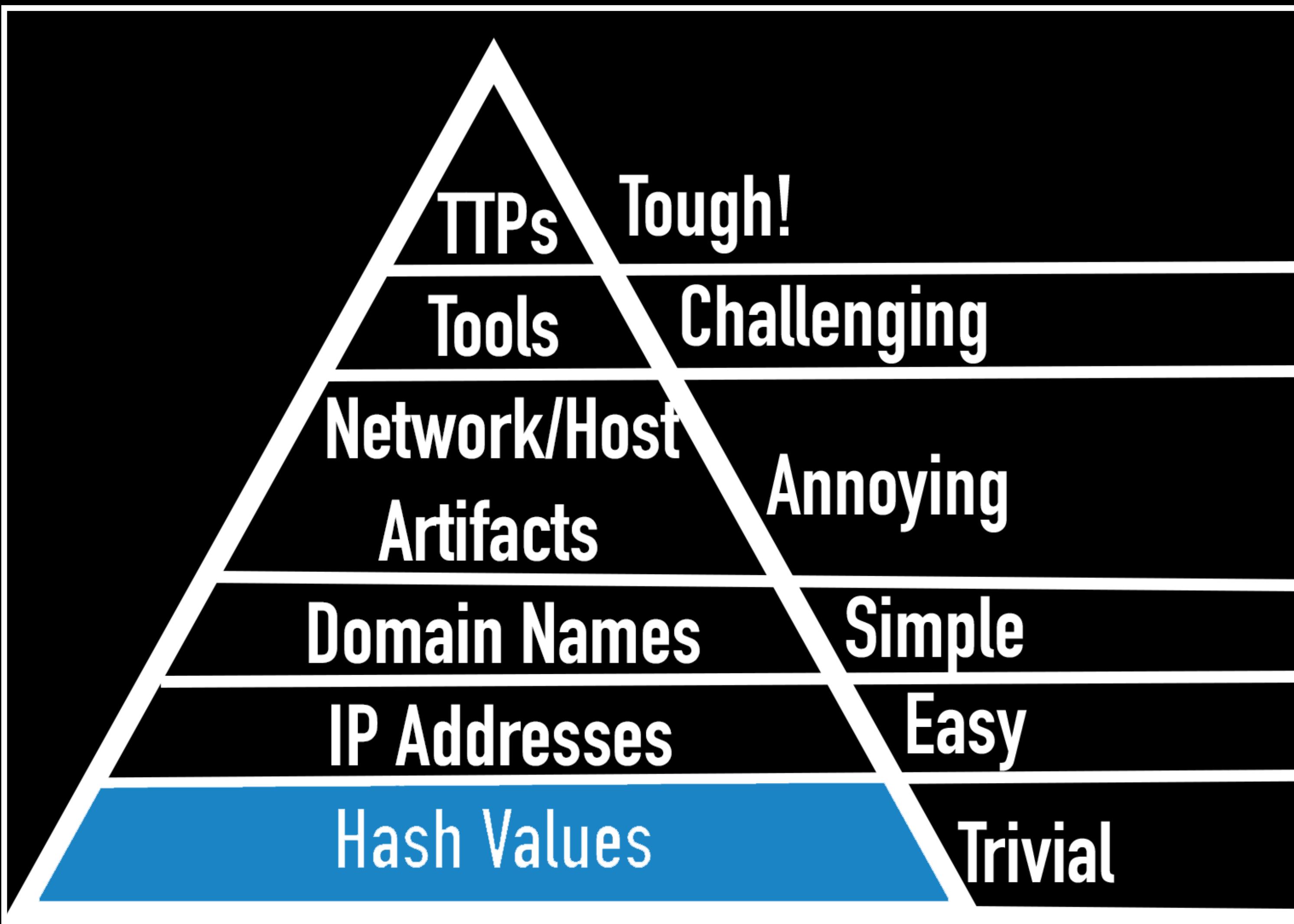
Example Malware Analysis Workflow



The Pyramid of Pain



The Pyramid of Pain



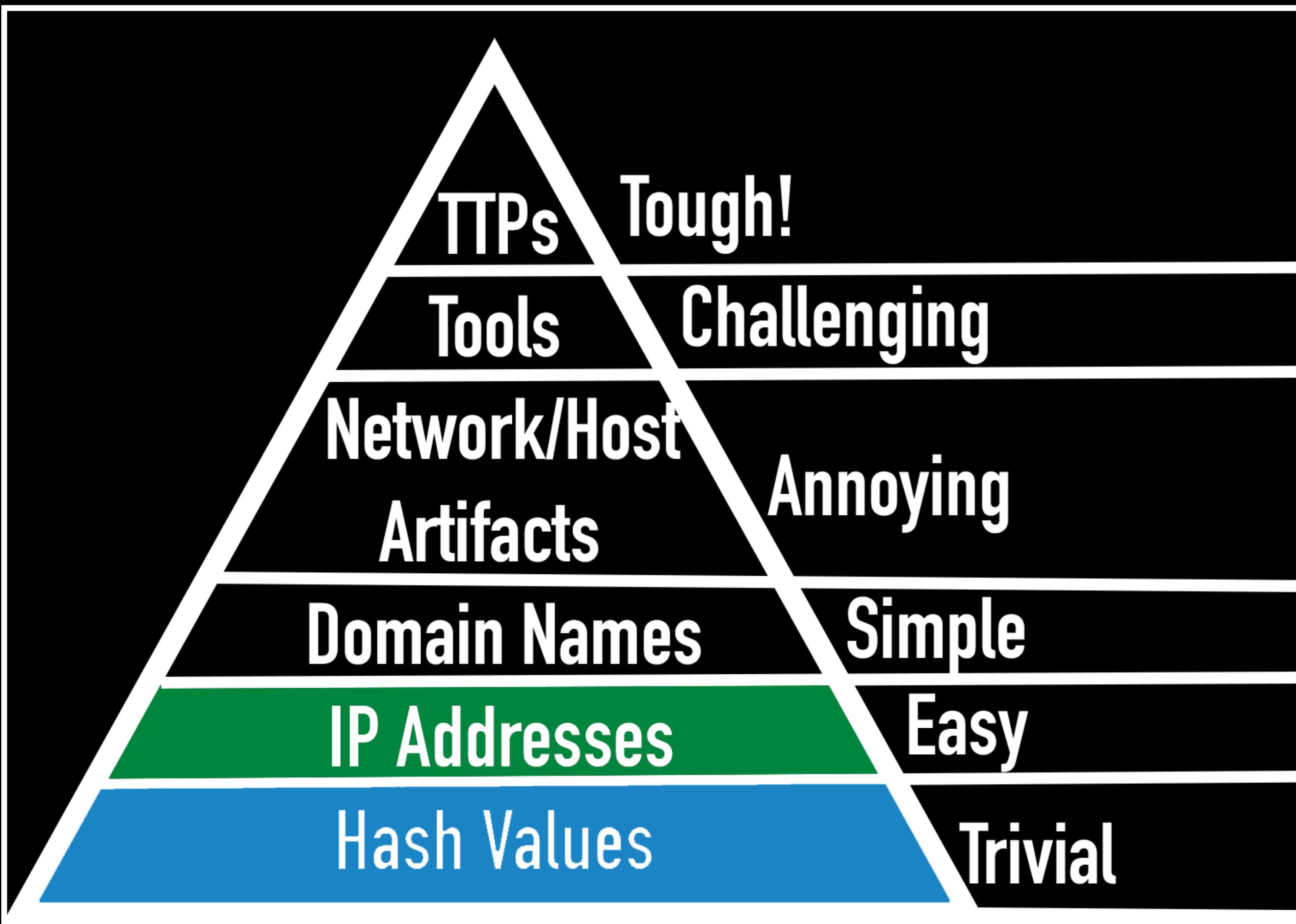
File Hash Example

```
PS C:\windows\system32> Get-FileHash calc.exe  


| Algorithm | Hash                                                             | Path                         |
|-----------|------------------------------------------------------------------|------------------------------|
| -----     | -----                                                            | -----                        |
| SHA256    | 3091E2ABFB55D05D6284B6C4B058B62C8C28AFC1D883B699E9A2B5482EC6FD51 | C:\windows\system32\calc.exe |


```

The Pyramid of Pain



IP Address Example

Capturing from ens33

File Edit View Go Capture Analyze Statistics Telephony Wireless Tools Help

Apply a display filter ... <Ctrl-/>

No.	Time	Source	Destination	Protocol	Length	Info
85	29.994175915	192.168.117.129	194.146.180.40	TCP	66	1590 → 80 [SYN] Seq=0 Win=64240 Len=0 MSS=1460 WS=256 SACK_PERM=1
86	29.994219853	194.146.180.40	192.168.117.129	TCP	66	80 → 1590 [SYN, ACK] Seq=0 Ack=1 Win=64240 Len=0 MSS=1460 SACK_P...
87	29.994672160	192.168.117.129	194.146.180.40	TCP	60	1590 → 80 [ACK] Seq=1 Ack=1 Win=2102272 Len=0
88	35.025341674	192.168.117.129	192.168.117.128	DNS	96	Standard query 0xdba9 A disc801.prod.do.dsp.mp.microsoft.com
89	35.025802169	192.168.117.128	192.168.117.129	DNS	112	Standard query response 0xdba9 A disc801.prod.do.dsp.mp.microsoft...
90	35.027504753	192.168.117.129	192.168.117.128	TCP	66	1591 → 443 [SYN] Seq=0 Win=64240 Len=0 MSS=1460 WS=256 SACK_PERM...

Frame 85: 66 bytes on wire (528 bits), 66 bytes captured (528 bits) on interface ens33, id 0

Ethernet II, Src: VMware_9e:75:df (00:0c:29:9e:75:df), Dst: VMware_30:d6:5c (00:0c:29:30:d6:5c)

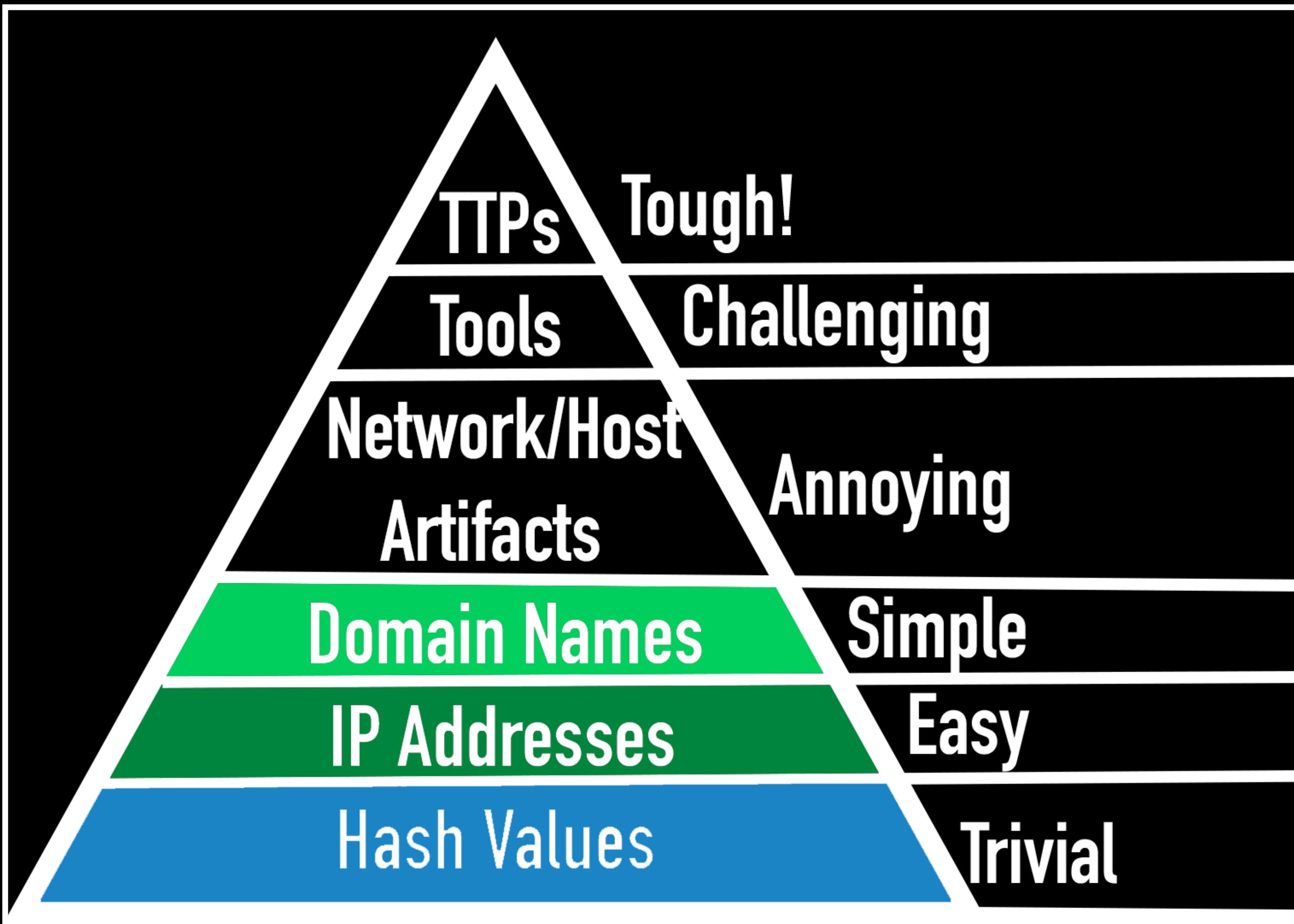
Internet Protocol Version 4, Src: 192.168.117.129, Dst: 194.146.180.40

Transmission Control Protocol, Src Port: 1590, Dst Port: 80, Seq: 0, Len: 0

0000	00 0c 29 30 d6 5c 00 0c 29 9e 75 df 08 00 45 00	...)0.\...) .u.. E.
0010	00 34 e0 8c 40 00 80 06 6d 52 c0 a8 75 81 c2 92	.4..@.... mR.. u...
0020	b4 28 06 36 00 50 6d 45 e3 27 00 00 00 00 80 02	.(.6.PmE ..'.....
0030	fa f0 70 47 00 00 02 04 05 b4 01 03 03 08 01 01	. .pG.....
0040	04 02	..

ens33: <live capture in progress> | Packets: 182 · Displayed: 182 (100.0%) | Profile: Default

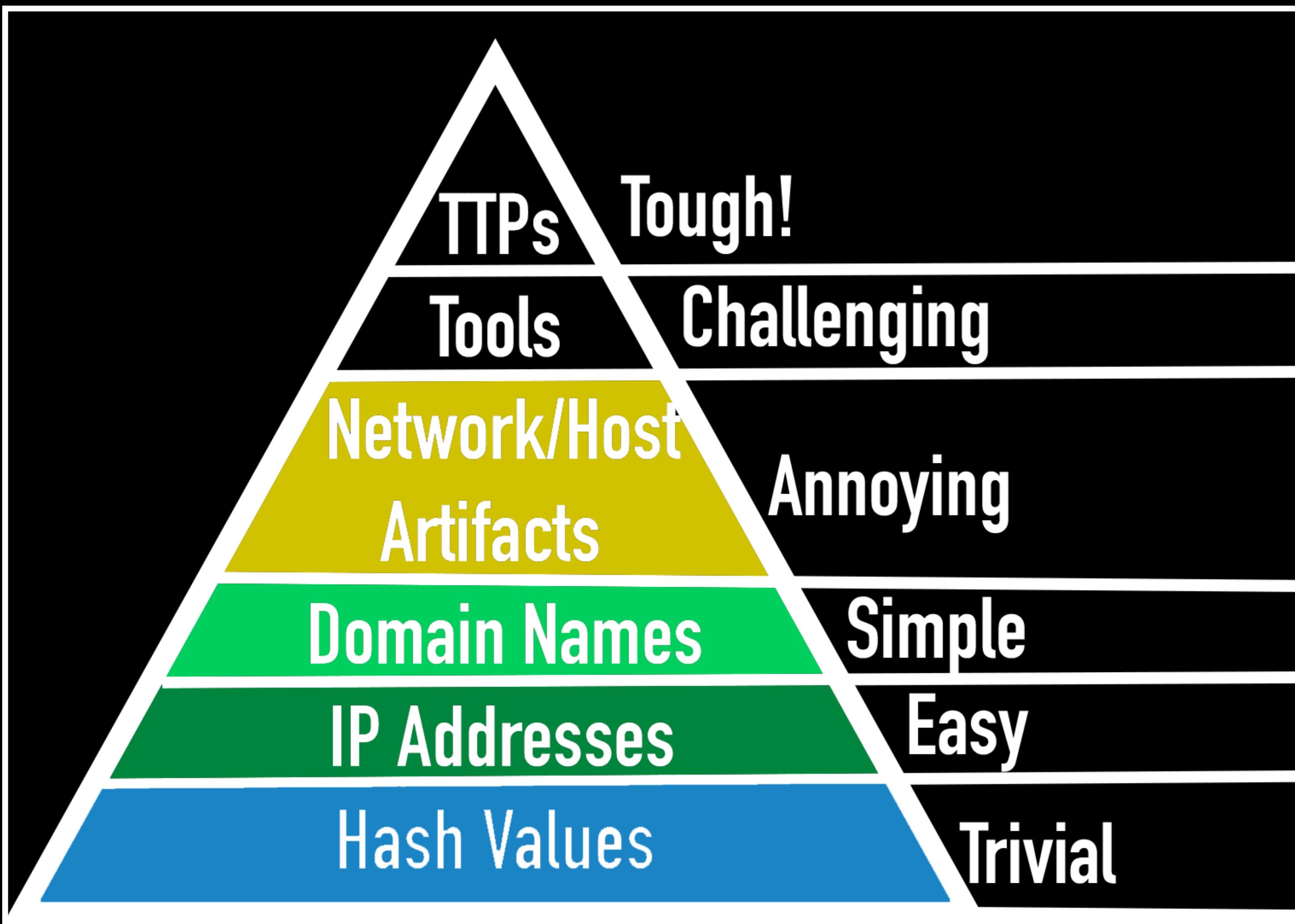
The Pyramid of Pain



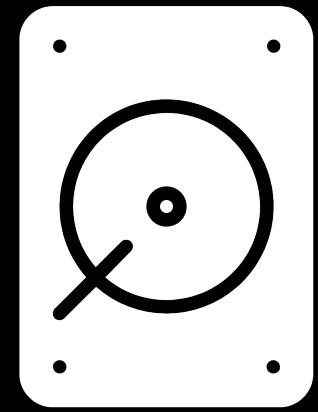
DNS Capture Example

```
remnux@remnux:~$ fakedns 2>&1 | grep -v "microsoft.com ->"  
fakedns[INFO]: dom.query. 60 IN A 192.168.117.128  
fakedns[INFO]: Response: plumberspro.us -> 192.168.117.128  
fakedns[INFO]: Response: ntp.ubuntu.com -> 192.168.117.128  
fakedns[INFO]: Response: ntp.ubuntu.com -> 192.168.117.128  
fakedns[INFO]: Response: godstar.hopto.org -> 192.168.117.128  
fakedns[INFO]: Response: grigori.ddns.net -> 192.168.117.128  
fakedns[INFO]: Response: checkip.dyndns.org -> 192.168.117.128  
fakedns[INFO]: Response: plumberspro.us -> 192.168.117.128  
fakedns[INFO]: Response: godstar.hopto.org -> 192.168.117.128  
fakedns[INFO]: Response: grigori.ddns.net -> 192.168.117.128
```

The Pyramid of Pain



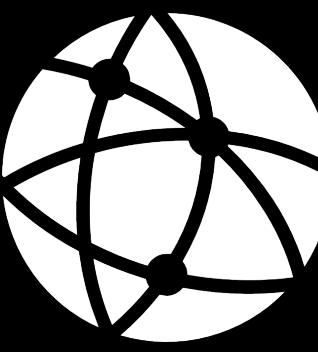
What are host and network artifacts?



Files residing on the filesystem



Registry hives, keys and values

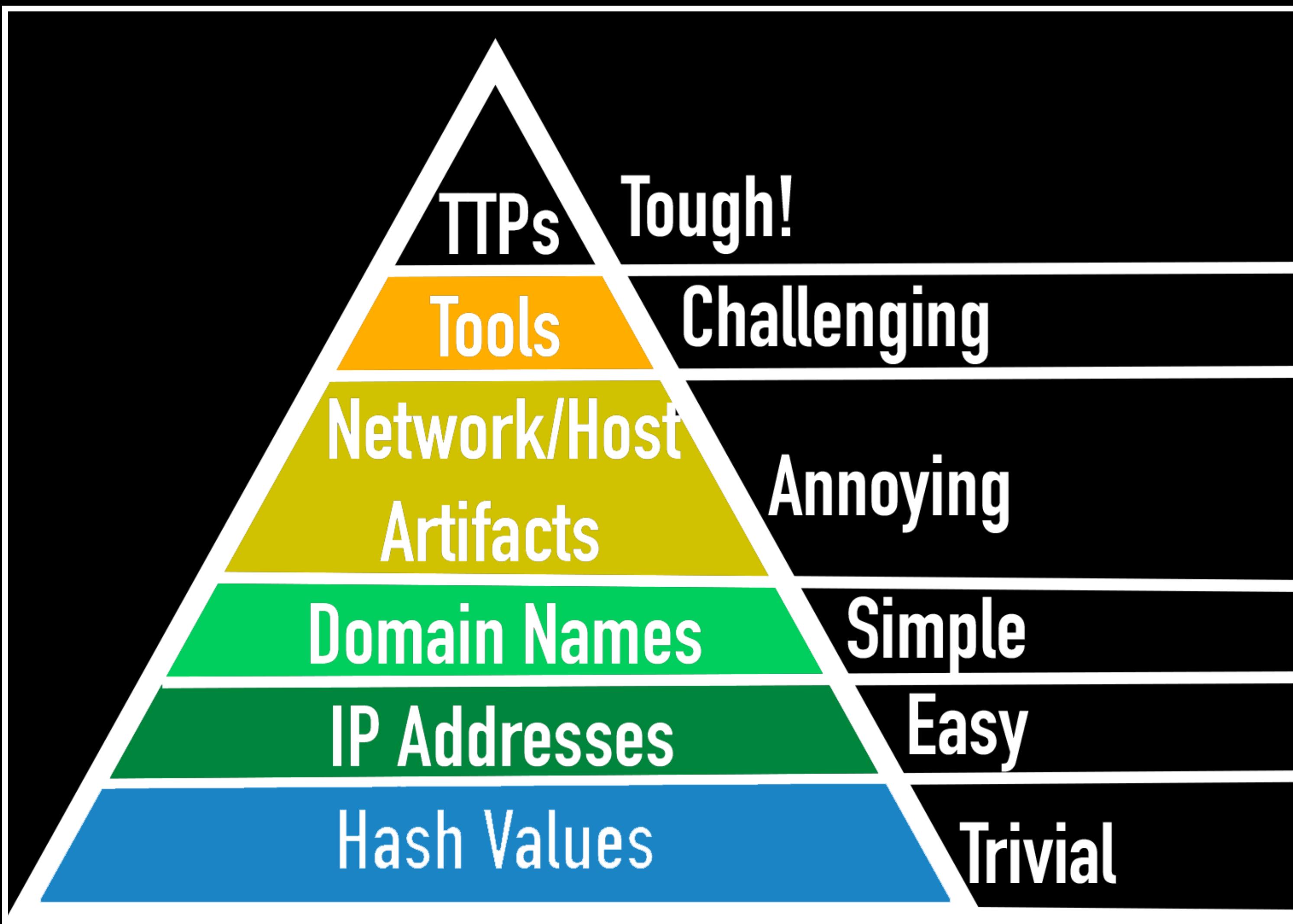


Network characteristics

Host Artifact Example

```
.text:012D7F29    push  eax      , _lpwdDisposition
• .text:012D7F2A    lea   eax, [ebp+phkResult]
• .text:012D7F2D    push  eax      ; phkResult
• .text:012D7F2E    push  0       ; lpSecurityAttributes
• .text:012D7F30    push  20006h   ; samDesired
• .text:012D7F35    push  0       ; dwOptions
• .text:012D7F37    push  0       ; lpClass
• .text:012D7F39    push  0       ; Reserved
• .text:012D7F3B    push  offset aSoftwareMicros_0 ; "SOFTWARE\\Microsoft\\Windows\\CurrentVe"...
• .text:012D7F40    push  HKEY_LOCAL_MACHINE ; hKey
• .text:012D7F45    call  ds:RegCreateKeyExA ; CHAR aSoftwareMicros_0[]
• .text:012D7F4B    test  eax, eax
• .text:012D7F4D    jnz   short loc_12D7FC6
• .text:012D7F4F    mov   ecx, esi
• .text:012D7F51    lea   edx, [ecx+1]
.text:012D7F54
• .text:012D7F54 loc_12D7F54:           ; CODE XREF: sub_12D7EF0+69↓j
• .text:012D7F54    mov   al, [ecx]
• .text:012D7F56    inc   ecx
• .text:012D7F57    test  al, al
• .text:012D7F59    jnz   short loc_12D7F54
• .text:012D7F5B    sub   ecx, edx
• .text:012D7F5D    lea   eax, [ecx+1]
• .text:012D7F60    push  eax      ; cbData
• .text:012D7F61    push  esi      ; lpData
• .text:012D7F62    mov   esi, ds:RegSetValueExA
• .text:012D7F68    push  1       ; dwType
• .text:012D7F6A    push  0       ; Reserved
• .text:012D7F6C    push  offset aSm1    ; "SM_1"
• .text:012D7F71    push  [ebp+phkResult] ; hKey
• .text:012D7F74    call  esi ; RegSetValueExA
• .text:012D7F76    mov   ecx, ebx
```

The Pyramid of Pain



Tool Identification Example

The screenshot shows the PEStudio interface with the following details:

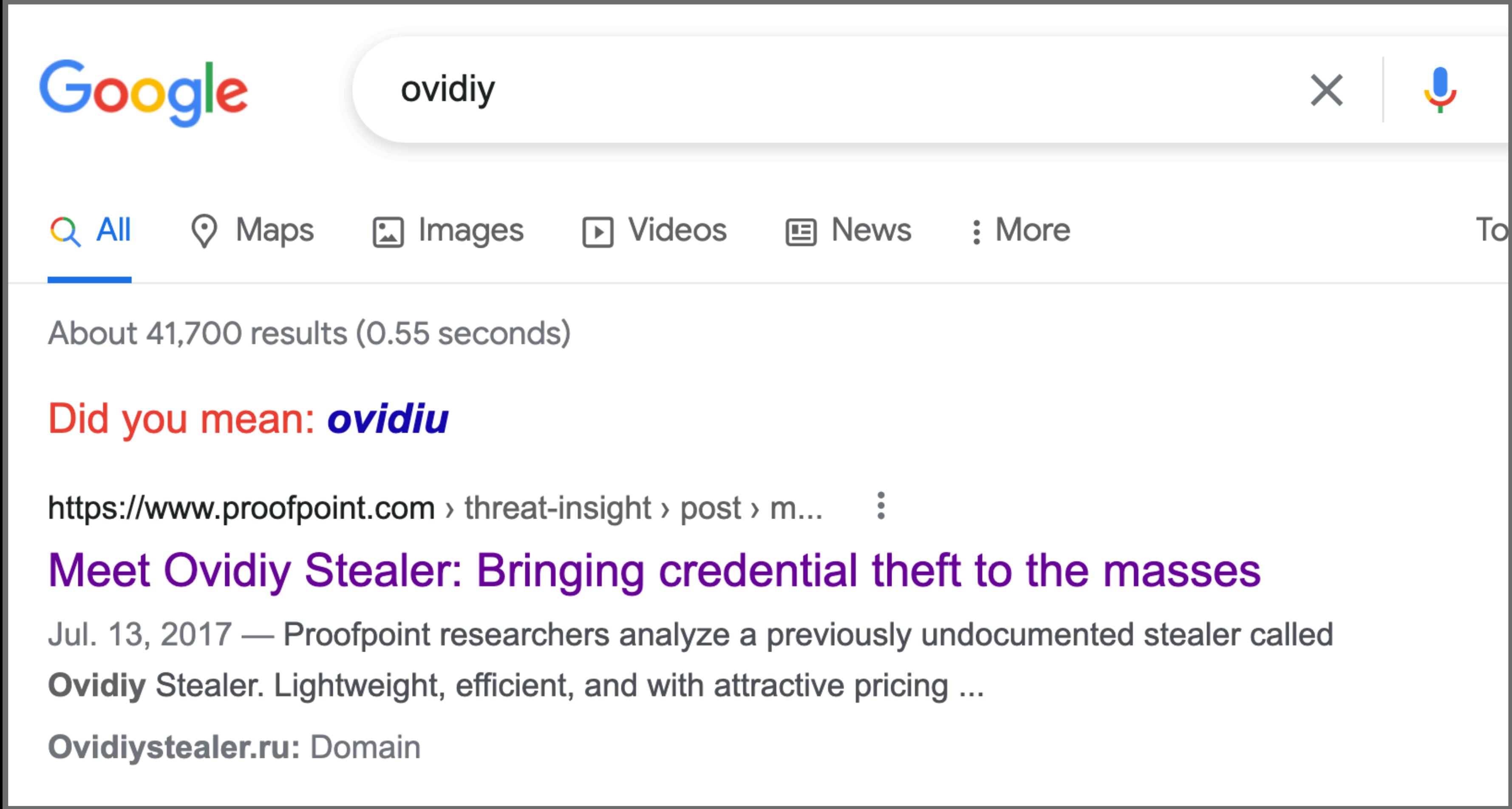
File Path: c:\users\ieuser\Desktop\c16408967de0ca4d3a1d28530453e1c395a5166b469893f14c47fc6683033cb3\c1640896

Tool Identification: The file is identified as Ovidiy.exe.

Properties Table:

property	value
md5	1363487F0DDD624330257C53550F5409
sha1	F88422158E9613FCEBF79D8A8F2A8D8C1EF60335
sha256	C01ED23EE0784BD4E6A50218788FE6FE8DB3429DFB42BE747DF0ECB029AF2C1B
file-type	executable
date	empty
language	neutral
code-page	Unicode UTF-16, little endian
FileDescription	Ovidiy Ovidiy
FileVersion	1.0.0.5
InternalName	Ovidiy.exe
LegalCopyright	n/a
LegalTrademarks	Ovidiy Ovidiy
OriginalFilename	Ovidiy.exe
ProductName	Ovidiy Ovidiy
ProductVersion	1.0.0.5
Assembly Version	1.0.0.5

Tool Identification Example



A screenshot of a Google search results page. The search bar at the top contains the query "ovidiy". Below the search bar, there are navigation links for "All", "Maps", "Images", "Videos", "News", and "More". A message "About 41,700 results (0.55 seconds)" is displayed. A red link labeled "Did you mean: ovidiu" is present. The first search result is a link to a Proofpoint blog post titled "Meet Ovidiy Stealer: Bringing credential theft to the masses", dated Jul. 13, 2017. The URL of the result is https://www.proofpoint.com/threat-insight/post/m... .

ovidiy

All Maps Images Videos News More

About 41,700 results (0.55 seconds)

Did you mean: **ovidiu**

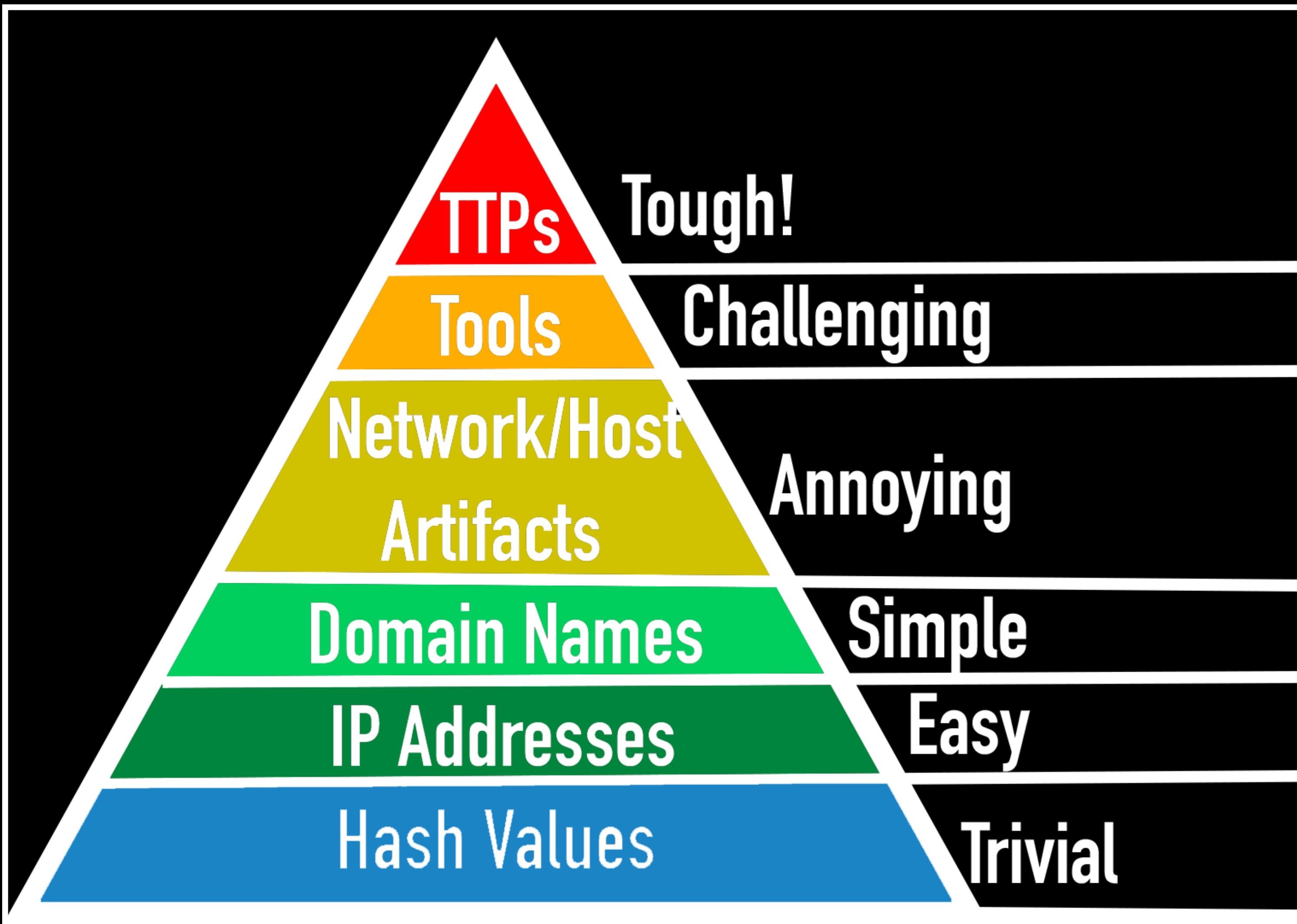
<https://www.proofpoint.com/threat-insight/post/m...>

Meet Ovidiy Stealer: Bringing credential theft to the masses

Jul. 13, 2017 — Proofpoint researchers analyze a previously undocumented stealer called Ovidiy Stealer. Lightweight, efficient, and with attractive pricing ...

Ovidiystealer.ru: Domain

The Pyramid of Pain



ATT&CK Framework

MITRE | ATT&CK®

Matrices Tactics ▾ Techniques ▾ Mitigations ▾ Groups Software Resources ▾ Blog ↗

Contribute Search

MATRICES

- Enterprise
- PRE
- Windows
- macOS
- Linux
- Cloud
- Network
- Containers
- Mobile
- ICS ↗

Home > Matrices > Windows

Windows Matrix

Below are the tactics and techniques representing the MITRE ATT&CK® Matrix for Enterprise. The Matrix contains information for the Windows platform.

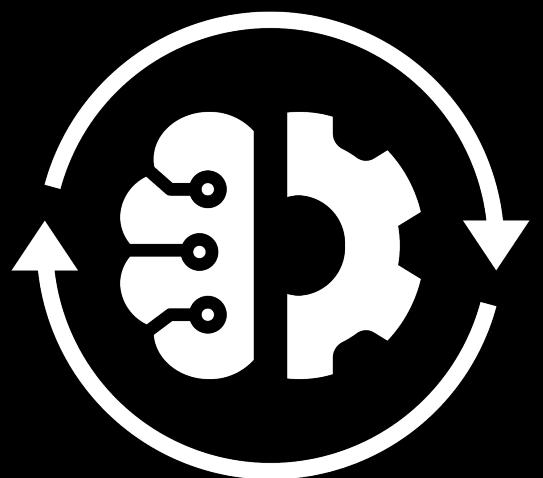
View on the ATT&CK® Navigator ↗
Version Permalink

layout: side ▾
show sub-techniques hide sub-techniques
help

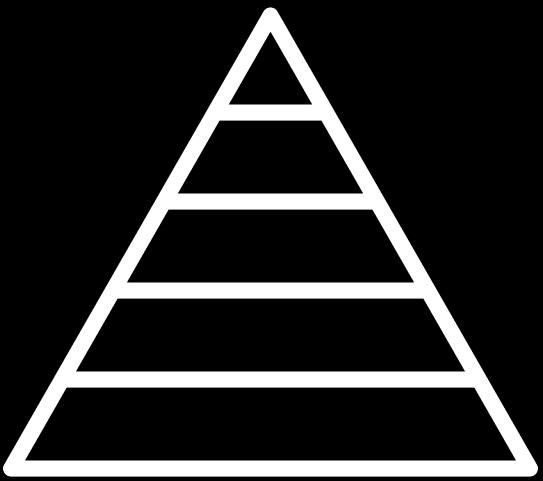
Initial Access	Execution	Persistence	Privilege Escalation	Defense Evasion	Credential Access
9 techniques	10 techniques	18 techniques	13 techniques	32 techniques	14 techniques
Drive-by Compromise Exploit Public-Facing Application External...	Command and Scripting Interpreter (5) Exploitation for Client Execution Inter Process...	Account Manipulation (1) BITS Jobs Boot or Logon Autostart Execution...	Abuse Elevation Control Mechanism (1) Access Token Manipulation (5) BITS Jobs	Abuse Elevation Control Mechanism (1) Access Token Manipulation (5) BITS Jobs	Brute Force (4) Credentials from Password Stores (3) Exploitation...

Automating Malware Analysis

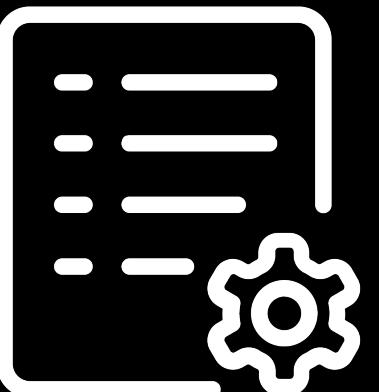
Automating Malware Analysis



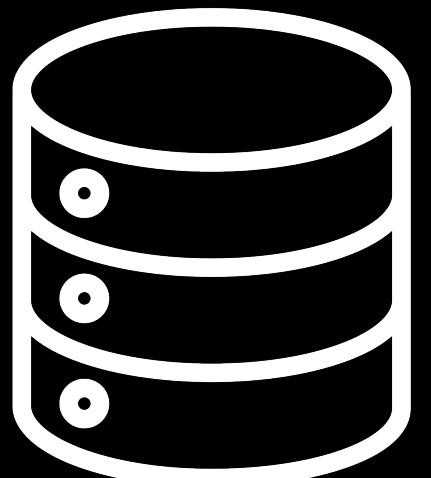
Goal of performing analysis automatically



Extract all levels of pyramid of pain



Extract malware configurations



Store all information for future analysis

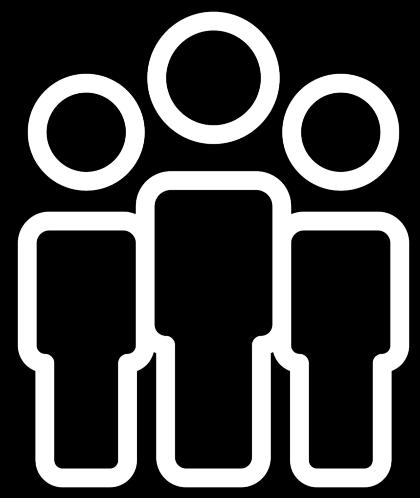
Considerations



Available budget - can someone do this for me?



Privacy - can we upload to cloud services?



Talent - Do we have in-house talent?

Assemblyline

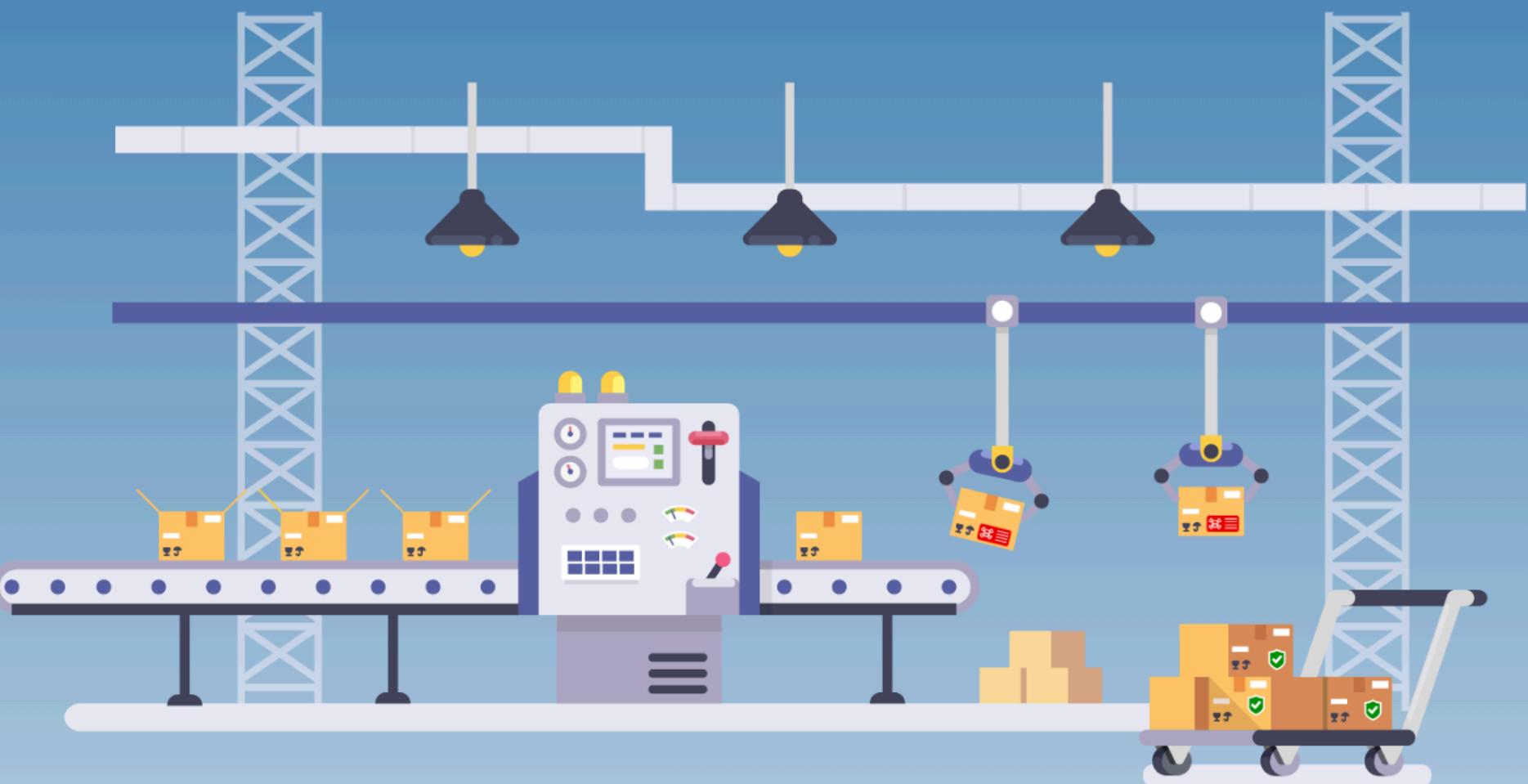
Assemblyline 4

A scalable file triage and malware analysis system integrating the cyber security community's best tools.

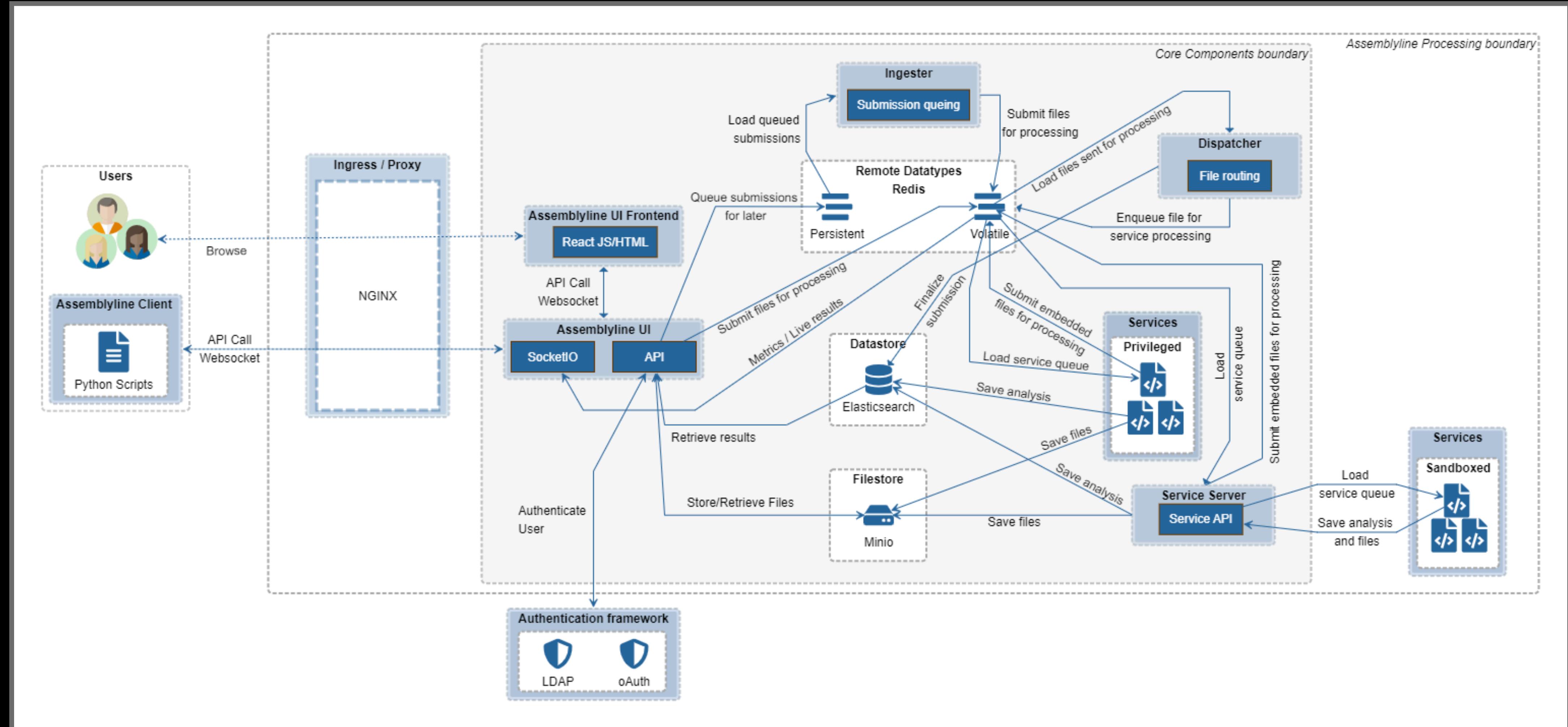
[How it works](#)

[View source code](#) 

[Join our community](#) 

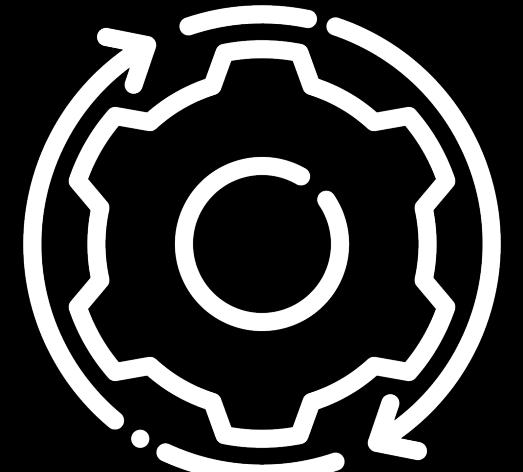


Assemblyline

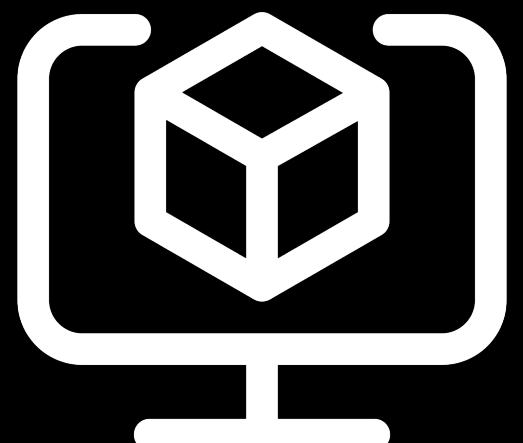


Source: https://cybercentrecanada.github.io/assemblyline4_docs/overview/architecture/

What is a sandbox?



Performs dynamic analysis of samples



Spins up a virtual machine and executes sample



Monitors activity during sample execution



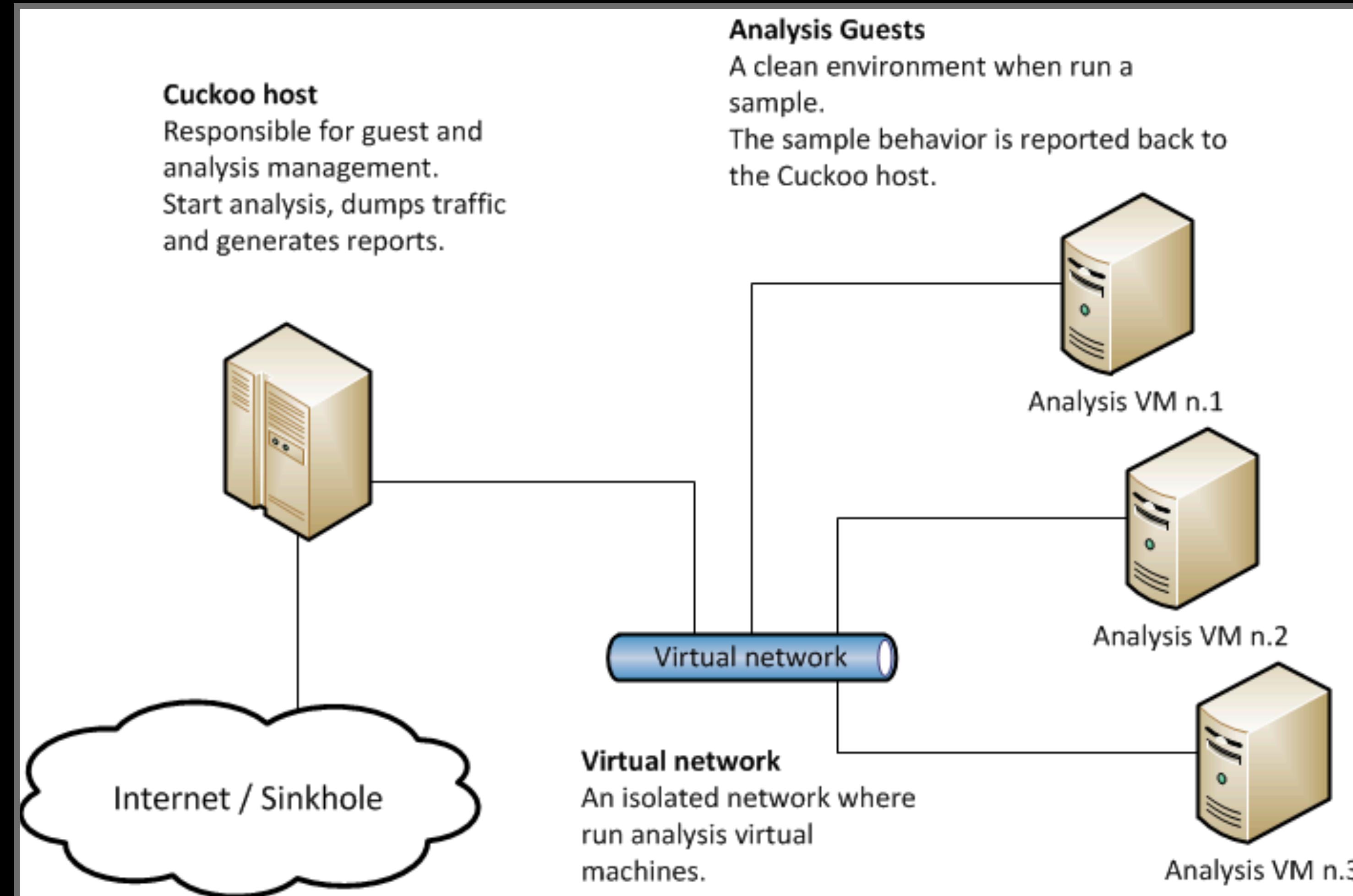
Uploads sandbox outputs to server for analysis

CapeV2

The screenshot shows the CapeV2 web application interface. At the top, there is a navigation bar with links for Dashboard, Recent, Pending, Search, API, Submit, Statistics, User, Docs, and Changelog. A search bar with a placeholder "Search term as regex" and a "Search" button is also present. Below the navigation bar, there are three tabs: Files (selected), Static, and PCAPs. A pagination bar with buttons for 1 through 10 and an "Older →" link is shown. The main area is titled "Recent Files" and contains a table with the following data:

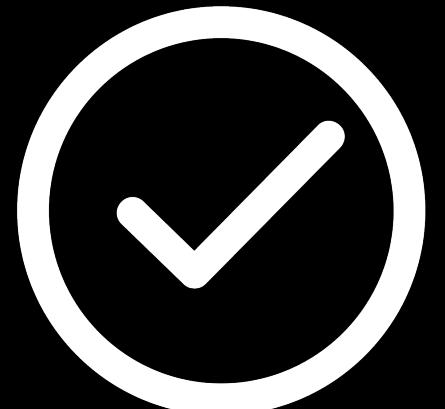
ID	Timestamp	Machine	Package	Filename	MD5	Detection	s	SuriAlert	VT	Status
460729	2023-11-27 15:25:07 (added on)	win7_4	exe	83fbaecfae4f7d45a950.exe	83fbaecfae4f7d45a950cf1d 8d8f604b	0	-	running		
460728	2023-11-27 15:25:07 (added on)	win7x64_6	exe	691533104da0cbd088b9.exe	691533104da0cbd088b947ab 823f77de	0	-	running		
460727	2023-11-27 15:20:33 (added on)	win7_2	exe	78c8ff3737983dee0591.exe	78c8ff3737983dee0591b47a 56f4784b	0	-	running		
460726	2023-11-27 15:20:32 (added on)	win7_3	exe	7e46d4436b93d1420eea.exe	7e46d4436b93d1420eea009 3aea2134	0	-	running		
460725	2023-11-27 15:20:32 (added on)	win7_4	exe	b3ff602ab07ea721e0ed.exe	b3ff602ab07ea721e0ed2df2 18c2566d	0	-	processing		

CapeV2

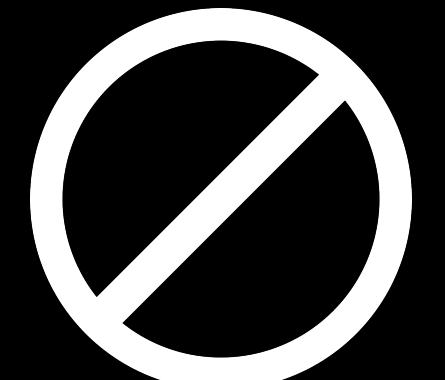


Source: <https://capev2.readthedocs.io/en/latest/introduction/what.html#architecture>

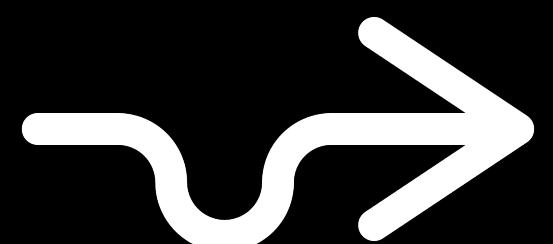
Limitations



Automation can result in false positives



Automation can result in false negatives

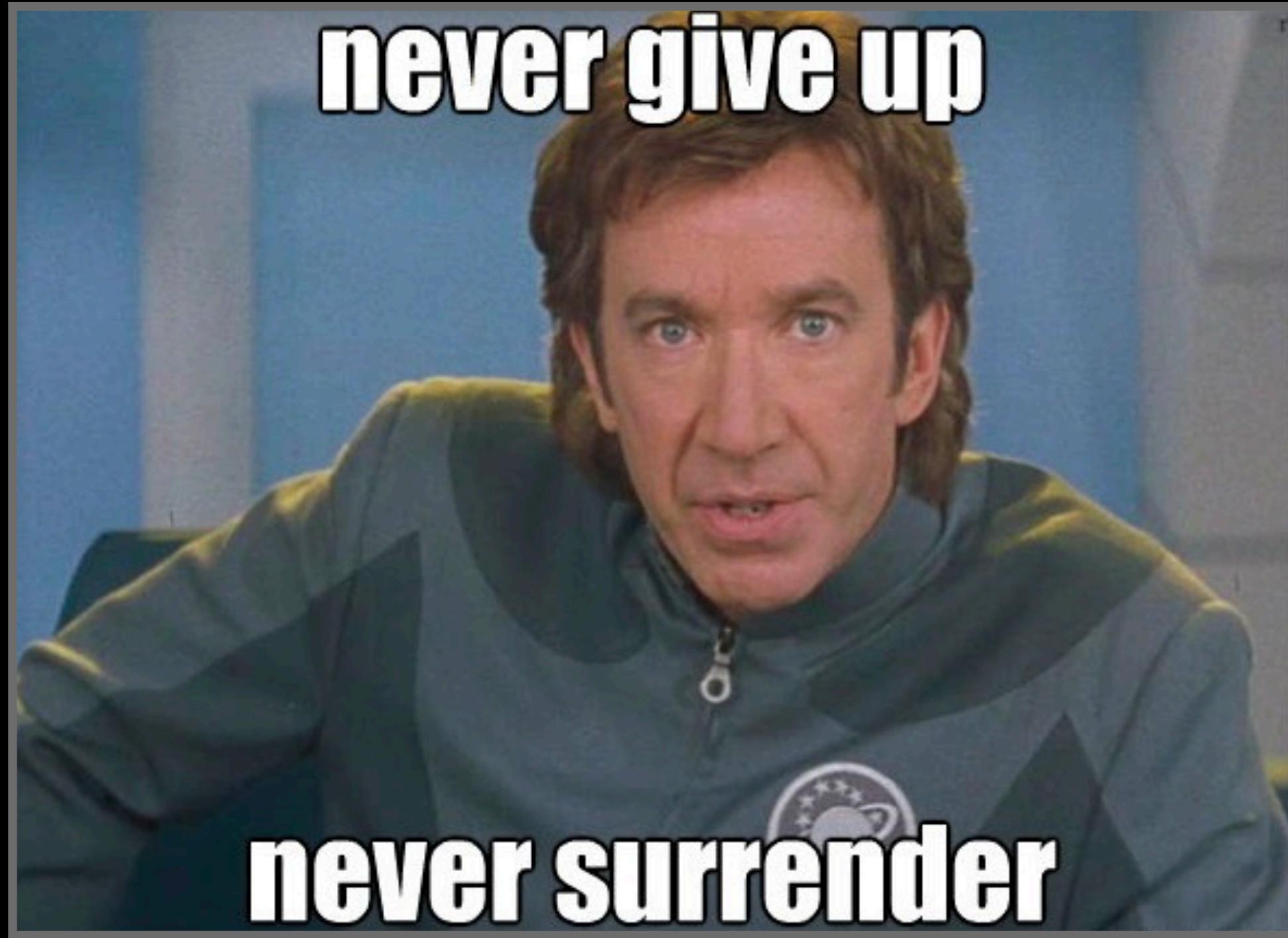


Automation can be evaded

Demo

Further Reading & References

- <https://detect-respond.blogspot.com/2013/03/the-pyramid-of-pain.html>
- <https://detect-respond.blogspot.com/2022/04/stop-using-hashes-for-detection-and.html>
- https://cybercentrecanada.github.io/assemblyline4_docs/
- <https://github.com/kevoreilly/CAPEv2>
- <https://capesandbox.com/analysis/>



Questions?

Contact

@jershmagersh / @InvokeReversing
info@invokere.com