

# BareCloud: Bare-metal Analysis-based Evasive Malware Detection

*Dhilung Kirat, Giovanni Vigna, Christopher Kruegel*

UC Santa Barbara



USENIX Security 2014

San Diego, CA



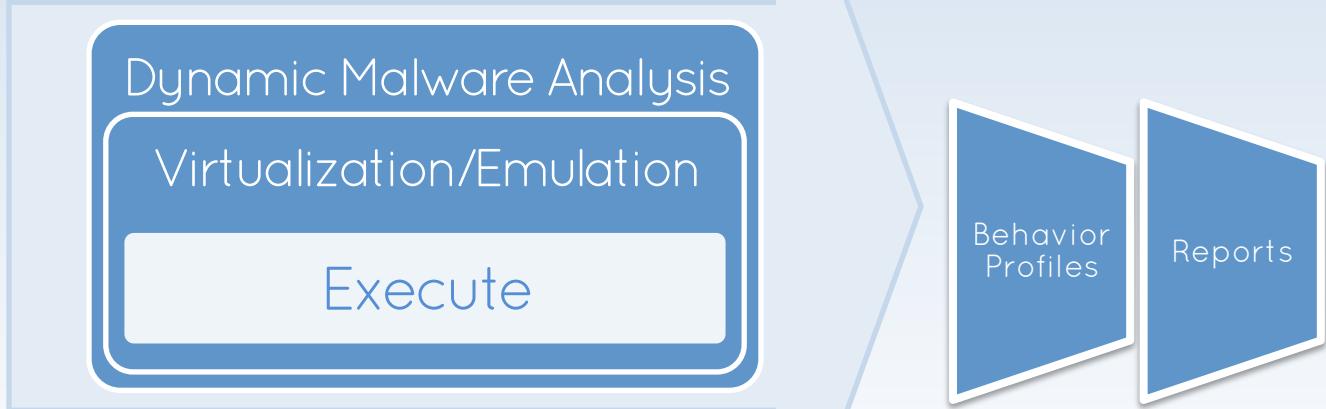
Dynamic Malware Analysis

Execute

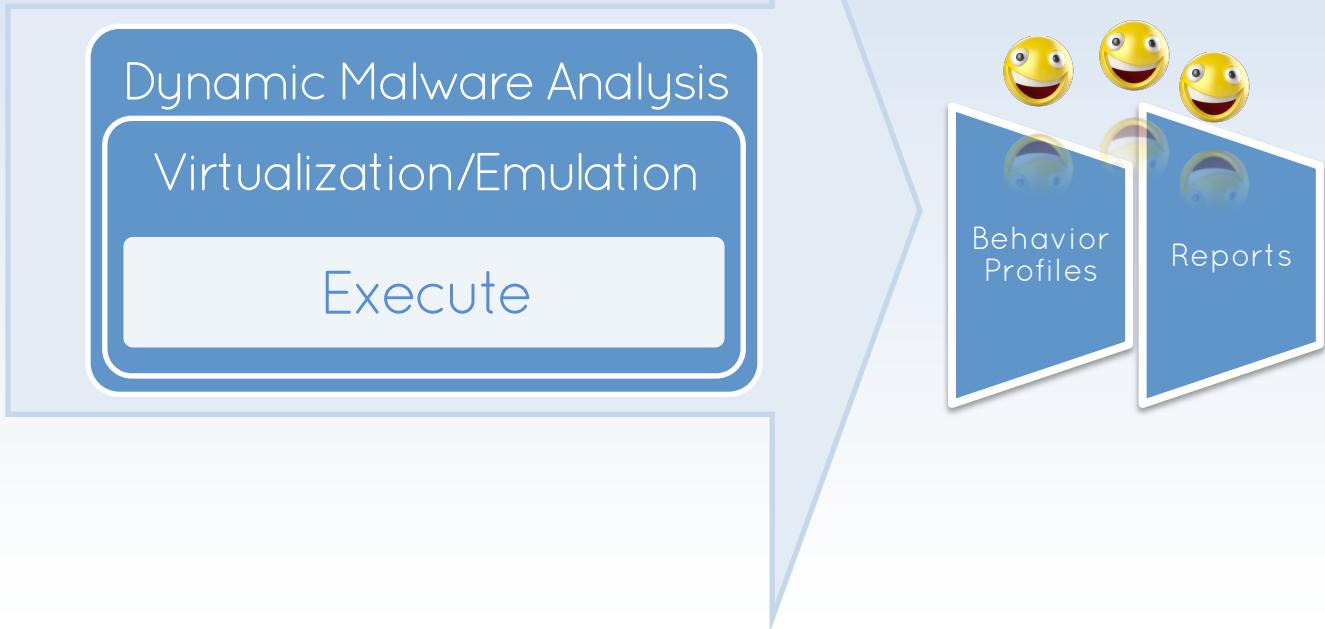


Behavior  
Profiles

Reports



# Evasive Malware



# Evasive Malware

Dynamic Malware Analysis

Virtualization/Emulation

Execute

# Detect Analysis Environment

- Disk
  - HKLM\Hardware\DeviceMap\Scsi
  - HKLM\System\CurrentControlSet\Services\Disk\Enum
- Bios
  - HKLM\Hardware\Description\System\SystemBiosVersion
- Keyboard/Mouse
  - Presence of mouse, keyboard layout
- User
  - Username, Windows Product ID
  - Active user

# Detect Analysis Environment

- CPU
    - SIDT instruction
    - CPU Emulation bug (including MMX instruction set)
  - Vulnerability
    - CVE-2012-3221 VirtualBox
  - Timing attack
    - The virtualization and emulation systems add some level of overhead
- ```
int swallow_redpill () {
    unsigned char m[2+4], rpill[] = "\x0f\x01\x0d\x00\x00\x00\x00\xc3";
    *((unsigned*)&rpill[3]) = (unsigned)m;
    ((void(*)())&rpill)();
    return (m[5]>0xd0) ? 1 : 0;
}
```
- ```
int crash() {
    asm (
        "int $0x8;"           // output: none
        : // input: none
        : "%eax", "%ebx", "%ecx", "%edx"    // clobbered register
        );
    return(0);
}
```

# Fully Undetectable (FUD)

The image displays four overlapping windows representing different FUD (Fully Undetectable) tools:

- Blackout AIO: Highly** (Left Window): Shows a list of checkboxes for various anti-detection features:
  - Anti-Sandboxie
  - Anti-Virtual PC
  - Anti-Norman Sandbox** (highlighted in green)
  - Anti-IDB Debugger
  - Anti-CWSandbox
  - Anti-Norman Sandbox
  - Anti-VirtualBox
  - Anti-Sunbelt Sandbox
  - Sleep Sec. (0)
  - Exceptions № (100)
  - Get All Privileges
  - Password protect
  - Execute w/ Command Line
  - Clone File Properties
- fEaRz Crypter 2.2.0 - by fEaRz** (Center Window): Shows a profile picture of Albert Einstein and a list of checkboxes:
  - Anti-Sandbox(s) (Username/Sleep)
  - Anti-Sandbox(s) (GetModuleHandle)
  - Anti-Sandbox(s) (ProcessEntry)
  - Anti-Sandbox(s) (Cd-Key)
  - Anti-Sandbox(s) (IDT base address)
  - Anti-Vmware
  - Detect IsDebuggerPresent
  - Detect Soft-ICE
  - Detect FileMon/RegMon
- Aegis Crypter 3.3** (Top Right Window): Shows a list of checkboxes for various evasion techniques:
  - Anti-Virtual Machine(WM | VPC | VBOX)
  - Anti-SandBox(Sandboxie | More)
  - Bypass UAC(Vista | Win7)
  - Disable Firewall
  - Hide Directory
  - Hide File
  - Add Junk Code
  - Add Startup
  - Inject Default browser
  - Disable EOF
- CRYPTONITE** (Bottom Right Window): Shows a list of checkboxes for anti-methods:
  - Anti VMWare
  - Anti Virtual Pc
  - Anti Anubis Sandbox
  - Anti OllyDebug
  - Anti VirtualBox
  - Anti Wireshark
  - Anti ThreatExpert
  - Anti Emulators
  - Anti Sunbelt Sandbox
  - Is Debugger Present
  - Anti Procmon
  - Anti Regmon
  - Anti Filemon
  - Anti Joebox
  - Anti Sandboxie
  - Anti CWSandbox
  - Anti Norman Sandbox
  - Anti Kaspersky

At the bottom of the image, there is a footer bar with the following text and icons:

Terminates the worm if it found in any of the selected environment's

ThreatExpert WireShark Sandboxie Anubis

Website Block's 1 Block's 2 Block's 3

Icon Support Select Icon

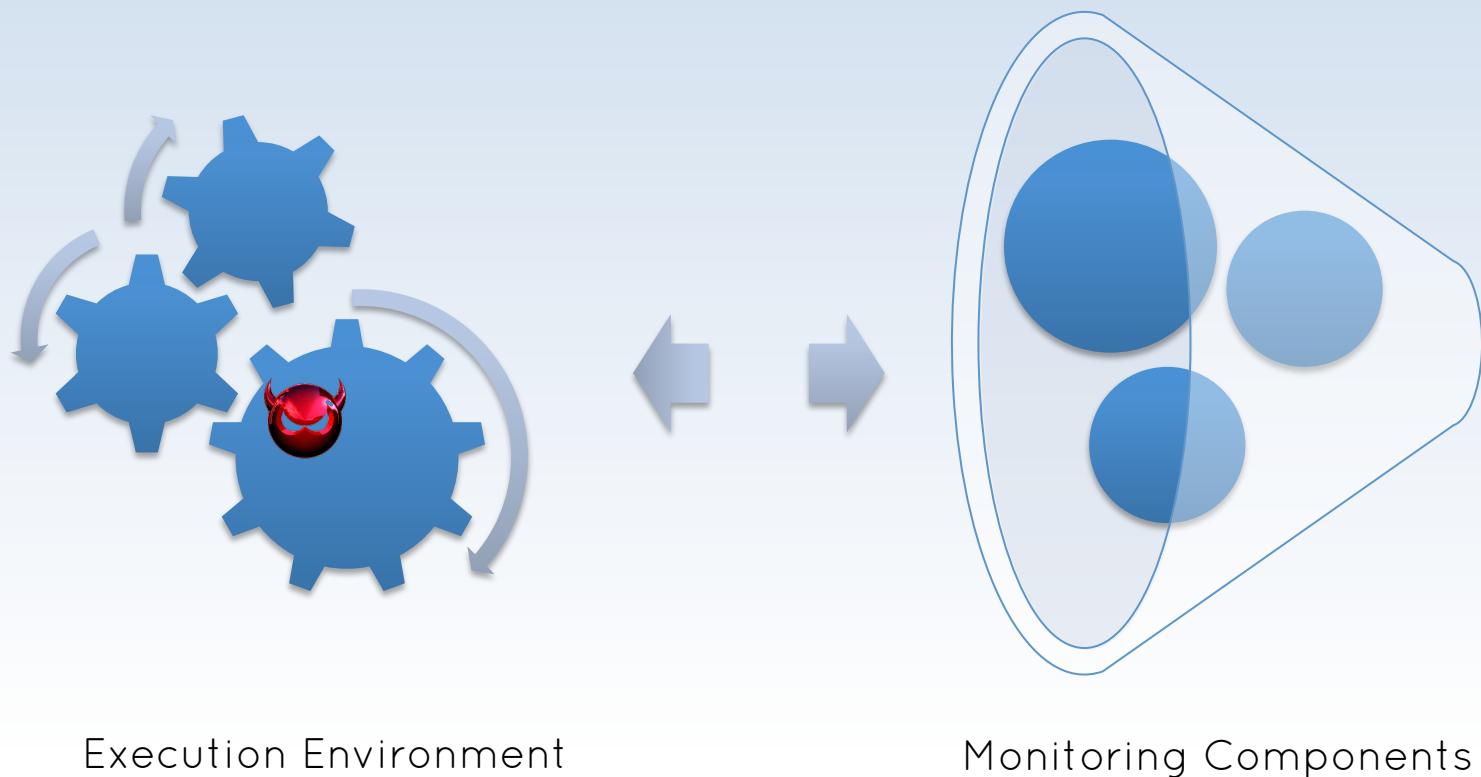
# Solutions?



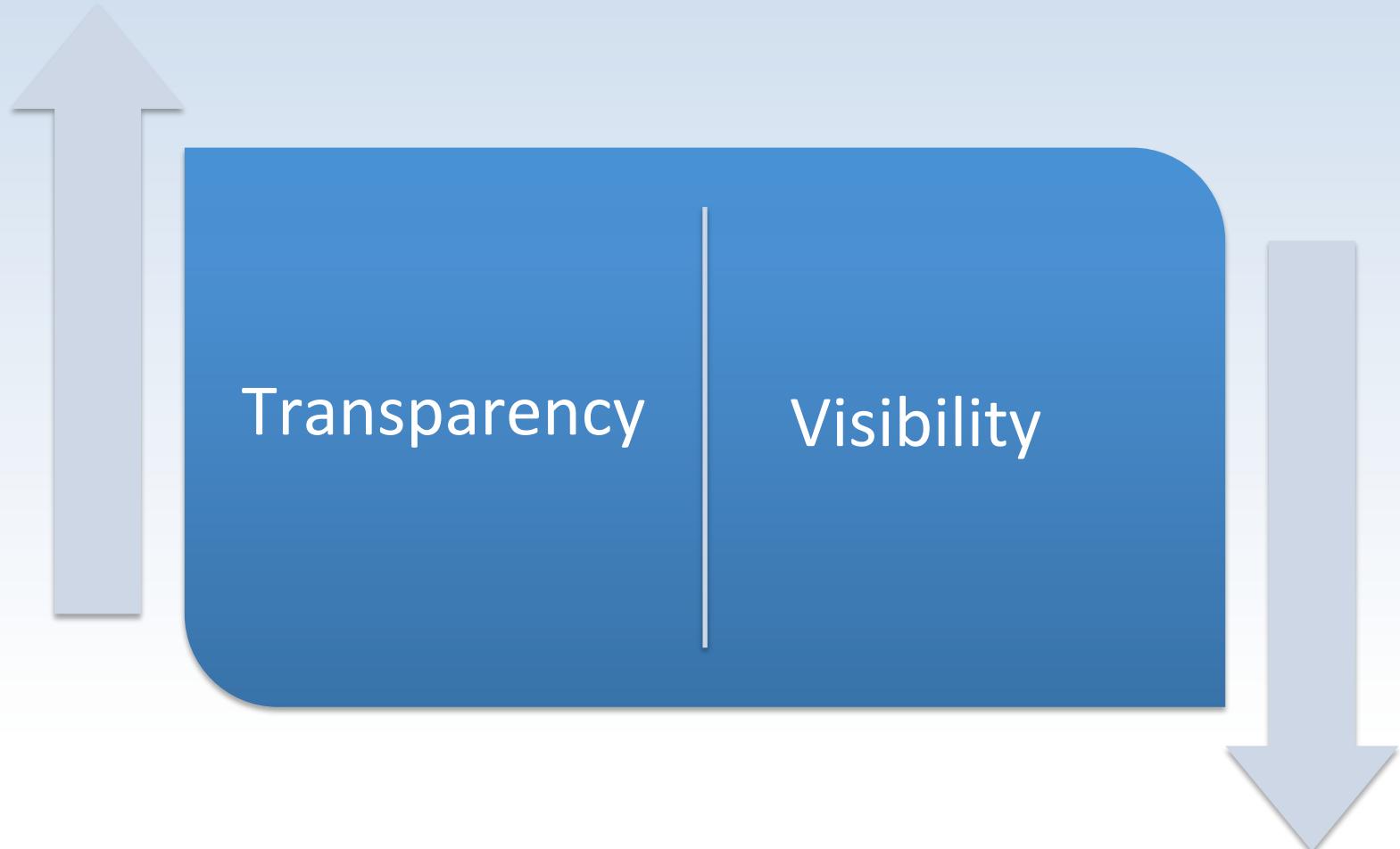
# Transparent Analysis



# Transparent Analysis

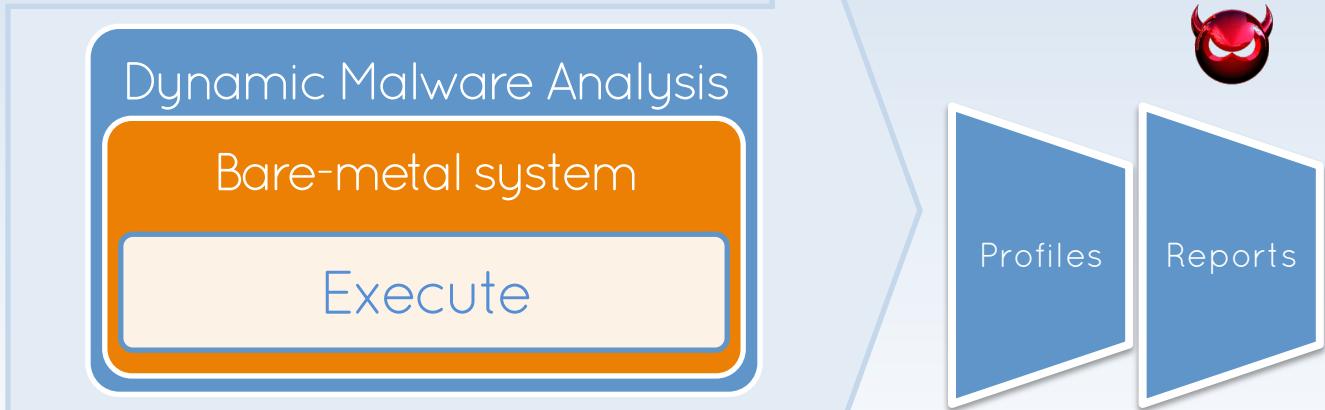


# Dynamic Malware Analysis

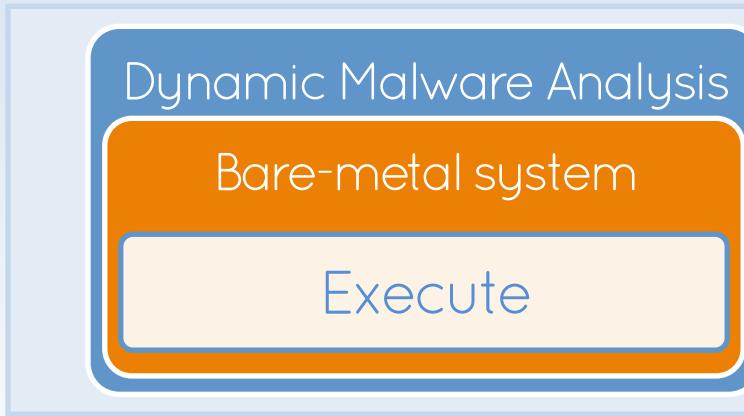


Can we automatically identify  
evasive malware under reduced  
visibility?

# BareCloud

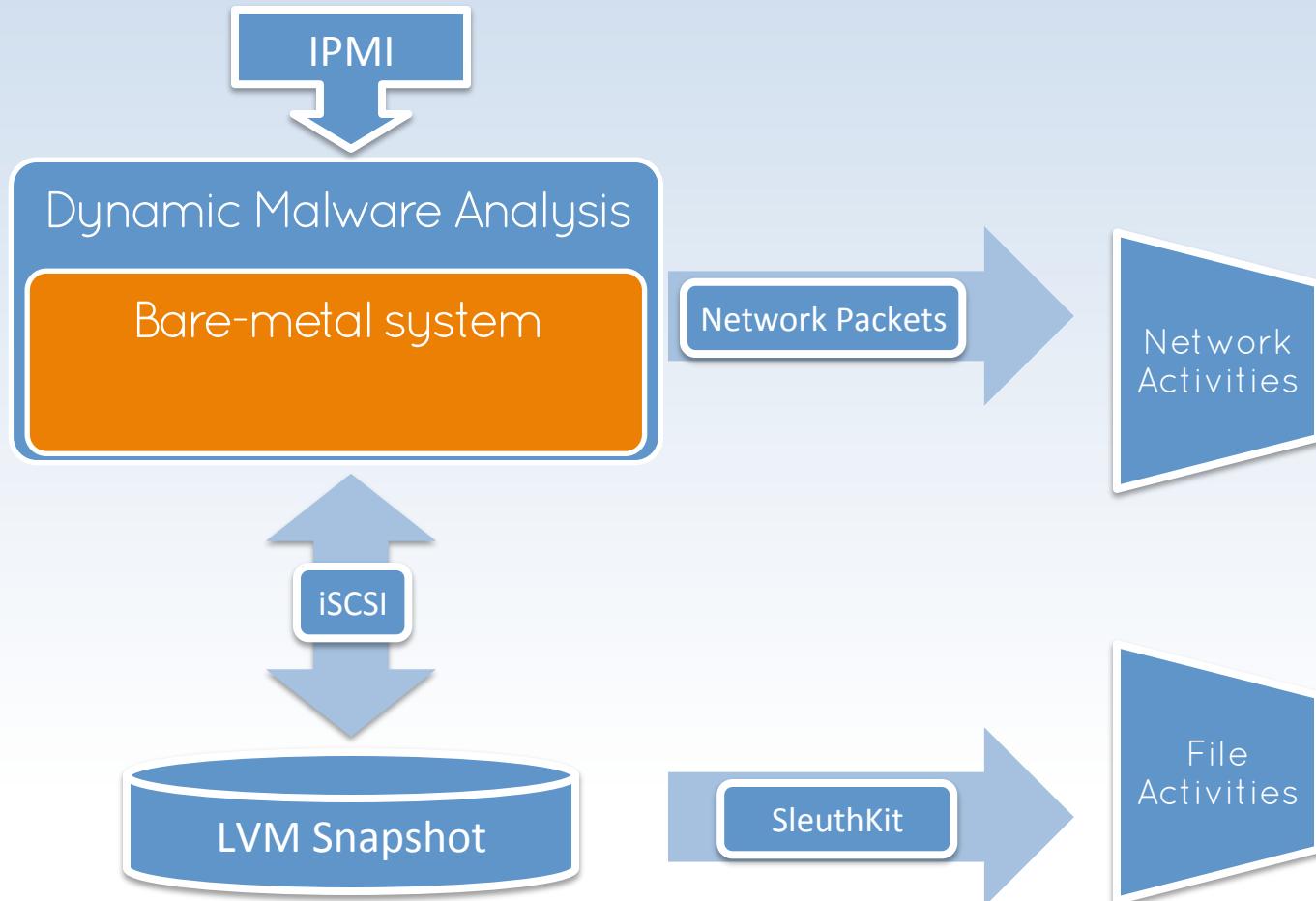


# BareCloud

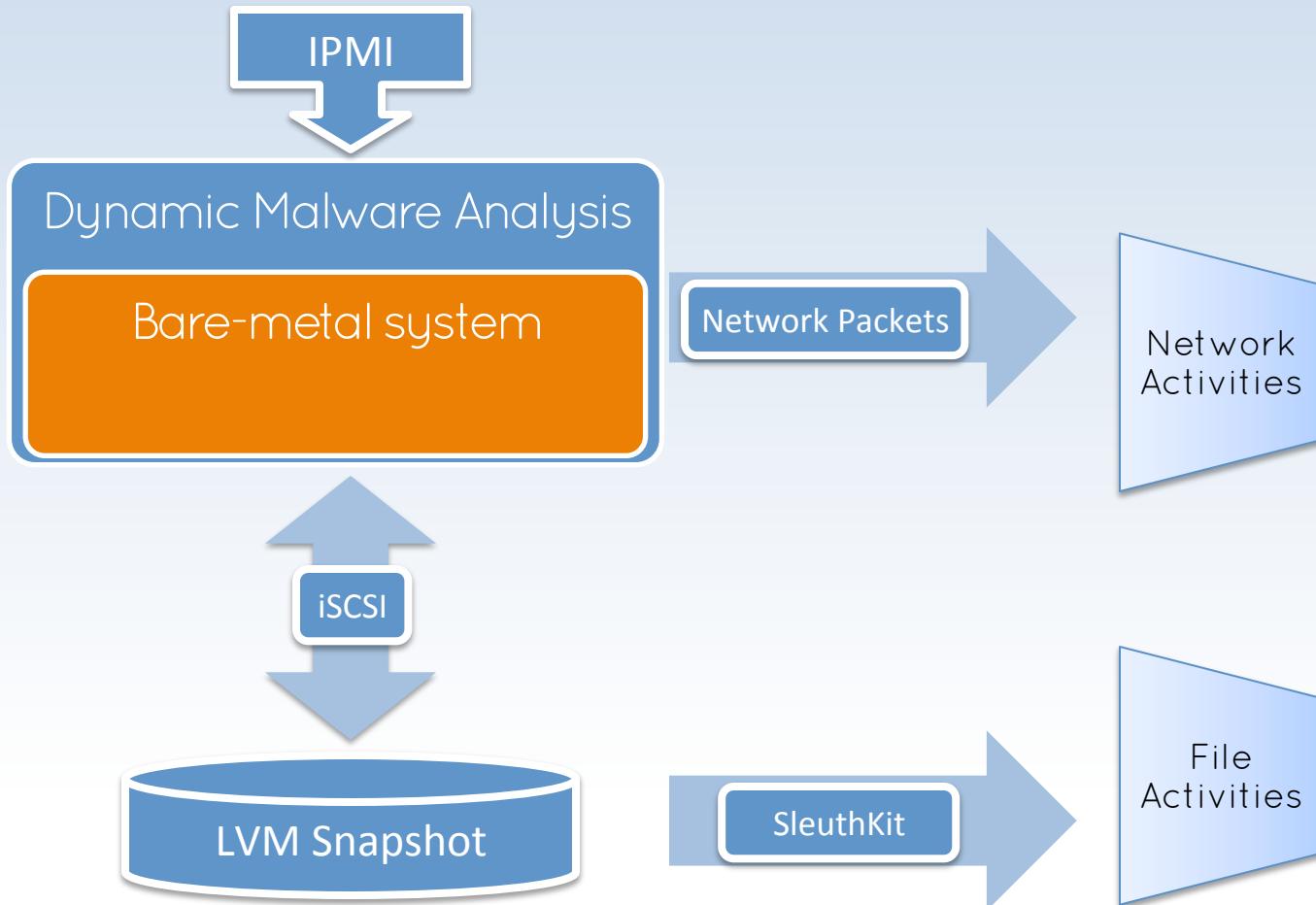


No in-guest  
monitoring component

# BareCloud

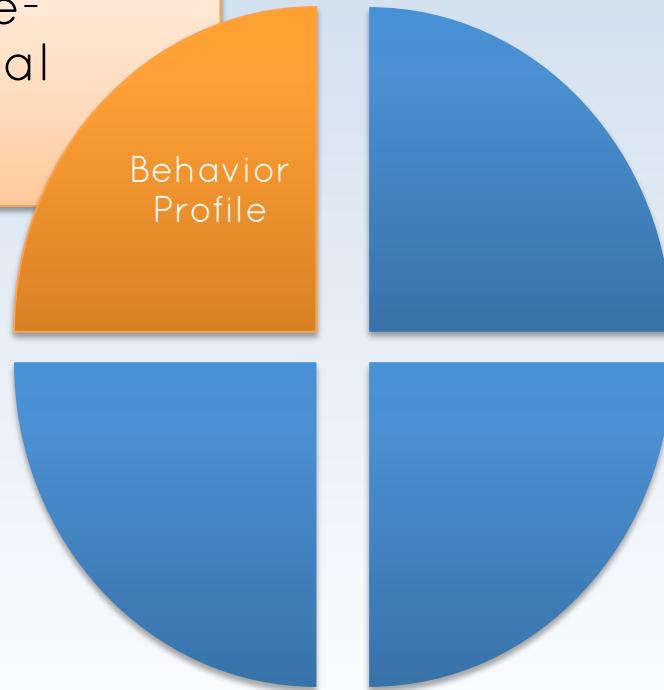


# BareCloud

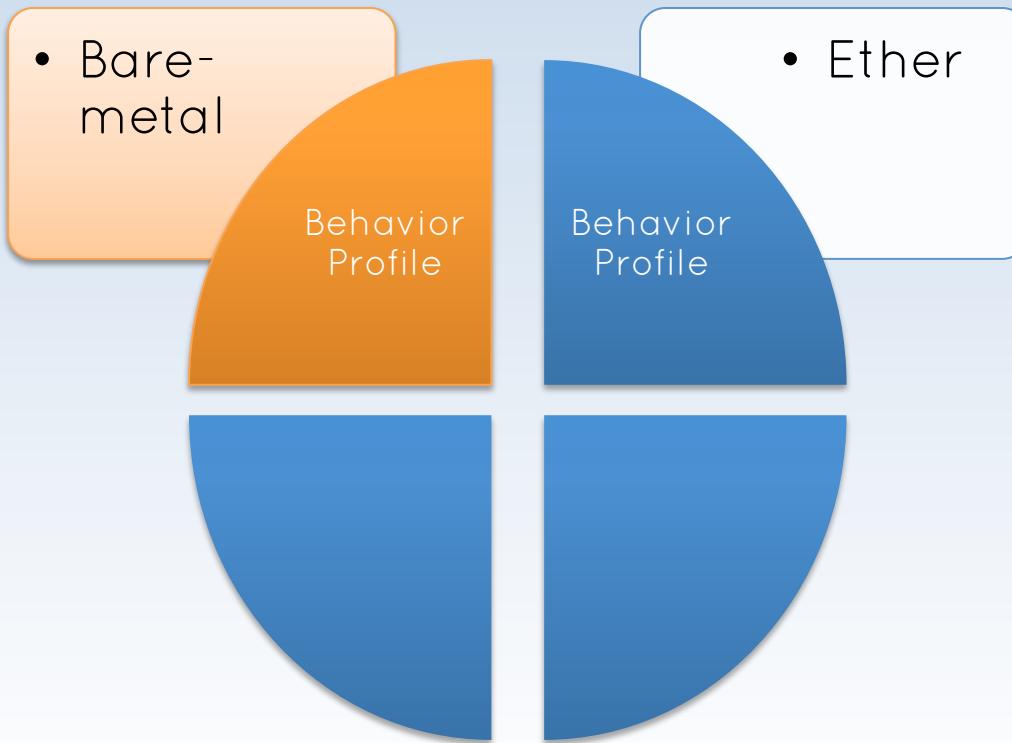


# BareCloud

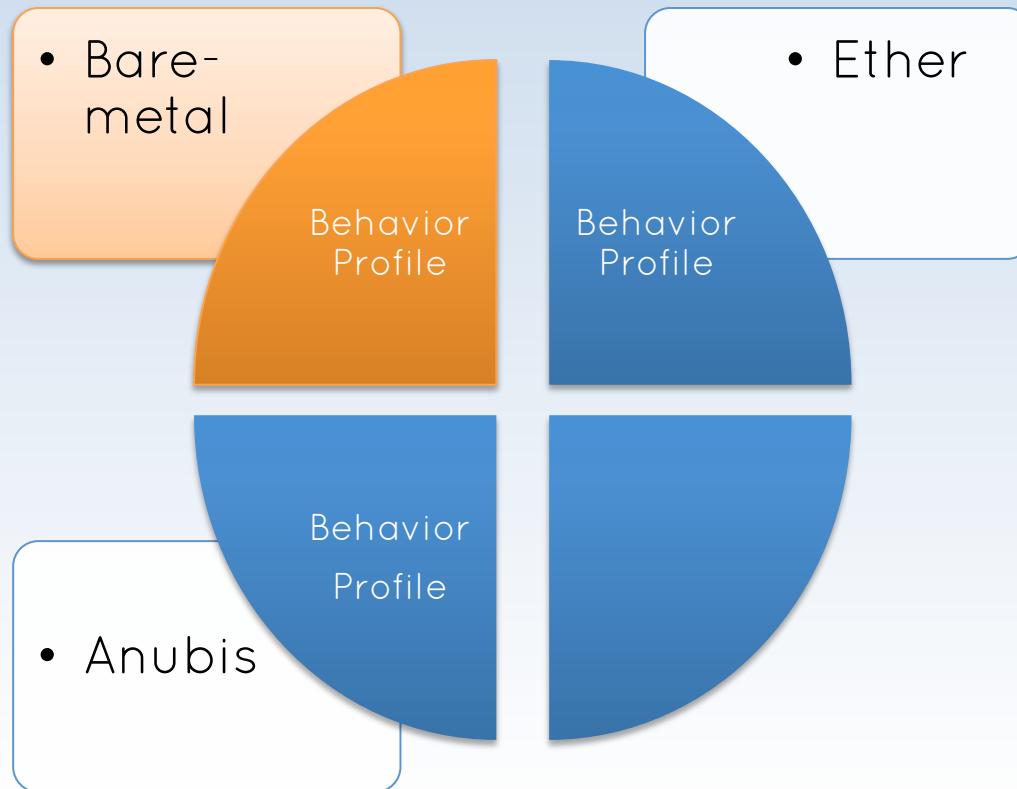
- Bare-metal



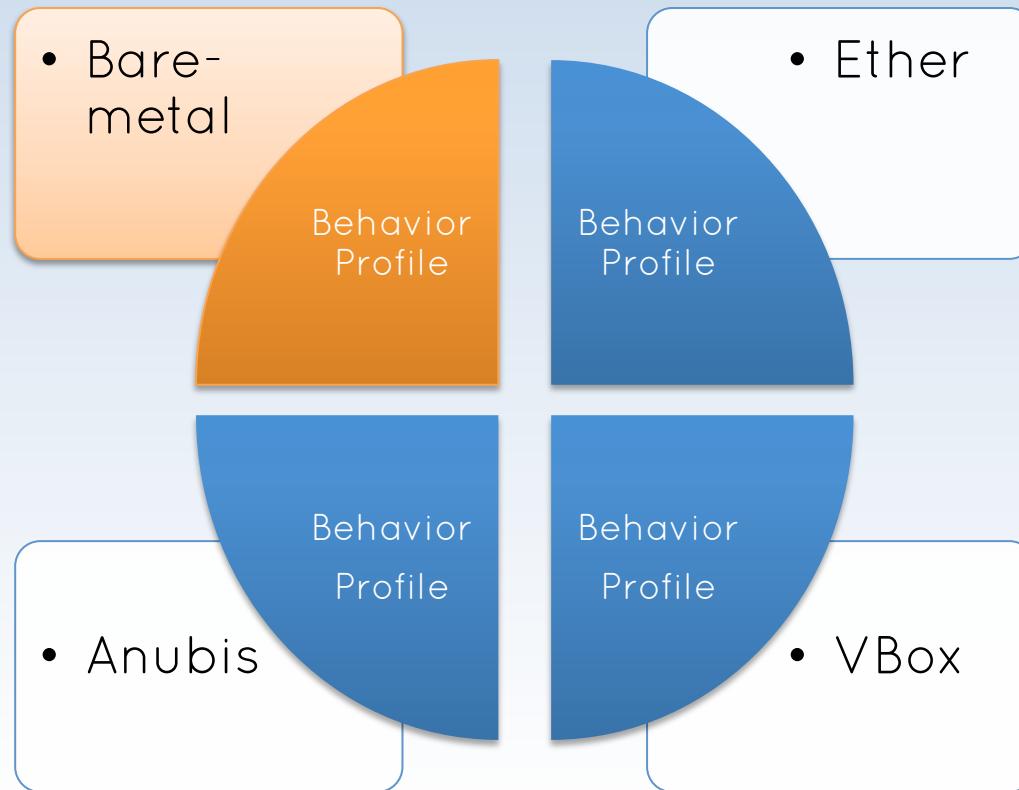
# BareCloud



# BareCloud



# BareCloud



# BareCloud

- Bare-metal

Behavior Profile

- Ether

Behavior Profile

- Anubis

Behavior Profile

- VBox

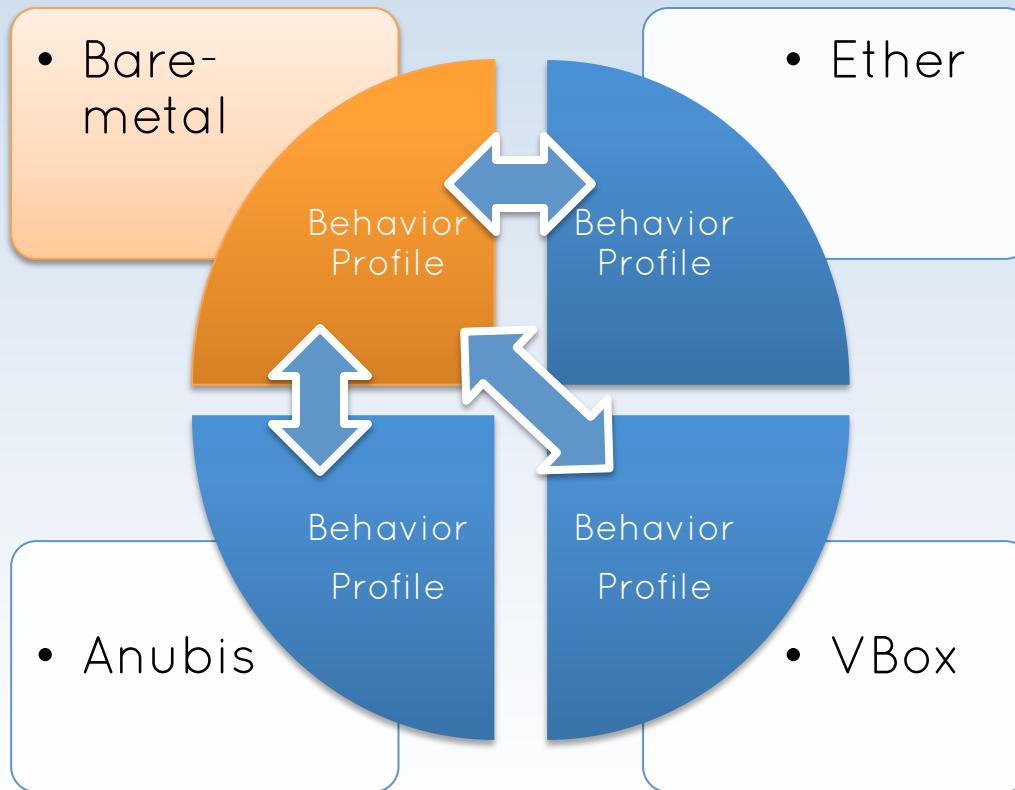
Behavior Profile

**QEMU**

open source processor emulator



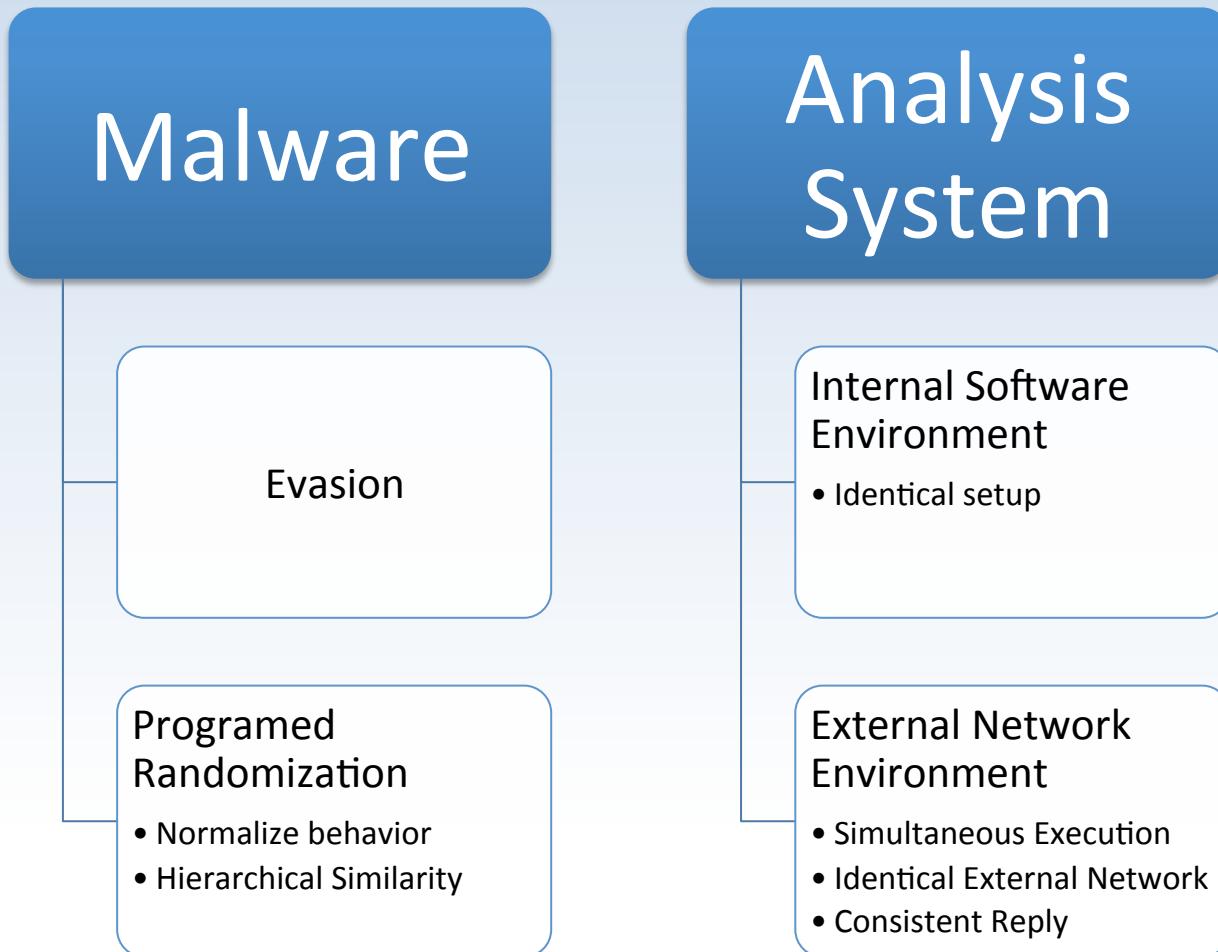
# BareCloud



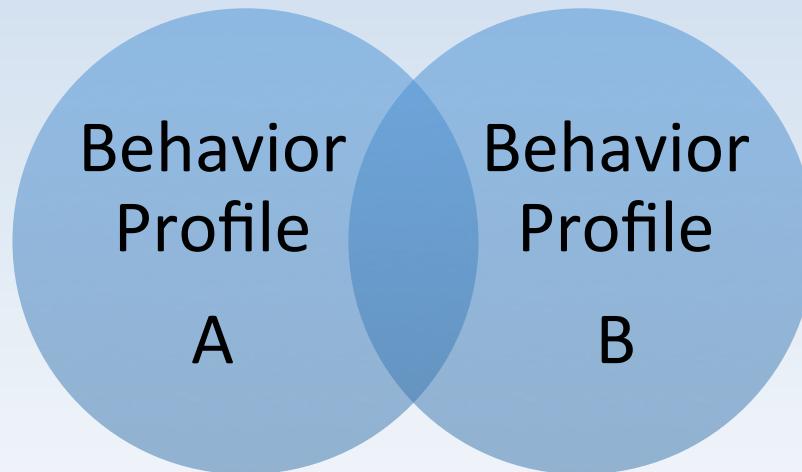
# Transient vs. Persistent



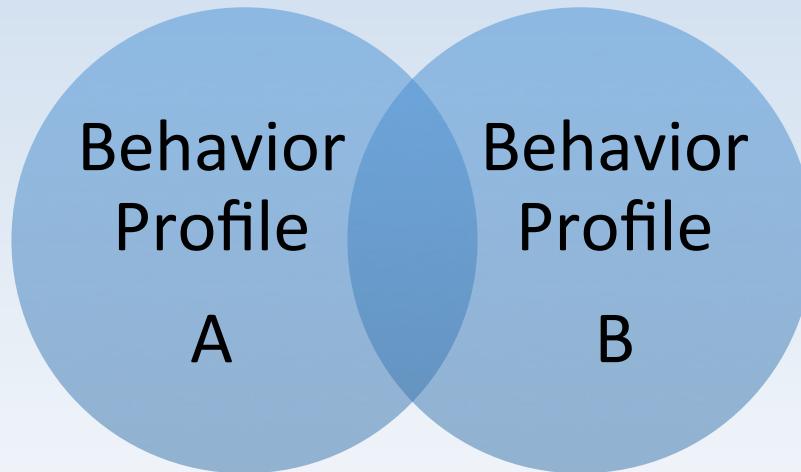
# Behavior Deviation



# Behavior Comparison



# Behavior Comparison



$$JaccardSimilarity = \frac{A \cap B}{A \cup B}$$

# Behavior Comparison

## Profile A

Create file X

Create file Y

Create file Z

## Profile B

Create file X

Create file Y

Modify file Z

## Profile C

Create file X

Create file Y

Connect to  
C&C

# Behavior Comparison

## Profile A

Create file X

Create file Y

Create file Z

## Profile B

Create file X

Create file Y

Modify file Z

## Profile C

Create file X

Create file Y

Connect to  
C&C

# Behavior Comparison

## Profile A

Create file X

Create file Y

Create file Z

## Profile B

Create file X

Create file Y

Modify file Z

## Profile C

Create file X

Create file Y

Connect to C&C

$$JaccardSimilarity(A, B) = 2/4 = JaccardSimilarity(A, C)$$

# Behavior Comparison

Behavior  
Profile

A

Behavior  
Profile

B

# Behavior Comparison

Behavior  
Profile

A

Behavior  
Profile

B

What type of events?

- Filesystem?
- Network?

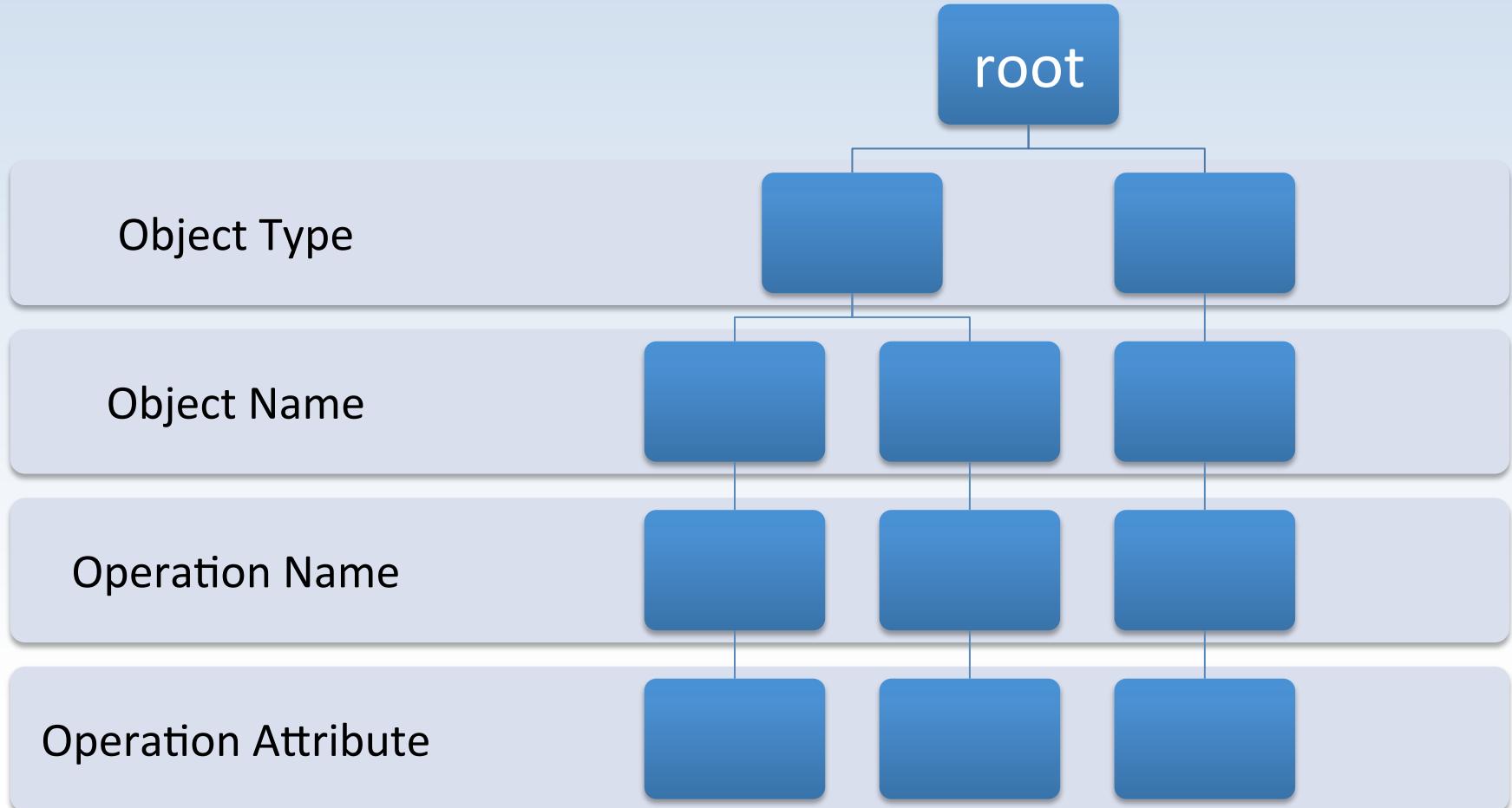
Are events related to  
the same object?

- Same file?
- Same network  
endpoint?

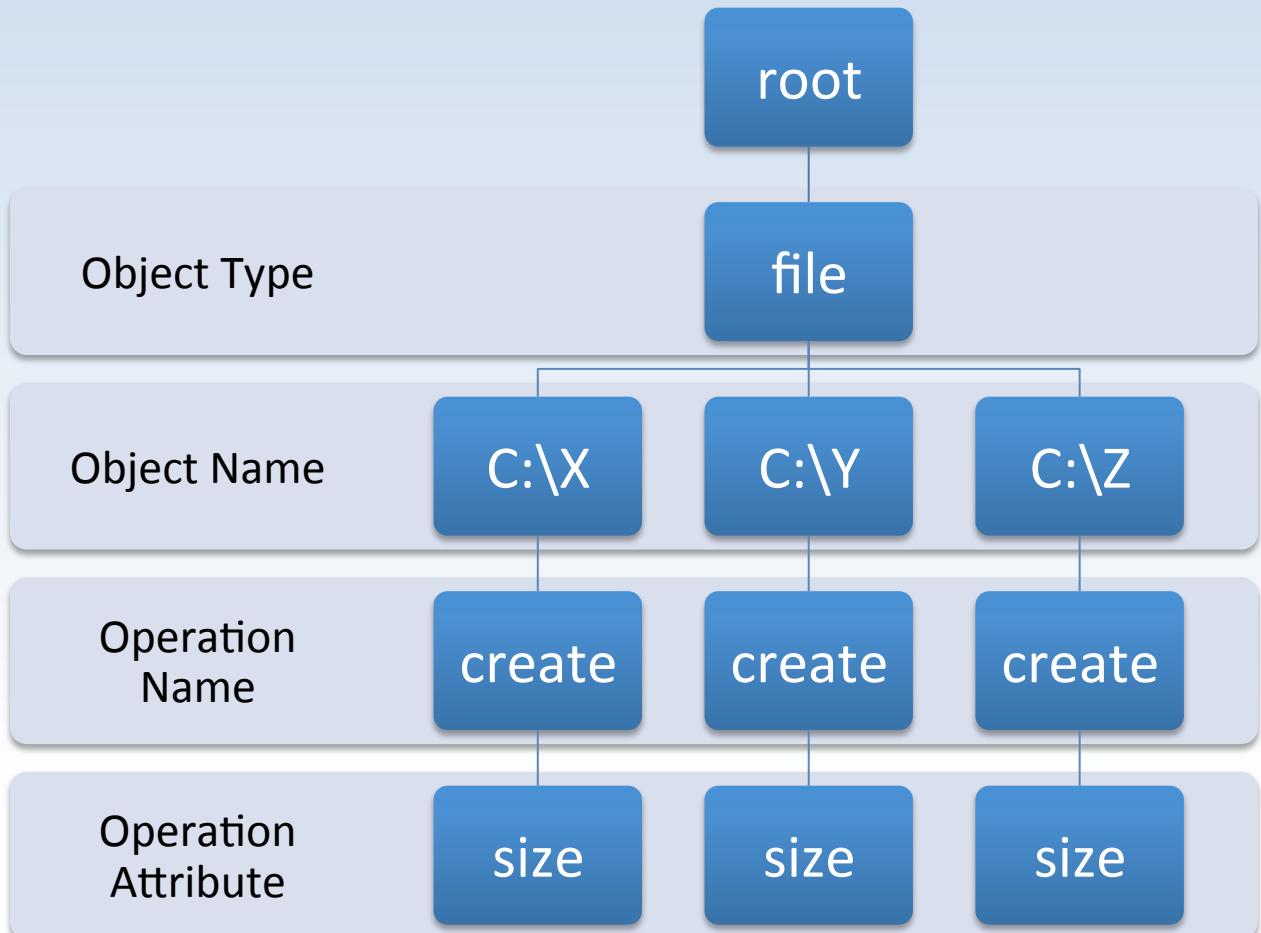
What type of  
operations?

- Create?
- Delete?
- HTTP?

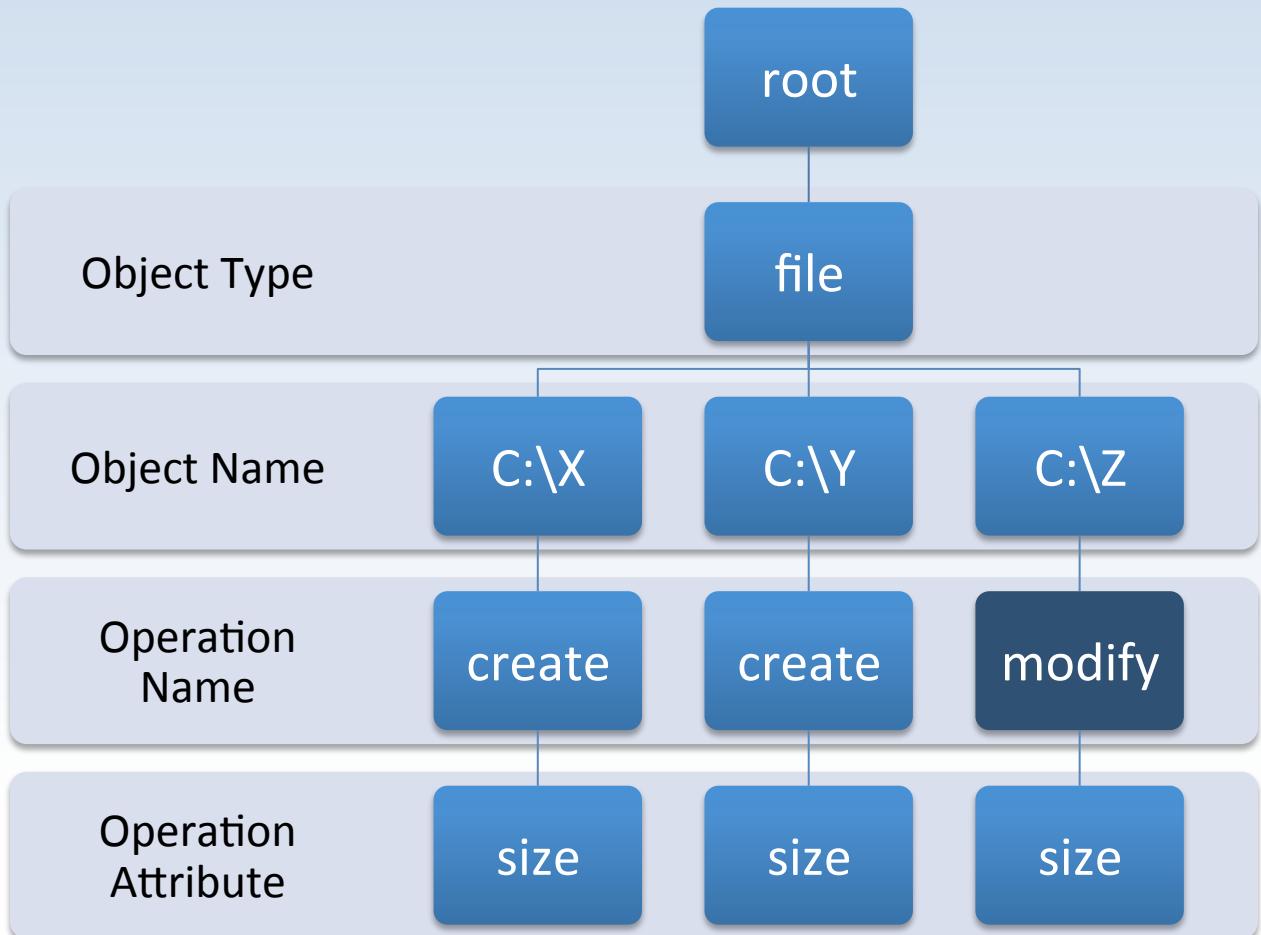
# Behavior Similarity Hierarchy



# Behavior Similarity Hierarchy



# Behavior Similarity Hierarchy



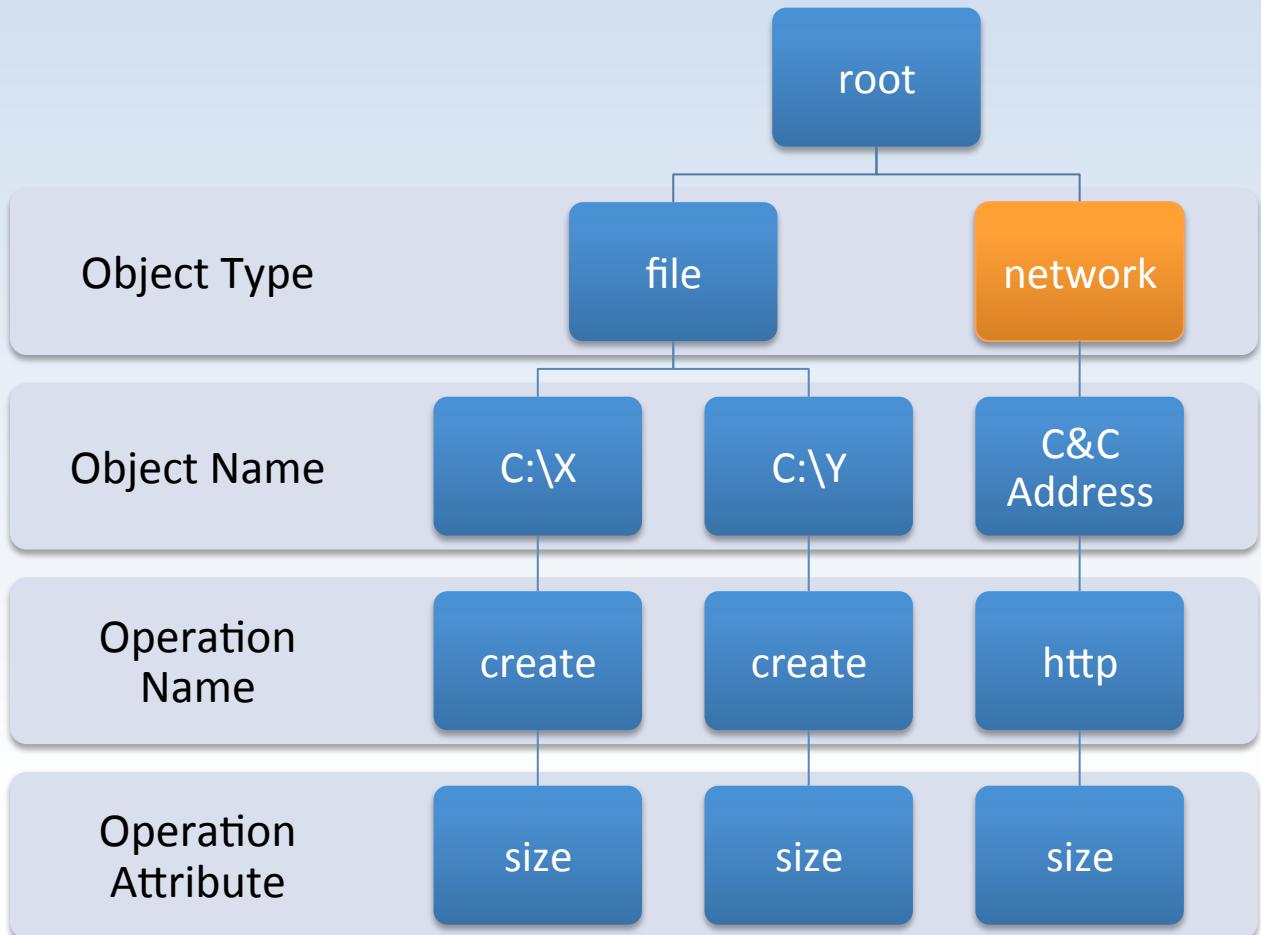
# Behavior Similarity Hierarchy

## Profile C

Create file X

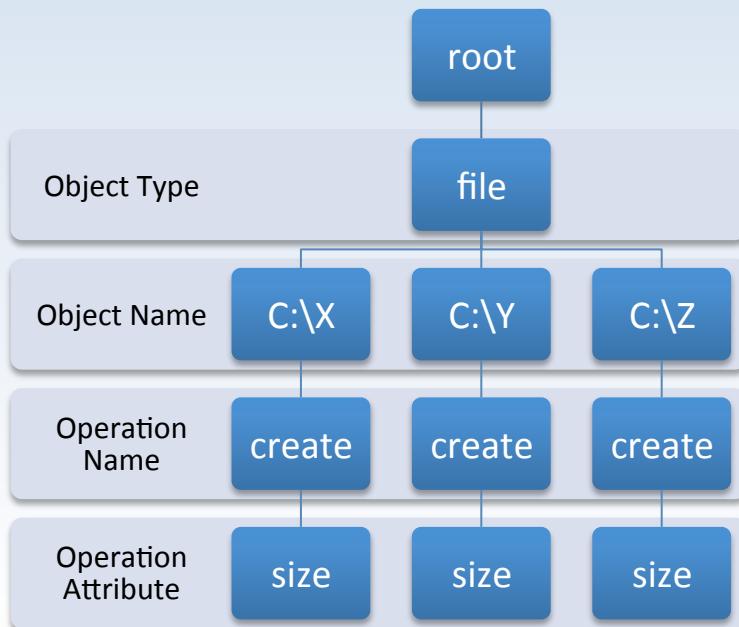
Create file Y

Connect to C&C

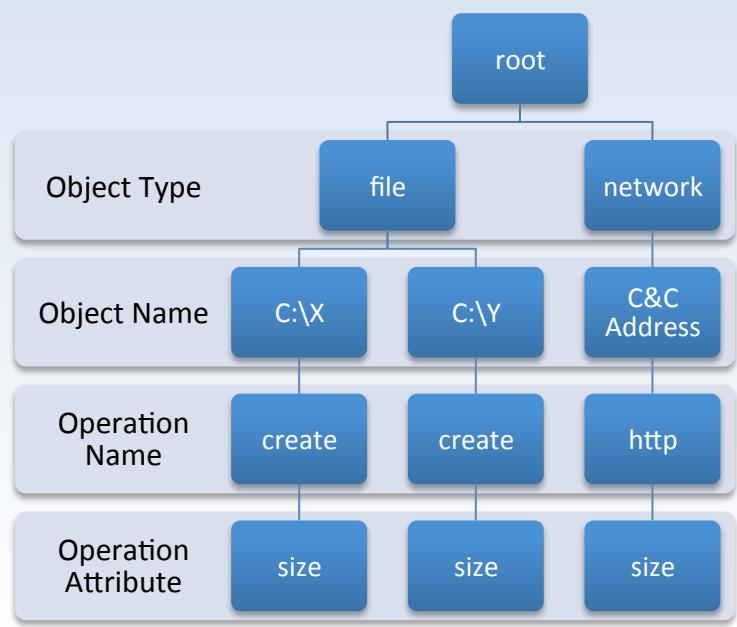


# Hierarchical Similarity

Profile A

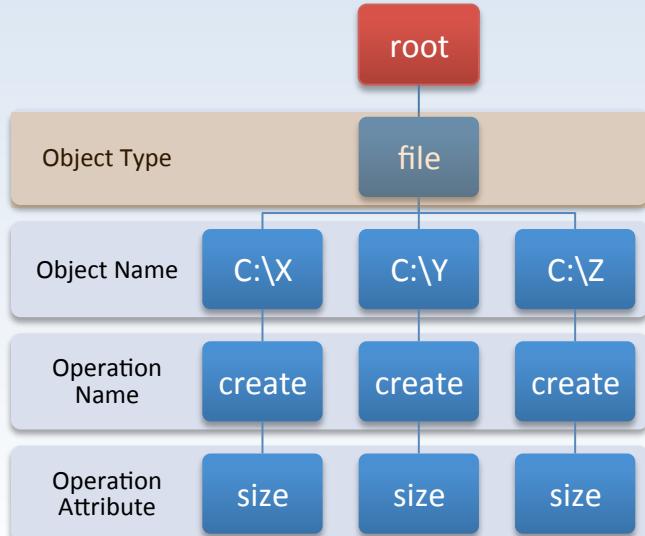


Profile C

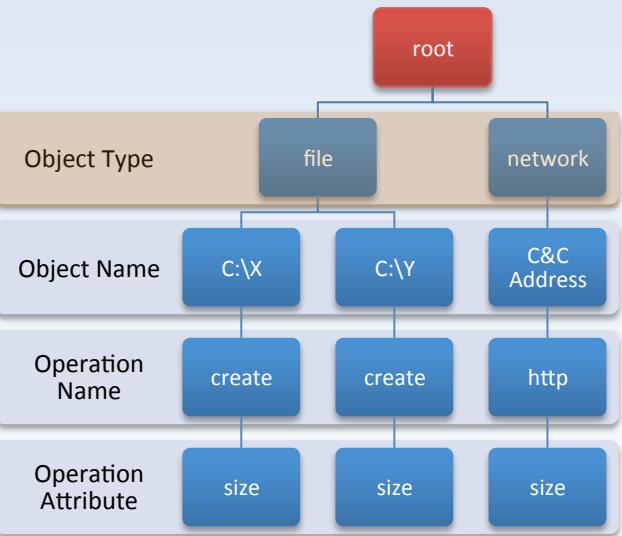


# Hierarchical Similarity

Profile A



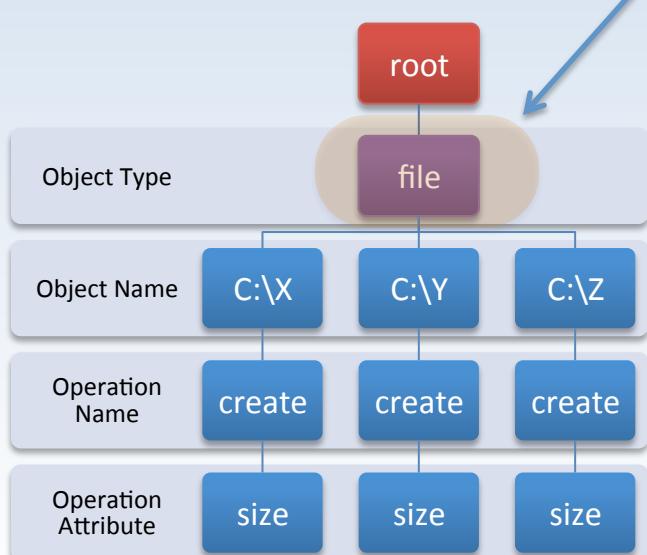
Candidate Sets



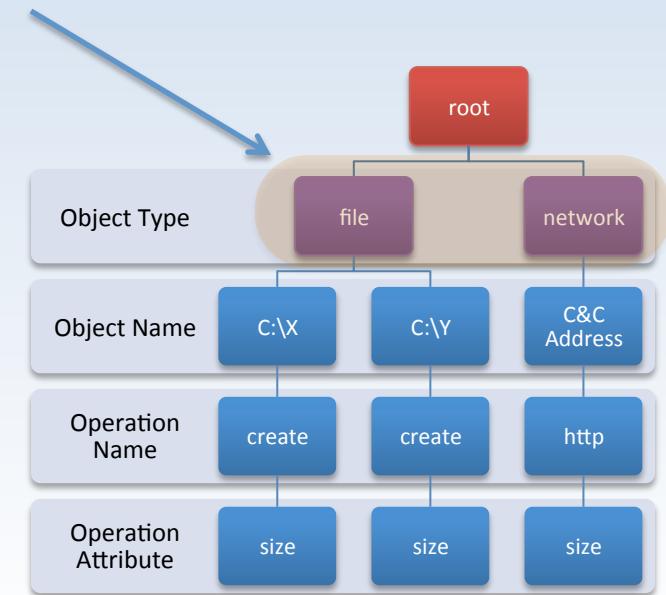
Profile C

# Hierarchical Similarity

Profile A



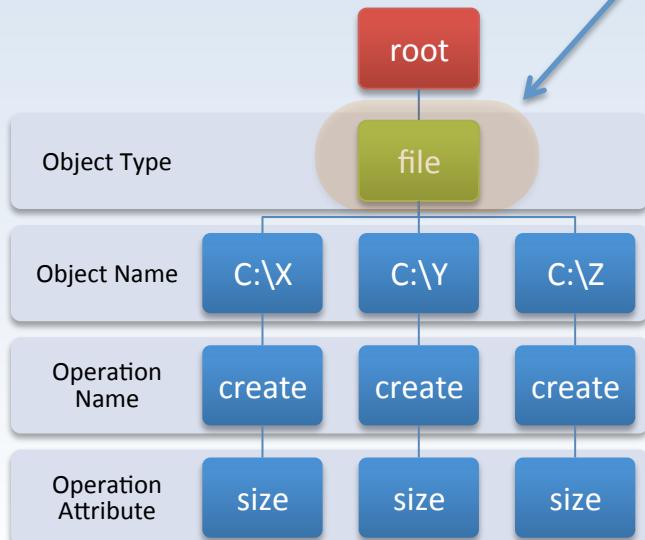
Candidate Sets



Profile C

# Hierarchical Similarity

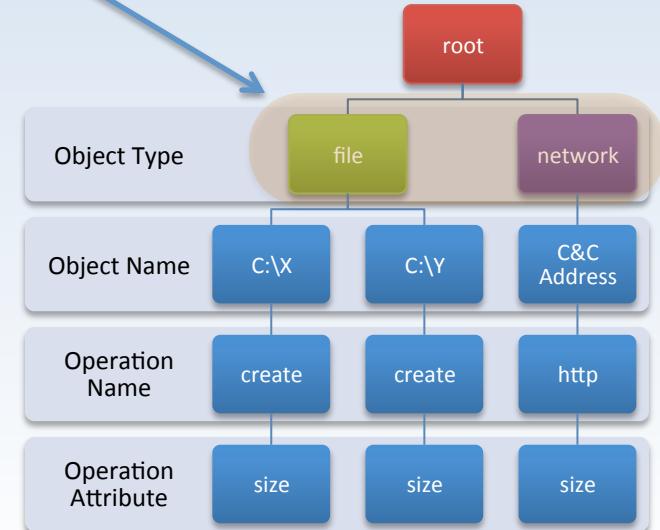
Profile A



Candidate Sets

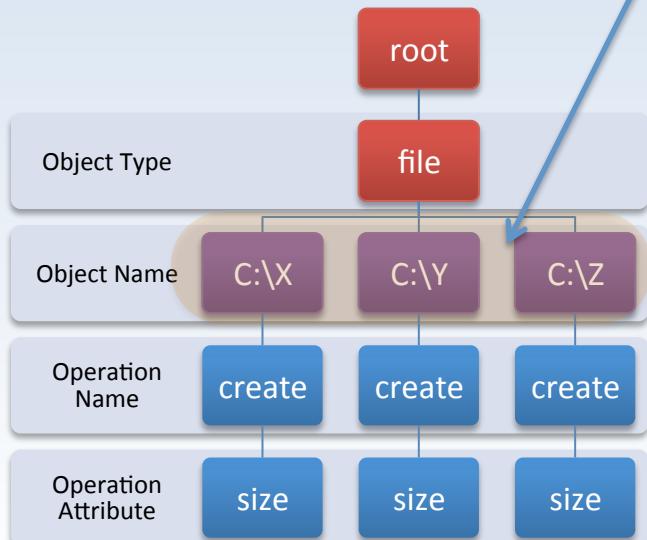
$$\text{Sim}_1 = 1/2$$

Profile C



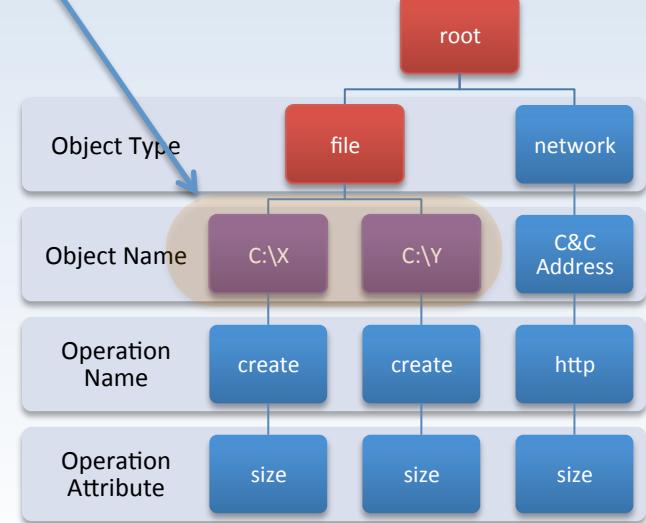
# Hierarchical Similarity

Profile A



Candidate Sets

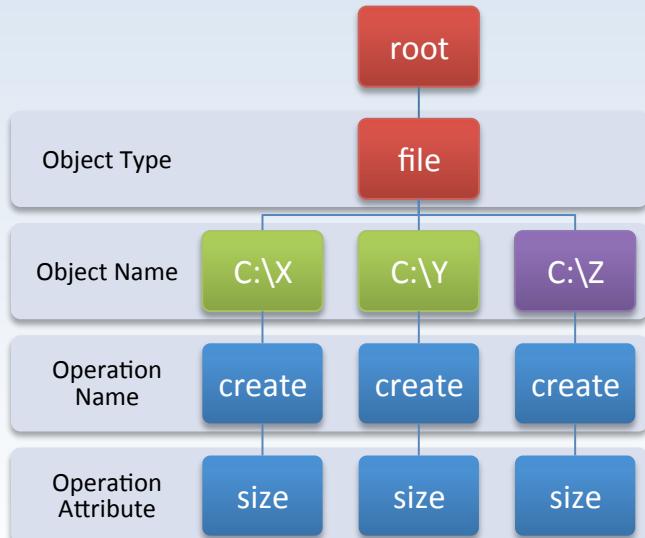
$$\text{Sim}_1 = 1/2$$



Profile C

# Hierarchical Similarity

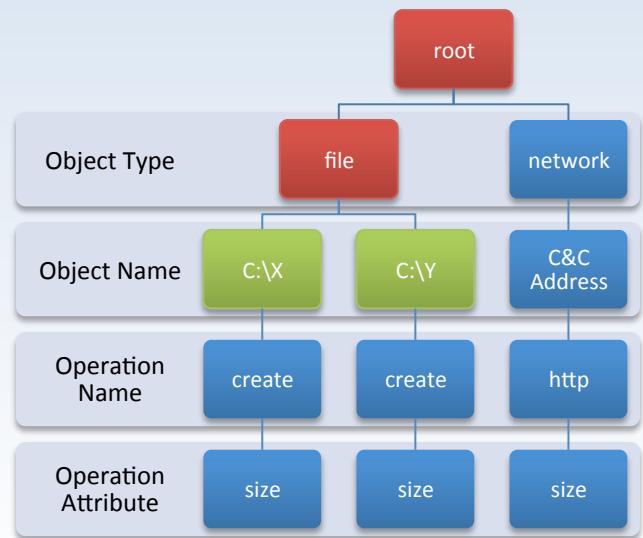
Profile A



$$\text{Sim}_1 = 1/2$$

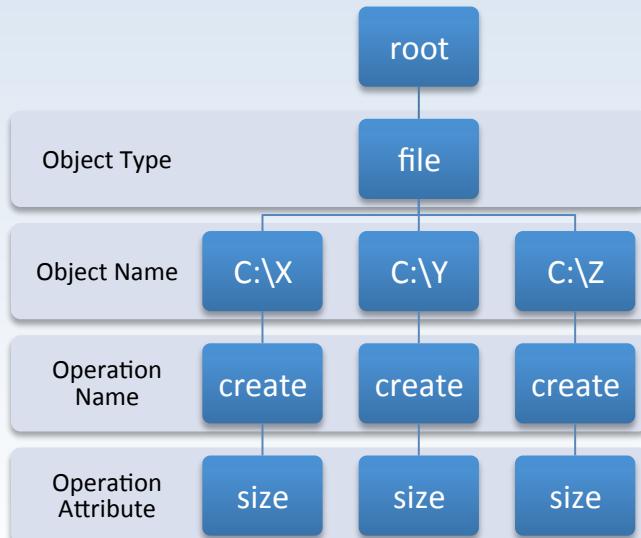
$$\text{Sim}_2 = 2/3$$

Profile C



# Hierarchical Similarity

Profile A



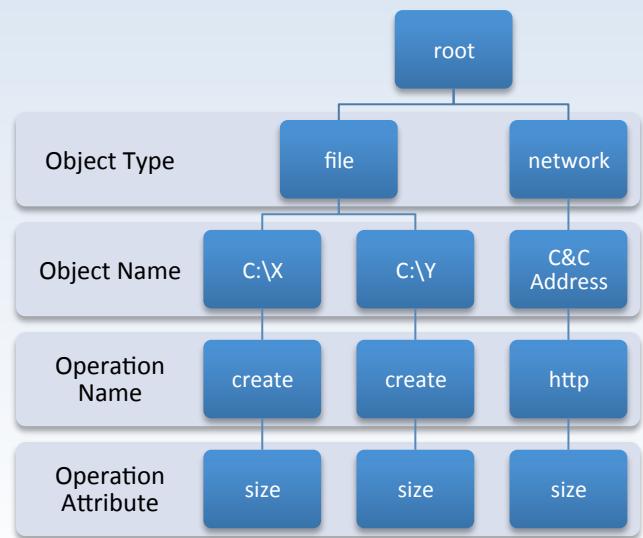
$$\text{Sim}_1 = 1/2$$

$$\text{Sim}_2 = 2/3$$

$$\text{Sim}_3 = 1$$

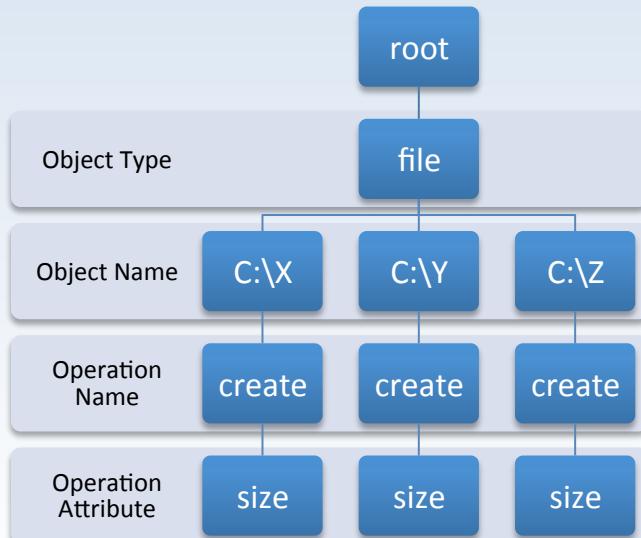
$$\text{Sim}_4 = 1$$

Profile C



# Hierarchical Similarity

Profile A



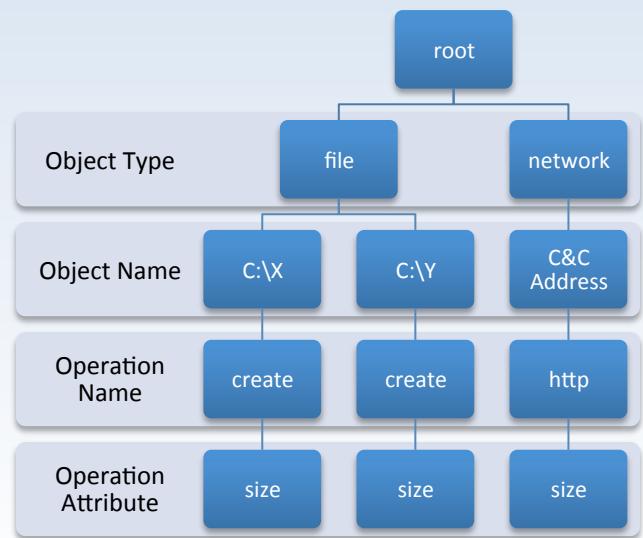
$$\text{Sim}_1 = 1/2$$

$$\text{Sim}_2 = 2/3$$

$$\text{Sim}_3 = 1$$

$$\text{Sim}_4 = 1$$

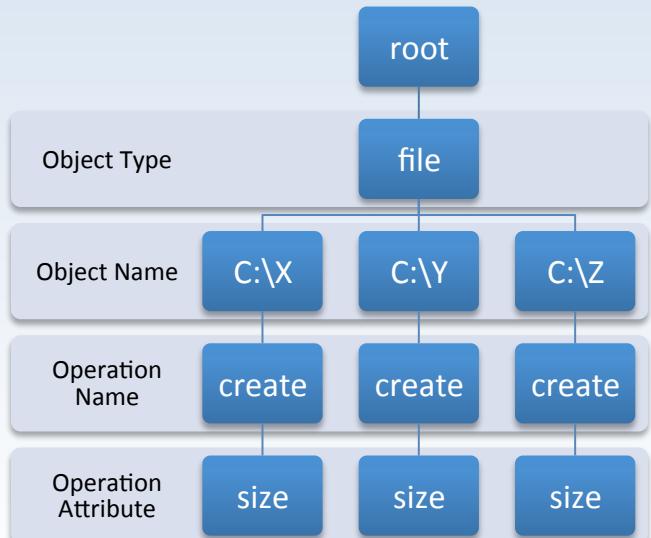
Profile C



$$\text{Sim}(A, C) = \text{AVG}(\text{Sim}_1 \dots \text{Sim}_4) = 0.79$$

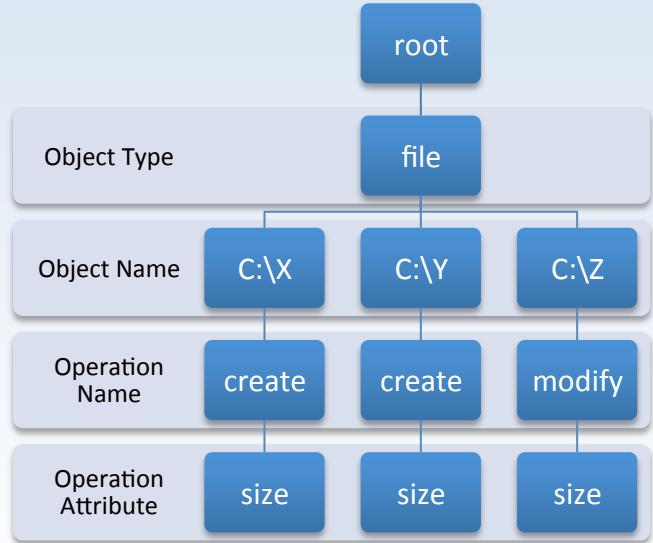
# Hierarchical Similarity

Profile A



$$\text{Sim}_1 = 1$$

Profile B



$$\text{Sim}_2 = 1$$

$$\text{Sim}_3 = 1/2$$

$$\text{Sim}_4 = 1$$

$$\text{Sim}(A, B) = \text{AVG}(\text{Sim}_1 \dots \text{Sim}_4) = 0.87$$

# Behavior Comparison

## Profile A

Create file X

Create file Y

Create file Z

## Profile B

Create file X

Create file Y

Modify file Z

## Profile C

Create file X

Create file Y

Connect to C&C

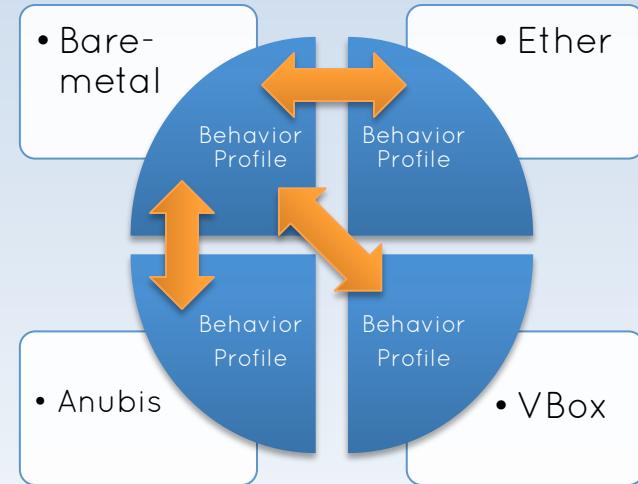
$$JaccardSimilarity(A, B) == JaccardSimilarity(A, C)$$

$$\begin{aligned} HierarchicalSim(A, B) &> HierarchicalSim(A, C) \\ 0.87 &> 0.79 \end{aligned}$$

# Deviation Score

- Behavior Distance

$$Distance(A, B) = 1 - Sim(A, B)$$



- Deviation Score  $D$

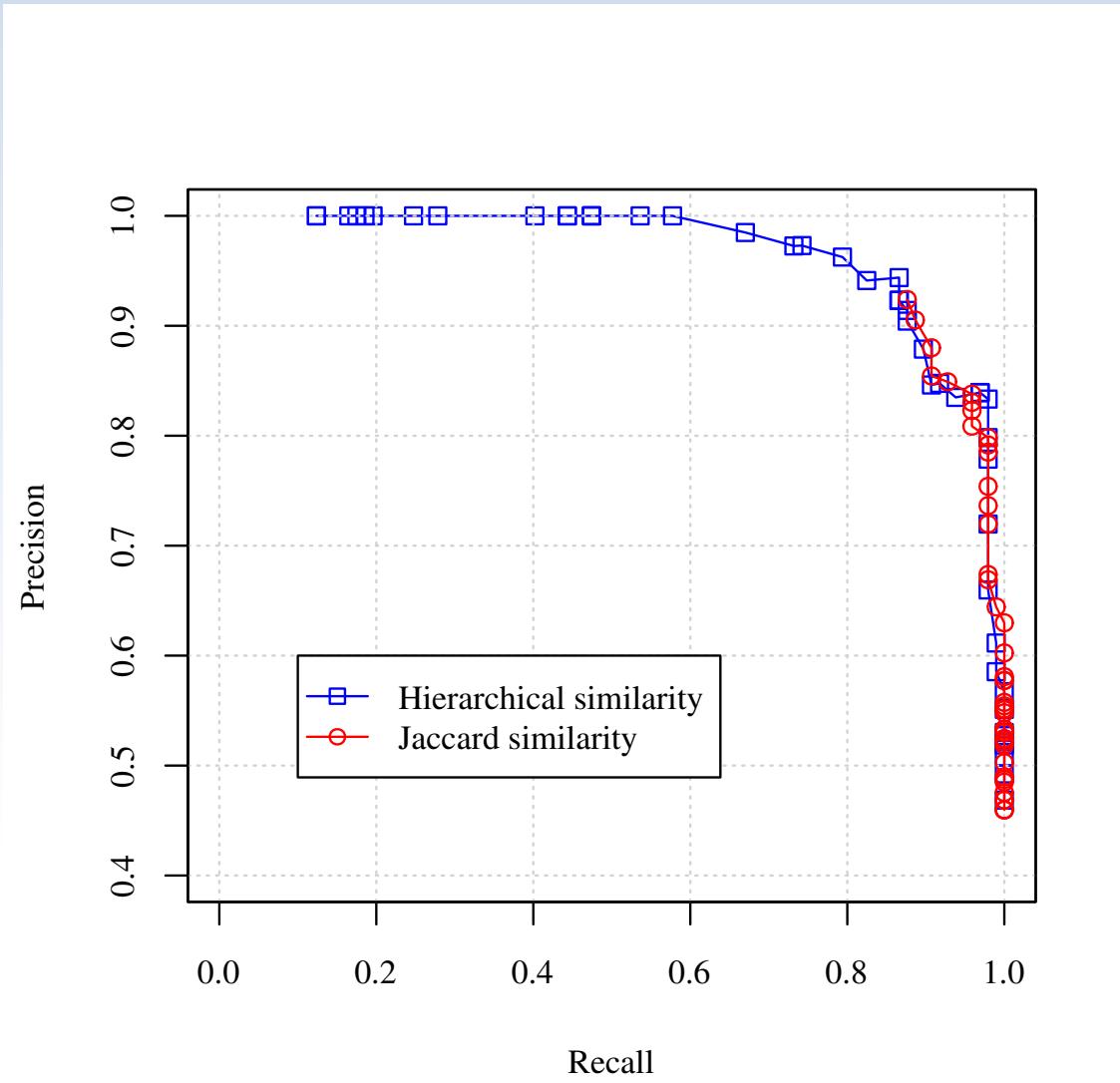
Quadratic mean of the behavior distances with respect to the bare-metal analysis

- Deviation Threshold  $t$ 
  - Evasive if  $D > t$

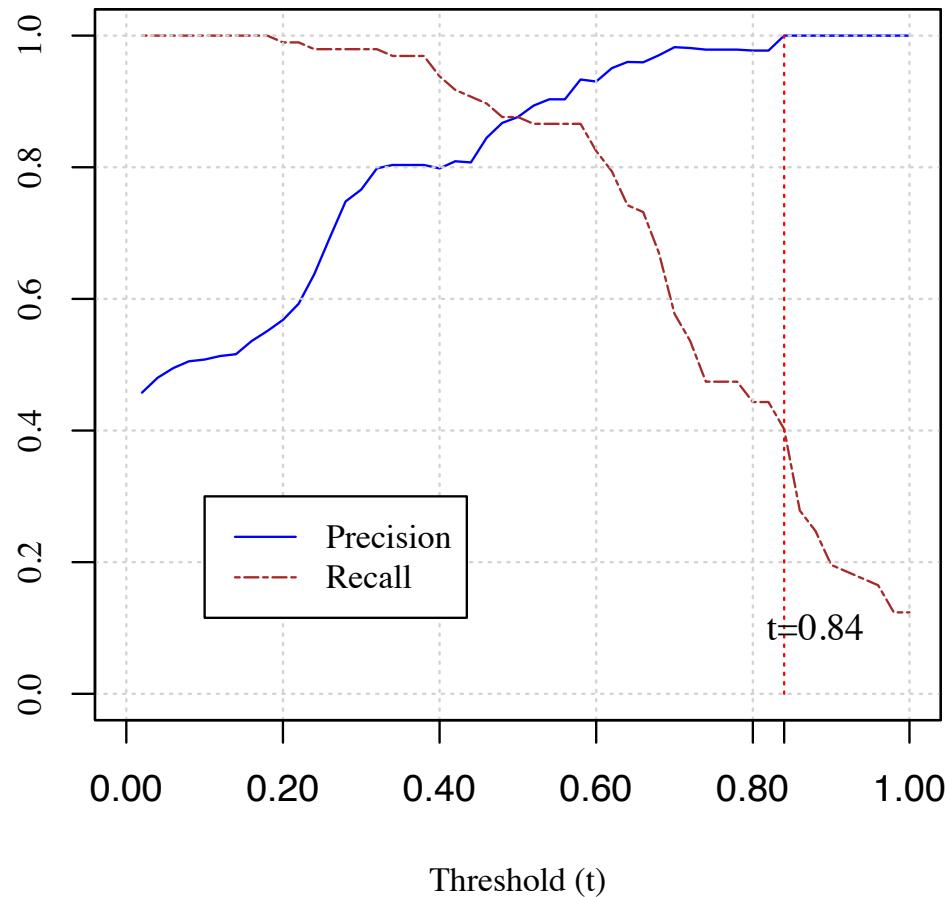
# Evaluation

- Ground truth
  - 111 evasive samples (29 families)
  - 119 non-evasive samples (49 families)
- Calculated behavior Deviation score **D**
- Calculate Jaccard distance-based deviation **JD**
  - Maximum Jaccard-distance among different behavior profiles of a malware
- Precision-recall analysis by varying the deviation threshold **t**

# Evaluation



# Evaluation



# Large-scale Evaluation

- Recent real-world malware feed observed by Anubis
- Randomly select samples with
  - low system and low network activity
  - high system and high network activity
  - high system but low network activity
  - Low system but high network activity
- 110,005 samples
  - 4 months period beginning from July 2013

# Large-scale Evaluation

Environment	Detection Count	Percentage
Anubis	4947	84.78
Ether	4562	78.18
VirtualBox	3576	61.28
All	2530	43.35

5,835 evasive malware out of 110,005 recent samples

# Limitations

- Hardware vs software iSCSI initiator
- Stalling code
  - Wait for user input
  - Advanced waiting
- Decoy reconnaissance
  - Real hardware ID not randomized

# Conclusions

- Evasive Malware is a real threat to the new wave of dynamic analysis based malware detection systems
- We presented a system that can detect these evasive malware automatically

Thank You!

# Questions