RSAConference2019
San Francisco | March 4–8 | Moscone Center

SESSION ID: MASH-R02

Threat Hunting Using 16th Century Math and Sesame Street

Vernon Habersetzer, CISSP, EnCE, GCFA, GCFE, GCIH, CISA, CFE, GREM, GCIA

Hunt Team Lead
Walmart Stores, Inc.
@HuntingNomad

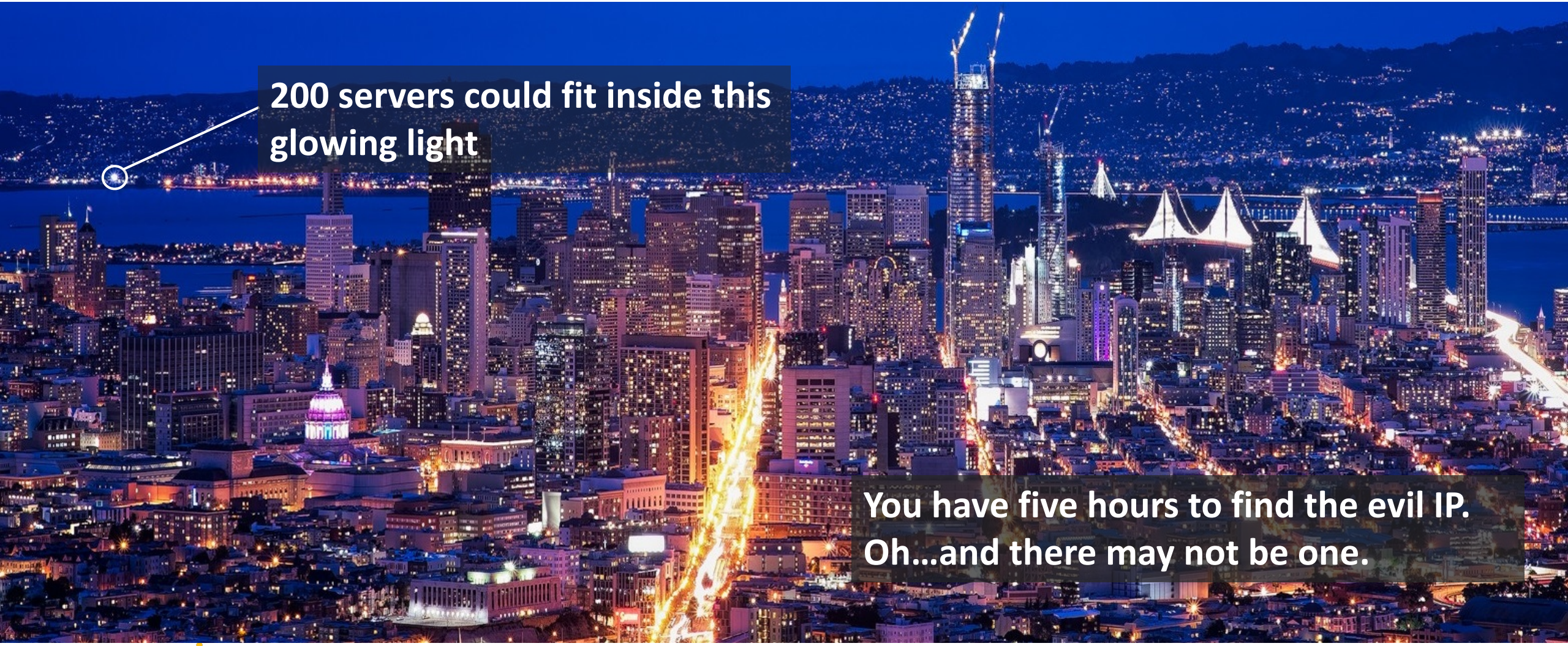
It’s 1983…

Fast Forward 16 years to 1996…

Networking, 2600 and phrack

THREAT HUNTING AT SCALE

Every pixel represents an IP in this 1MP image…



THREAT HUNTER'S GOAL

Make it difficult for adversaries to hide!

HOW?

Sometimes we get “Hunter’s Block”™

We need a new approach!

HOW?

We tend to focus too much on granular threat detection.

We need to look at threat detection more openly.

HOW?

Our focus should be on hunting methods that:

- Are scalable and sustainable
- Transcend attack specifics (as much as possible)
  - Try as you may, you'll never create enough granular detection logic to catch every variation of malicious behavior
  - TTP-based hunting should not be overlooked, but try to keep logic open

THREAT HUNTING AT SCALE

Think about the scale

THREAT HUNTING AT SCALE

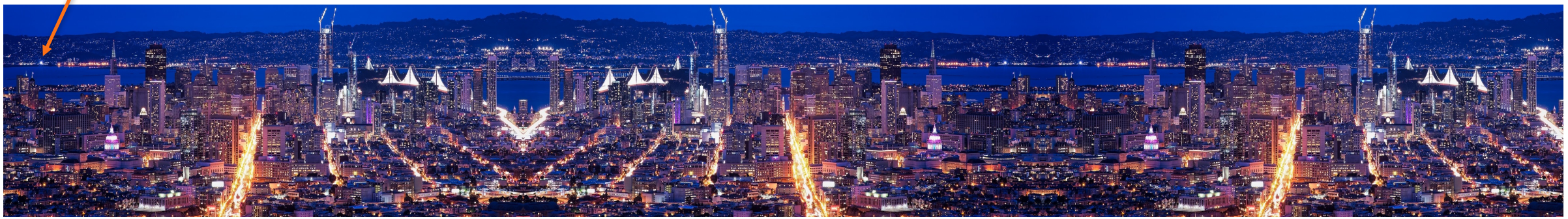
Let's **double** the size of the network….2 million IPs?

THREAT HUNTING AT SCALE

Let's **triple** the size….3 million IPs, 8B connections per day!

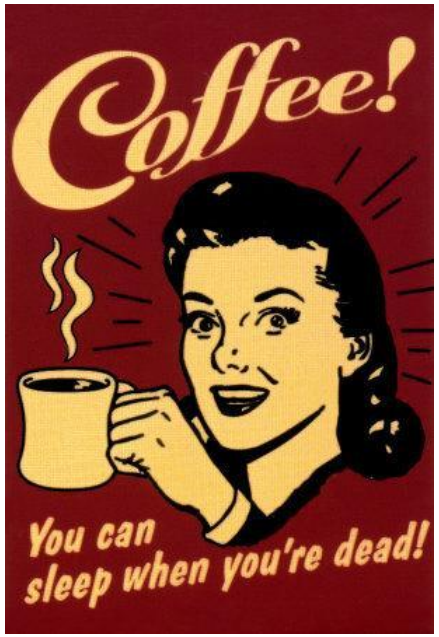
The numbers *seem* to be against you!

200 servers



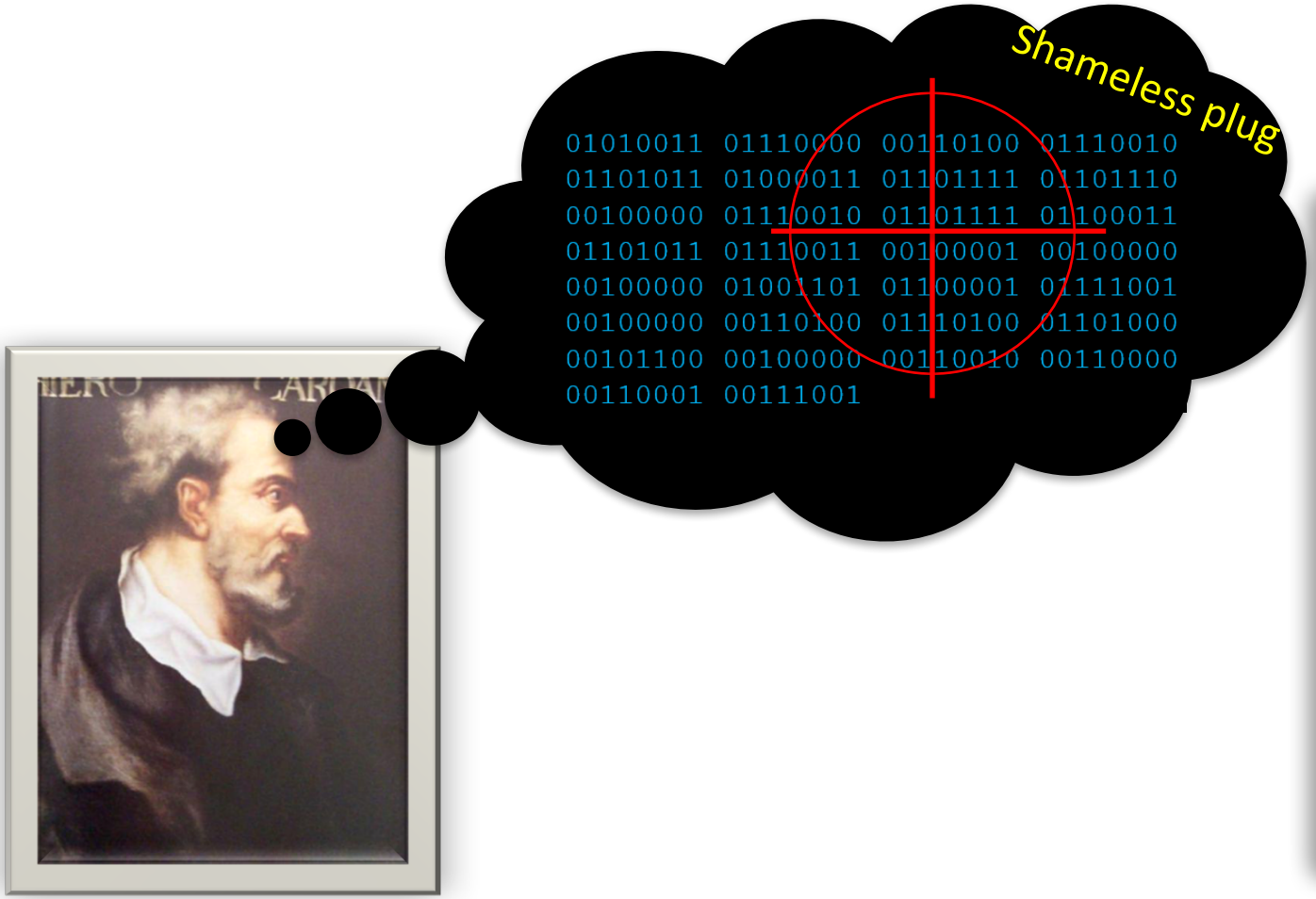
Unless you use them to to your advantage…

Begin caffeine intake now!



PROBABILITY THEORY

Gerolamo Cardano knew we'd be threat hunting someday!



"The mathematical theory of probability has its roots in attempts to analyze games of chance by Gerolamo Cardano in the sixteenth century"

-wikipedia

THE LAW OF LARGE NUMBERS

In probability theory…

“The law of large numbers is a theorem that describes the result of performing the same experiment a large number of times” - wikipedia

THE LAW OF LARGE NUMBERS

According to the law…

The average of the results obtained from a large number of trials should be close to the expected value, and will tend to become closer as more trials are performed.

TWEAKING THE LAW OF LARGE NUMBERS

Let's change it slightly for our purposes…

The average of the results obtained from a large number of ~~trials~~ events should be close to the expected value (benign), and will tend to become closer as more ~~trials are performed~~ events are analyzed.

If we apply the LoLN to the average network, most events *should* be benign! Otherwise…

FOCUS ON RARE EVENTS

Here's the point: •

If we believe most events are benign, shouldn't we be looking for **rare** events?

FOCUS ON RARE EVENTS

- Most attacks introduce something new (rare) to the environment:
  - Domains
  - IPs
  - Hashes
  - Registry Values
  - Services
  - Scheduled jobs

FOCUS ON RARE EVENTS

Early childhood threat hunting course

Identifying anomalies immediately without knowing what to look for ahead of time

FOCUS ON RARE EVENTS

What if you could tell whenever a critical asset started acting differently than the rest?

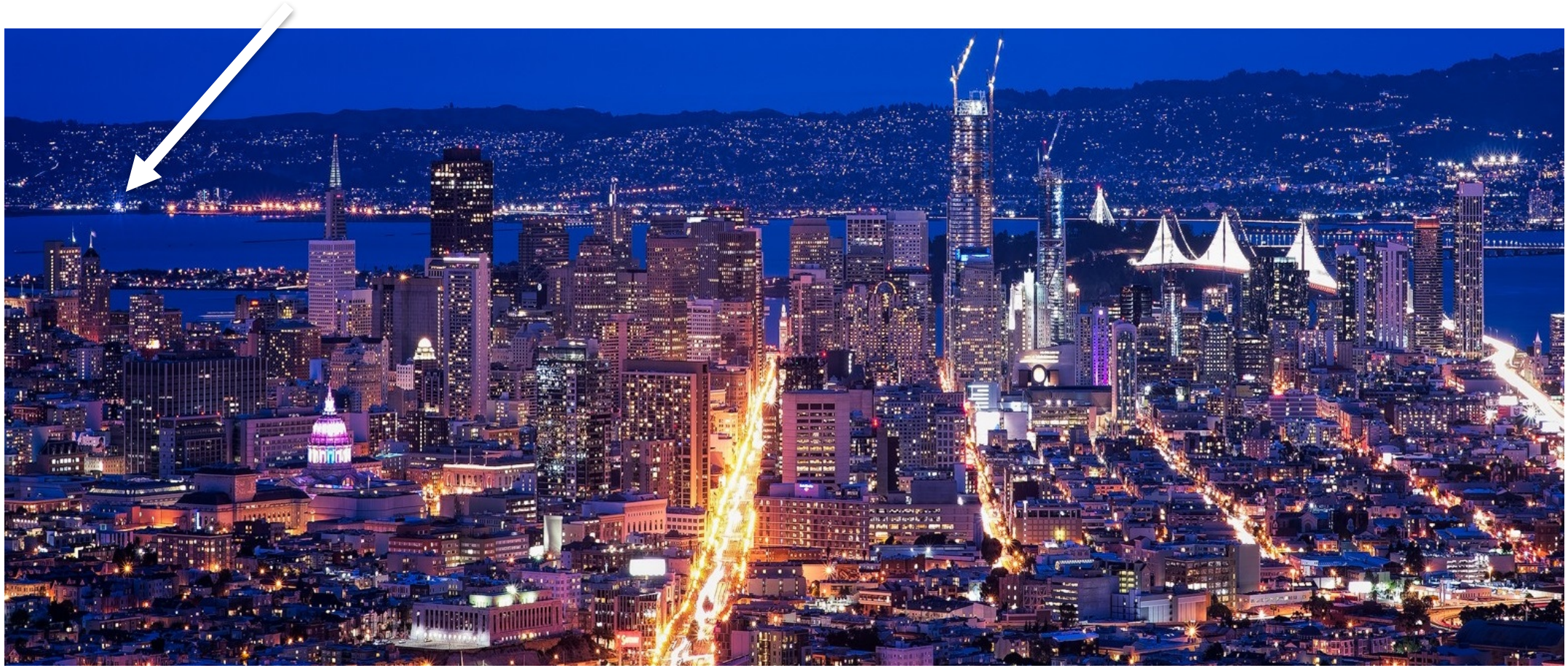
What if 1 out of 200 Domain Controllers started:

- Using a new network protocol?
- Talking to a new network segment?
- Running a new process?

ATMs, POS devices, mail servers, web servers, DNS servers, etc.

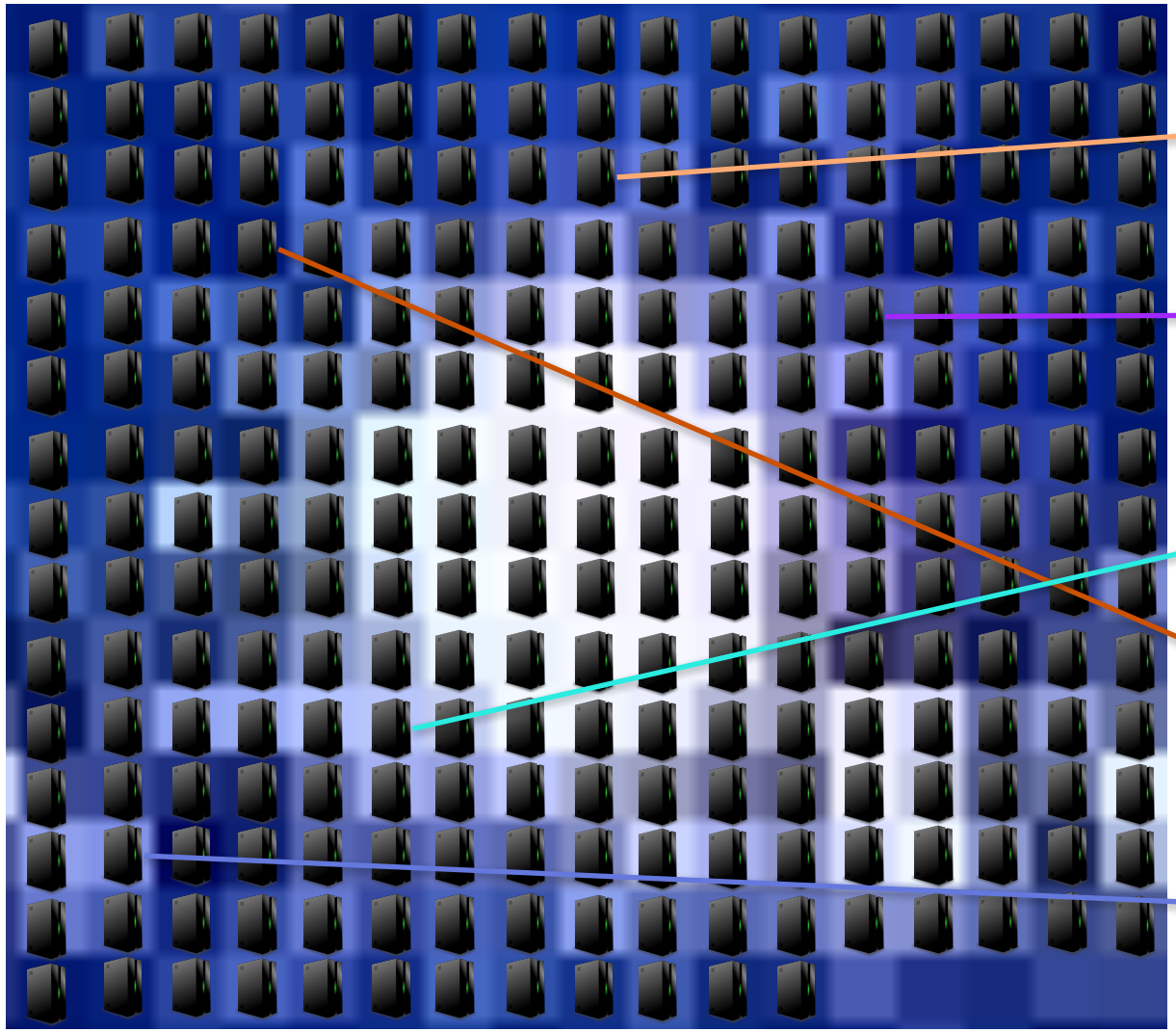
FOCUS ON RARE EVENTS

Remember the glowing light



FOCUS ON RARE EVENTS

Which one of these DCs is not like the others?



HTTP to 10.x.x.x:4444

ZIP File leaving via SMB

Large upload to DropBox

1.exe copied to the server

RDP from a dev workstation

RSAConference2019

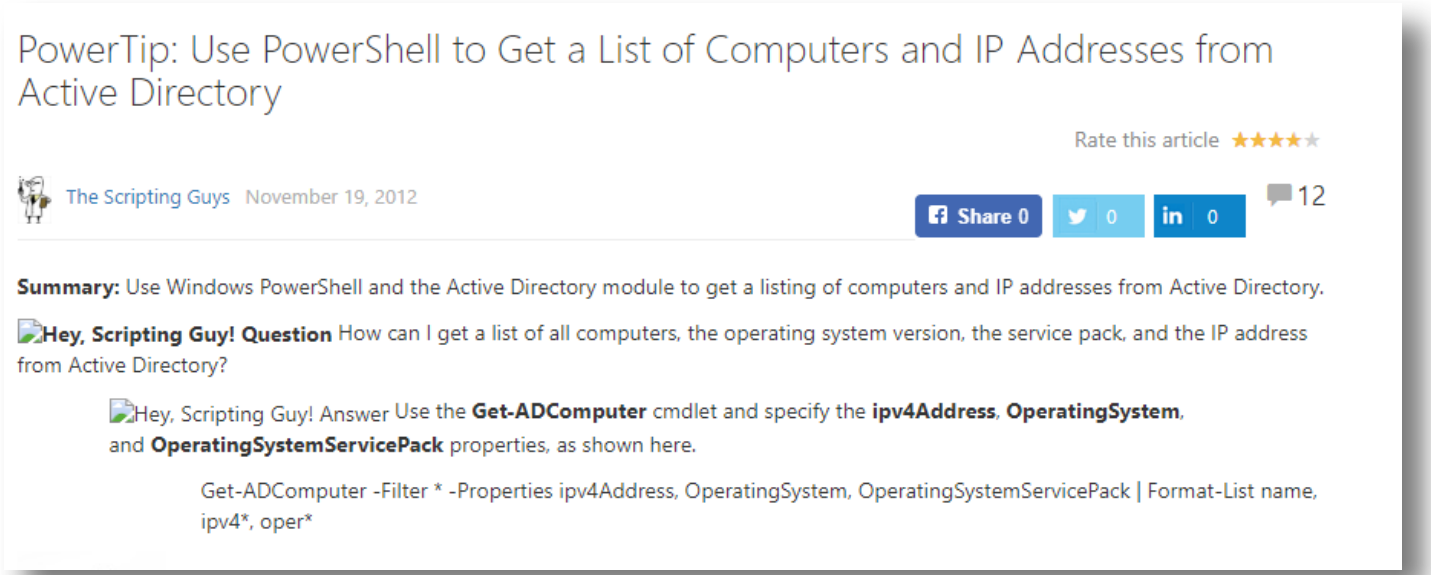
How to Find Rare Events

REQUIREMENTS

- You probably have one or more of these:
  - Proxy logs
  - Full packets
  - Netflow
  - Bro logs
  - Centralized Endpoint logs (or a way to query them in mass)
    - Registry values
    - Scheduled tasks
    - Security event logs
    - Running processes, etc.
  - **Any logs that record connections or running processes!**

REQUIREMENTS

- Asset Tagging (Describing the host type for each IP)
  - Tag as many as possible, focusing on critical assets
  - Sources: Active Directory, DNS, asset management solutions, internal wikis and databases, vulnerability management solutions, etc.
  - How to scrape AD for hostnames and IPs:
  - Nltest /dclist:<domain>
  - Powershell script:
    - https://bit.ly/2Kjeyqc



PowerTip: Use PowerShell to Get a List of Computers and IP Addresses from Active Directory

Rate this article ★★★★☆

The Scripting Guys November 19, 2012

Share 0 | 0 | in 0 | 12

Summary: Use Windows PowerShell and the Active Directory module to get a listing of computers and IP addresses from Active Directory.

Hey, Scripting Guy! Question How can I get a list of all computers, the operating system version, the service pack, and the IP address from Active Directory?

Hey, Scripting Guy! Answer Use the **Get-ADComputer** cmdlet and specify the **ipv4Address**, **OperatingSystem**, and **OperatingSystemServicePack** properties, as shown here.

```
Get-ADComputer -Filter * -Properties ipv4Address, OperatingSystem, OperatingSystemServicePack | Format-List name, ipv4*, oper*
```

HOW TO FIND RARE EVENTS

Let’s focus on using network traffic logs / packets

CORRELATING ASSET LISTS AND LOGS

- Asset tagging is extremely valuable!

Basic correlation: Join asset and log tables by IP fields

List of Asset IPs

A	B
10.10.10.3	Domain Controller
10.10.10.4	Domain Controller
10.10.10.5	Domain Controller
10.10.10.6	Domain Controller
10.10.10.7	Domain Controller
10.10.10.8	Domain Controller
10.10.10.9	Domain Controller
10.10.10.10	Domain Controller
10.10.10.11	Domain Controller
10.10.10.12	Domain Controller
10.10.10.13	Domain Controller
10.10.10.14	Domain Controller
192.168.5.12	Exchange Server
192.168.5.13	Exchange Server
192.168.5.14	Exchange Server
192.168.5.15	Exchange Server
192.168.5.16	Exchange Server
172.16.22.5	DNS Server
172.16.22.6	DNS Server
172.16.22.7	DNS Server
172.16.22.8	DNS Server
172.16.22.9	DNS Server
172.16.22.10	DNS Server

Network Logs

A	B	C	D
05/23/2018 16:41:03	10.10.10.3	10.10.10.7	SMB
05/23/2018 16:41:04	10.10.10.7	192.168.5.13	FTP
05/23/2018 16:41:05	192.168.5.15	10.10.10.14	RDP
05/23/2018 16:41:06	10.10.10.6	192.168.5.15	RDP
05/23/2018 16:41:07	10.10.10.7	192.168.5.13	SMB
05/23/2018 16:41:08	10.10.10.8	192.168.5.14	SMB
05/23/2018 16:41:09	172.16.22.10	10.4.5.44	SSH
05/23/2018 16:41:10	10.10.10.3	8.8.8.8	DNS
05/23/2018 16:41:11	172.16.22.8	10.10.10.7	SMB
05/23/2018 16:41:12	10.10.10.7	192.168.5.15	SMB
05/23/2018 16:41:13	10.10.10.7	10.10.10.6	FTP
05/23/2018 16:41:14	10.10.10.14	192.168.5.14	SSH
05/23/2018 16:41:15	192.168.5.15	192.168.5.14	SSH
05/23/2018 16:41:16	192.168.5.13	192.168.5.14	SSH
05/23/2018 16:41:17	192.168.5.14	10.10.10.14	HTTP
05/23/2018 16:41:18	192.168.5.15	172.16.22.6	HTTP
05/23/2018 16:41:19	192.168.5.16	192.168.5.13	HTTP
05/23/2018 16:41:20	172.16.22.5	192.168.5.13	HTTP
05/23/2018 16:41:21	172.16.22.6	192.168.5.14	HTTP
05/23/2018 16:41:22	172.16.22.6	10.4.5.44	SMB
05/23/2018 16:41:23	172.16.22.8	10.10.10.7	SMB
05/23/2018 16:41:24	172.16.22.9	10.10.10.14	SMTP
05/23/2018 16:41:25	172.16.22.7	172.16.22.6	FTP

CORRELATION QUERY EXAMPLE

- Query results show protocols used by each asset type

Date	Asset_Type	Source	Protocol
05/23/2018 16:41:03	Domain Controller	10.10.10.3	SMB
05/23/2018 16:41:04	Domain Controller	10.10.10.7	FTP
05/23/2018 16:41:05	Exchange Server	192.168.5.15	RDP
05/23/2018 16:41:06	Domain Controller	10.10.10.6	RDP
05/23/2018 16:41:07	Domain Controller	10.10.10.7	SMB
05/23/2018 16:41:08	Domain Controller	10.10.10.8	SMB
05/23/2018 16:41:09	DNS Server	172.16.22.10	SSH
05/23/2018 16:41:10	Domain Controller	10.10.10.3	DNS
05/23/2018 16:41:11	DNS Server	172.16.22.8	SMB
05/23/2018 16:41:12	Domain Controller	10.10.10.7	SMB
05/23/2018 16:41:13	Domain Controller	10.10.10.7	FTP
05/23/2018 16:41:14	Domain Controller	10.10.10.14	SSH
05/23/2018 16:41:15	Exchange Server	192.168.5.15	SSH
05/23/2018 16:41:16	Exchange Server	192.168.5.13	SSH
05/23/2018 16:41:17	Exchange Server	192.168.5.14	HTTP
05/23/2018 16:41:18	Exchange Server	192.168.5.15	HTTP
05/23/2018 16:41:19	Exchange Server	192.168.5.16	HTTP
05/23/2018 16:41:20	DNS Server	172.16.22.5	HTTP
05/23/2018 16:41:21	DNS Server	172.16.22.6	HTTP
05/23/2018 16:41:22	DNS Server	172.16.22.6	SMB
05/23/2018 16:41:23	DNS Server	172.16.22.8	SMB
05/23/2018 16:41:24	DNS Server	172.16.22.9	SMTP

BRO USERS

If you use Zeek (Bro):

- SQLite filter option beginning in version 2.2

```
event bro_init()
    {
    local filter: Log::Filter =
        [
        $name="sqlite",
        $path="/var/db/conn",
        $config=table(["tablename"] = "conn"),
        $writer=Log::WRITER_SQLITE
        ];

     Log::add_filter(Conn::LOG, filter);
    }
```

https://docs.zeek.org/en/stable/frameworks/logging-input-sqlite.html

HOW TO FIND RARE EVENTS

- Multiple ways to find rare events. Here are three that work!
  1. Define a bad behavior, find outliers exhibiting that behavior
  2. Tag assets by type, find outliers by specific types of artifacts (protocol, IP, network, files, running processes, etc.)
  3. Tag assets by type, define many characteristics, find outliers

The last two require asset tagging, which is WELL worth the effort!

HOW TO FIND RARE EVENTS

Let's start with this one:

Define a behavior, find outliers exhibiting that behavior

(Doable without asset tagging)

HOW TO FIND RARE EVENTS

- Table named 'proxy_logs' with millions of rows:

ip.src	domain	method
192.168.21.54	taboola.com	POST
10.4.22.24	facebook.com	POST
10.4.22.24	facebook.com	GET
172.16.23.187	ff5ee.com	POST
10.42.5.77	disorderstatus.ru	POST

HOW TO FIND RARE EVENTS

Utilize the **COUNT**, **DISTINCT**, and **GROUP BY** functions of SQL

```
SELECT domain, COUNT (DISTINCT ip.src) FROM proxy_logs
WHERE <insert bad behavior query>
GROUP BY domain
ORDER BY COUNT (DISTINCT ip.src) DSC;
```

Example “bad behavior” criteria for basic C2 hunting:

- HTTP POST without a GET
- No referer
- Content type = application / octet-stream
- No cookies

HOW TO FIND RARE EVENTS

- The unicorns start appearing! Benign domains are easier to ignore.

domain	COUNT (DISTINCT ip.src)
vertamedia.com	6095
rubiconproject.com	5730
cat.sv.us.criteo.com	5380
taboola.com	4507
adsnxs.com	2033
disorderstatus.ru	5 
ff5ee.com	3 

HOW TO FIND RARE EVENTS

Finding new unicorns for a given behavior

C2 Domains - POST

domain	COUNT (DISTINCT ip.src)
vertamedia.com	6095
rubiconproject.com	5730
cat.sv.us.criteo.com	5380
taboola.com	4507
adsnxs.com	2033
disorderstatus.ru	5 🦄
ff5ee.com	3 🦄

Append Rare Values

previous-c2-domains

domain
disorderstatus.ru
ff5ee.com

Isolate new values NOT in the historical list

method = “post” AND referer IS NULL etc. etc. AND WHERE domain NOT IN (SELECT domain FROM previous-c2-domains)

New potential C2 domains are more likely to be spotted now!

HOW TO FIND RARE EVENTS

Now this one:

Tag assets by type, find outliers by specific types of artifacts

(protocol, IP, network, files, running processes, etc.)

Let’s pick on protocols!

HOW TO FIND RARE EVENTS

```
SELECT asset.type, protocol, COUNT (DISTINCT ip.src)
FROM netflow_log
WHERE asset.type='us domain controller'
ORDER BY COUNT (DISTINCT ip.src) DSC;
```

Asset.Type	Protocol	COUNT (DISTINCT ip.src)
US DOMAIN CONTROLLER	RPC	150
US DOMAIN CONTROLLER	KERBEROS	150
US DOMAIN CONTROLLER	LDAP	150
US DOMAIN CONTROLLER	HTTP	134
US DOMAIN CONTROLLER	SSL	122
US DOMAIN CONTROLLER	SMB	2
US DOMAIN CONTROLLER	DNS	1

HOW TO FIND RARE EVENTS

Taking it to the next level!

Instead of you defining the specific behavior or picking the right artifact, have the behaviors reveal themselves to you!

<insert spooky music here>

HOW TO FIND RARE EVENTS

Let’s work with full packets now…

HOW TO FIND RARE EVENTS

Define Traffic Characteristics

Examples:

- HTTP Content type
- File types (by extension and header)
- Byte count ranges (0-100K, 100K-1MB, etc.)
- Directionality (inbound, outbound, lateral)
- SSL self-signed certs
- Payload entropy levels
- HTTP lacking a specific header value
- HTTP direct to IP request
- Proxy blocked traffic
- Single sided TCP
- Traffic over non-standard ports
- Destination network descriptions
- Transmit payload byte size
- Receive payload byte size
- IP address exists

Store these in a field named ‘traffic_characteristics’

^^^ this is a dummy rule to give you total counts per asset group

HOW TO FIND RARE EVENTS

```
SELECT asset.type, traffic_characteristics, COUNT
(DISTINCT ip.src) FROM packet_log

WHERE asset.type='us domain controller'

ORDER BY COUNT (DISTINCT ip.src) DSC
```

Asset.Type	traffic_characteristics	COUNT (DISTINCT ip.src)
US DOMAIN CONTROLLER	IP address exists	150
US DOMAIN CONTROLLER	medium ratio payload	150
US DOMAIN CONTROLLER	session size 100k-1000k	143
US DOMAIN CONTROLLER	session size 1K-100K	54
US DOMAIN CONTROLLER	http direct to ip request	2
US DOMAIN CONTROLLER	SSL self-signed cert	2
US DOMAIN CONTROLLER	host header contains port	2

https://X.X.X.X:8080/evil.php

HOW TO FIND RARE EVENTS

SELECT asset.type, traffic_characteristics, COUNT (DISTINCT ip.src) from packet_log ~~WHERE asset.type='us domain controller'~~ ORDER BY asset.type, COUNT (DISTINCT ip.src) DSC

Rare events are now easily identified across every type of asset you can tag!

The key is to define a multitude of behaviors using the logs and artifacts available to you.

Asset.Type	traffic_characteristics	COUNT (DISTINCT ip.src)	
US DOMAIN CONTROLLER	IP address exists	150	
US DOMAIN CONTROLLER	medium ratio payload	150	
US DOMAIN CONTROLLER	session size 100k-1000k	143	
US DOMAIN CONTROLLER	session size 1K-100K	54	
US DOMAIN CONTROLLER	http direct to ip request	2	
US DOMAIN CONTROLLER	SSL self-signed cert	2	
US DOMAIN CONTROLLER	host header contains port	2	
EXCHANGE SERVER	IP address exists	23	
EXCHANGE SERVER	session size 100k-1000k	22	
EXCHANGE SERVER	session size 1MB-5MB	18	
EXCHANGE SERVER	RDP from non-IT subnet	3	
EXCHANGE SERVER	EXE file copied to asset	1	
ATM	IP address exists	4500	
ATM	session size 100k-1000k	4013	
ATM	session size 1K-100K	3554	
ATM	IRC protocol	8	
ATM	zip file copied from asset	1	
DNS SERVER	IP address exists	72	
DNS SERVER	session size 100k-1000k	72	
DNS SERVER	session size 1K-100K	56	
DNS SERVER	HTTP protocol	1	
POS CONTROLLER	IP address exists	3751	
POS CONTROLLER	session size 100k-1000k	3236	
POS CONTROLLER	session size 1K-100K	2754	
POS CONTROLLER	proxy server destination	2	
POS CONTROLLER	http direct to ip request	2	

HOW TO FIND RARE EVENTS

Yes, it really does work!

protocol	asset.type	count (distinct ip.src)
FTP	<redacted>	2
NETBIOS		3
SSL		5
DNS		5
138		6
NTP		13
RPC		45662
HTTP		45789
389		47052
88		47206
SMB		47315
OTHER		47316

HOW TO FIND RARE EVENTS

Yes, it really does work!

asset.type	traffic_characteristics	count (distinct ip.src)
<redacted>	unknown service over http port	1
	ssl certificate missing issuer organizational name	1
	http post missing content-type	597
	named pipe	671
	http1.1 without accept header	815
	host header contains port	846
	http1.1 without referer header	847
	smb session on non-smb port	847
	rpc over non-standard port	849
	hostname consecutive consonants	851
		851

HOW TO FIND RARE EVENTS

Example findings from hunting for rare behaviors

- Rare autoruns entry revealed new, custom malware
- Rare running process analysis reveals packet capture utility running on 1 out of 400 critical assets of one type
- C2 activity from malware

APPLICATION

Applies to many types of artifacts, such as:

- Internal country source / destinations pairs
- Asset type source / destinations pairs
- Registry autorun values
- Running processes and services
- Listening ports
- Hash values
- Scheduled Tasks / cron jobs

APPLICATION

Query Timeframes

- Too short, things may look rare when they aren't
- Too long, you may not see an event in a timely manner
- 24 hours is a good starting point, experiment from there

CHALLENGES

Benign contributors to rare behaviors

- Misconfigured machines (someone plugged a dummy IP into software)
- Troubleshooting activities (left an odd executable running)

Automated follow-up queries can enrich results to help you determine:

- If a rare behavior the next day belongs to the same machine
- If several rare behaviors belong to the same machine
- If the rare artifact is benign or needs further review

TIP: Start small! Start with Domain Controllers, expand from there.

APPLY

1. Identify assets in your environment (automate!):
   - Domain Controllers
   - Mail servers
   - Critical systems (POS, ATMs, SCADA, R&D, etc.)
   - IoT
2. Join the asset description table with your event logs (packets, netflow, endpoint, etc.) by IP address
3. Craft queries to group events by asset type and count unique source (or dest) IPs for given sets of behaviors

YOU CAN DO THIS!

HOW TO FIND RARE EVENTS

Happy Hunting!

Contact:

vernon.habersetzer@walmart.com

 @HuntingNomad