Looking for Needles in Needlestacks

with the Threat Hunting Toolkit

About Us

- Derek and Ethan
- Threat Hunters with Black Hills Info Sec
- More at the end if you're still here

BLACK HILLS
Information Security
• 2008 •

Roadmap

- What is Threat Hunting
- Types of Data Sources
- Example Hunt for C2

What is Threat Hunting Anyway?

> **Proactive** approach to identifying threats
> - Josh Liburdi, BroCon 2015

> **Human-driven**, **proactive** and iterative search through networks, endpoints, or datasets in order to detect malicious, suspicious, or risky **activities that have evaded detection** by existing automated tools.
> - Hunt Evil: Your Practical Guide to Threat Hunting, Sqrrl

> **Human-centric** process of **proactively** searching through networks for evidence of attacks that **evade existing security monitoring** tools.
> - Chris Sanders, Practical Threat Hunting

What is Threat Hunting Anyway?

> **Act** of tracking and eliminating cyber adversaries from your network **as early as possible**.
> - Dr. Eric Cole, 2017

> Threat hunting is just the new term for "farting around on the network"
> - Anonymous

What is Threat Hunting Anyway?

Common Themes

- It's proactive
- Taking a large amount of data and finding a subset
- Find what existing protections miss
- Investigate the weird (anomalous does not necessarily mean bad)

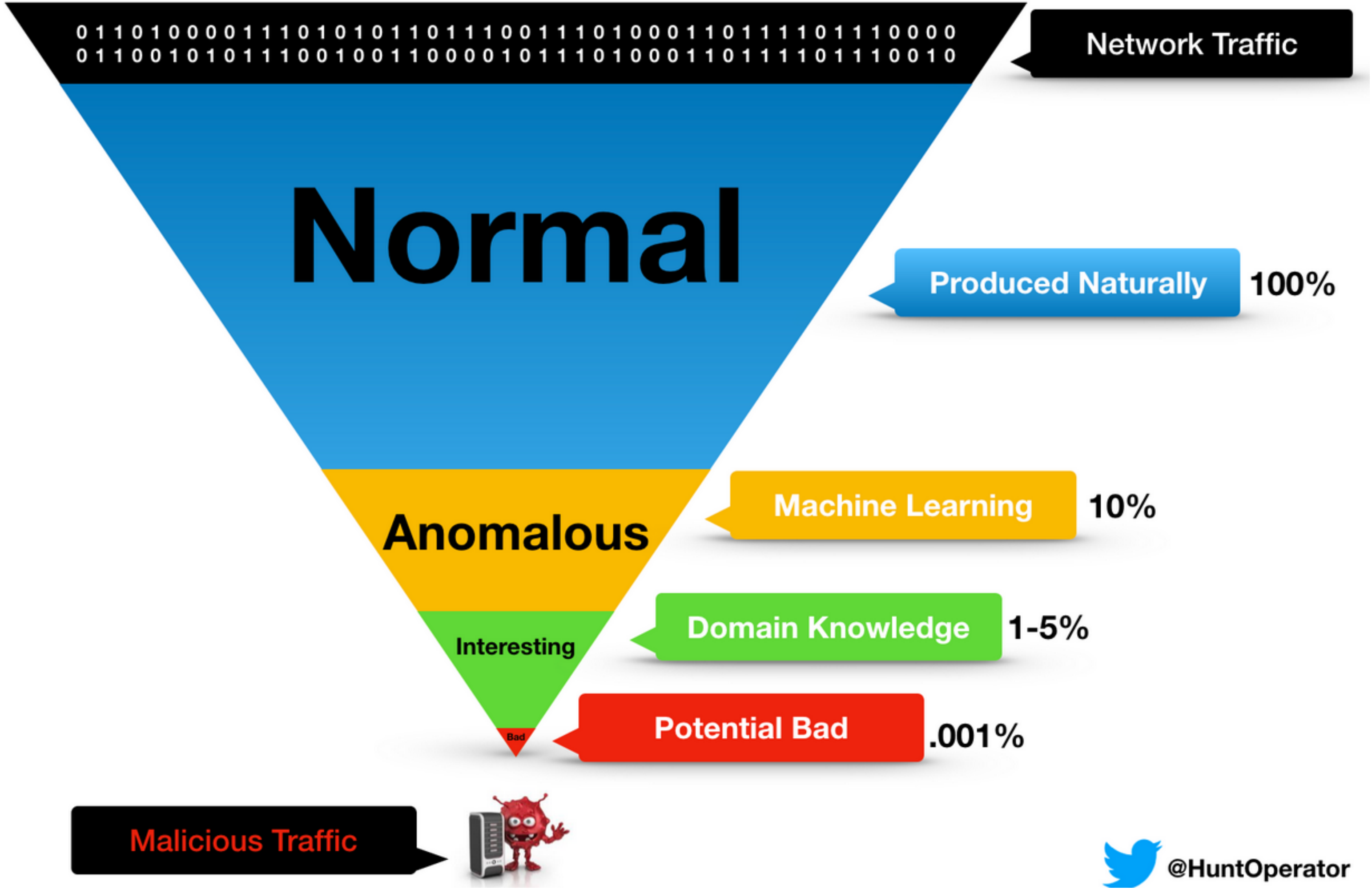


Image Credit: Austin Taylor

Modern Threat Hunting Challenges

- Ever growing traffic volume
- Log correlation
- More sophisticated attackers
  - CDNs
  - Fileless malware
- Remote workforce
  - VPNs
  - Split tunnels

- Encrypted traffic
  - TLS 1.3
  - Encrypted Server Name Indicator (ESNI)
  - DNS over TLS/HTTPS
- Cloud
  - Containers
  - Serverless

Hypothesis Based Threat Hunting

- Attempting to prove or disprove a question of interest
- Data + Technique
    - Stack counting
    - Anomaly detection, outlier discovery
    - Set theory
    - Beaconing

Data Sources

Big Three

- Host
- Network
- Active Directory (Azure)

Other Types

- Appliance logs (Proxies)
- Firewall logs
- Cloud resource logs
- Application logs (Web Servers, etc)
- Intrusion Detection System

Data Sources

Host Logs

Examples

- Process execution
- Network connection
- Login attempt

Sources

- Sysmon
- Osquery
- Wazuh
- Elastic Agent
- OpenEDR

Data Sources

Host Logs

Pros

- Increased visibility

Cons

- Can be difficult to deploy
- Compromised host can hide
- No de facto standard
- No IoT

Data Sources

Active Directory Logs

Examples

- Authentication attempts
- Process logging
- Powershell script block

Sources

- Windows events
- Azure AD (may require extra $$)

Data Sources

Active Directory Logs

Pros

- Holisic picture of Windows environment

Cons

- Windows only; missing Linux, OSX, IoT
- Not originally designed for security

Data Sources

Network Logs

Examples

- IP network flows (layers 3 & 4)
- DNS queries
- Protocols
- Amount of data transferred

Sources

- Zeek
- Netflow
- Tcpdump
- Proxy/firewall

Data Sources

Network Logs

Pros

- Very difficult to hide from
- May be easier to deploy
- Mature open source and free options

Cons

- Little visibility for encrypted traffic
- Hardware & storage costs
- Limited/immature support for cloud, PaaS, and containers

Network Log Sources

Netflow

Pros

- Already supported by existing network devices
- Tooling is mature
- Small storage cost

Cons

- Every vendor has their own nuanced implementation
- Not designed for security

Network Log Sources

Full Packet Capture

Pros

- All content passess over the wire
- See everything, know everything

Cons

- High storage requirements
- High disk I/O requirements
- Time consuming to search

Network Log Sources

Zeek (formerly Bro)

Pros

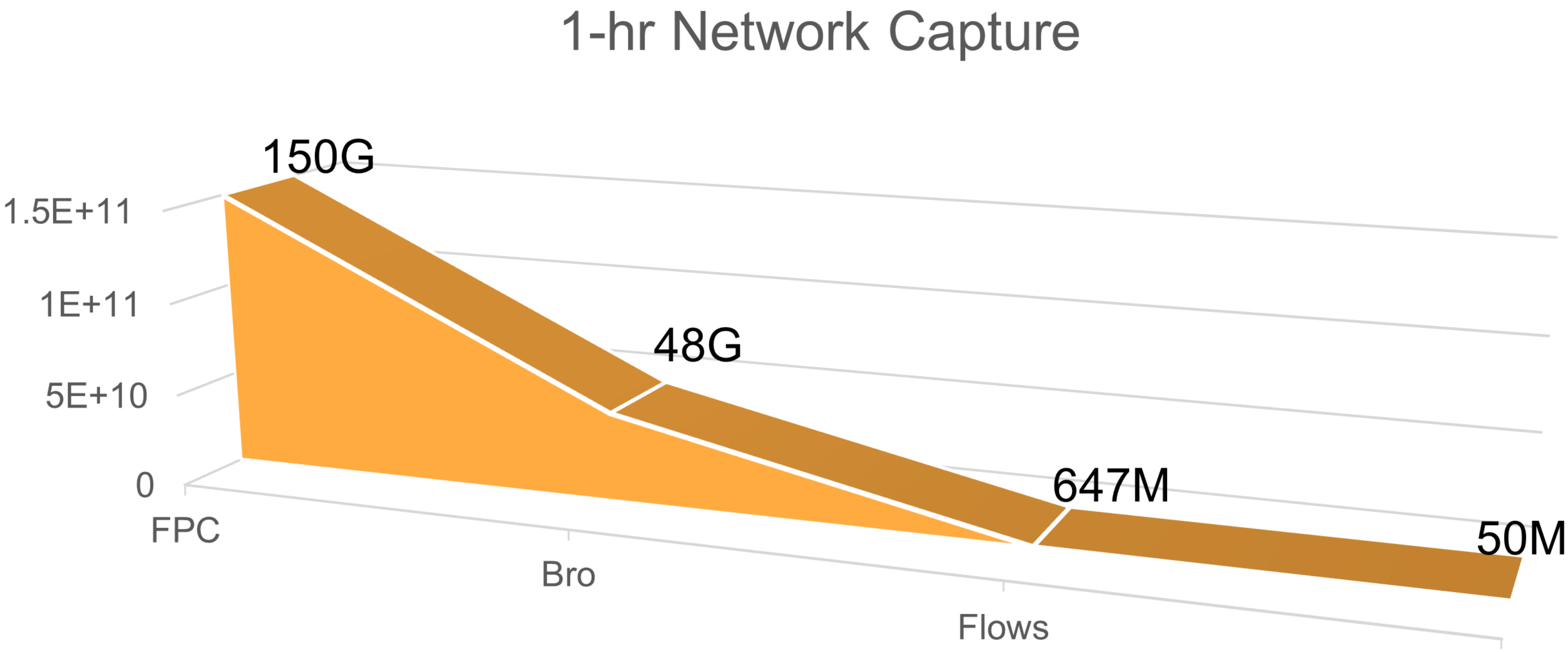
- Records interesting metadata
- Extensible
- Stiches flows (unidirectional) into connection events (bidirectional)

Cons

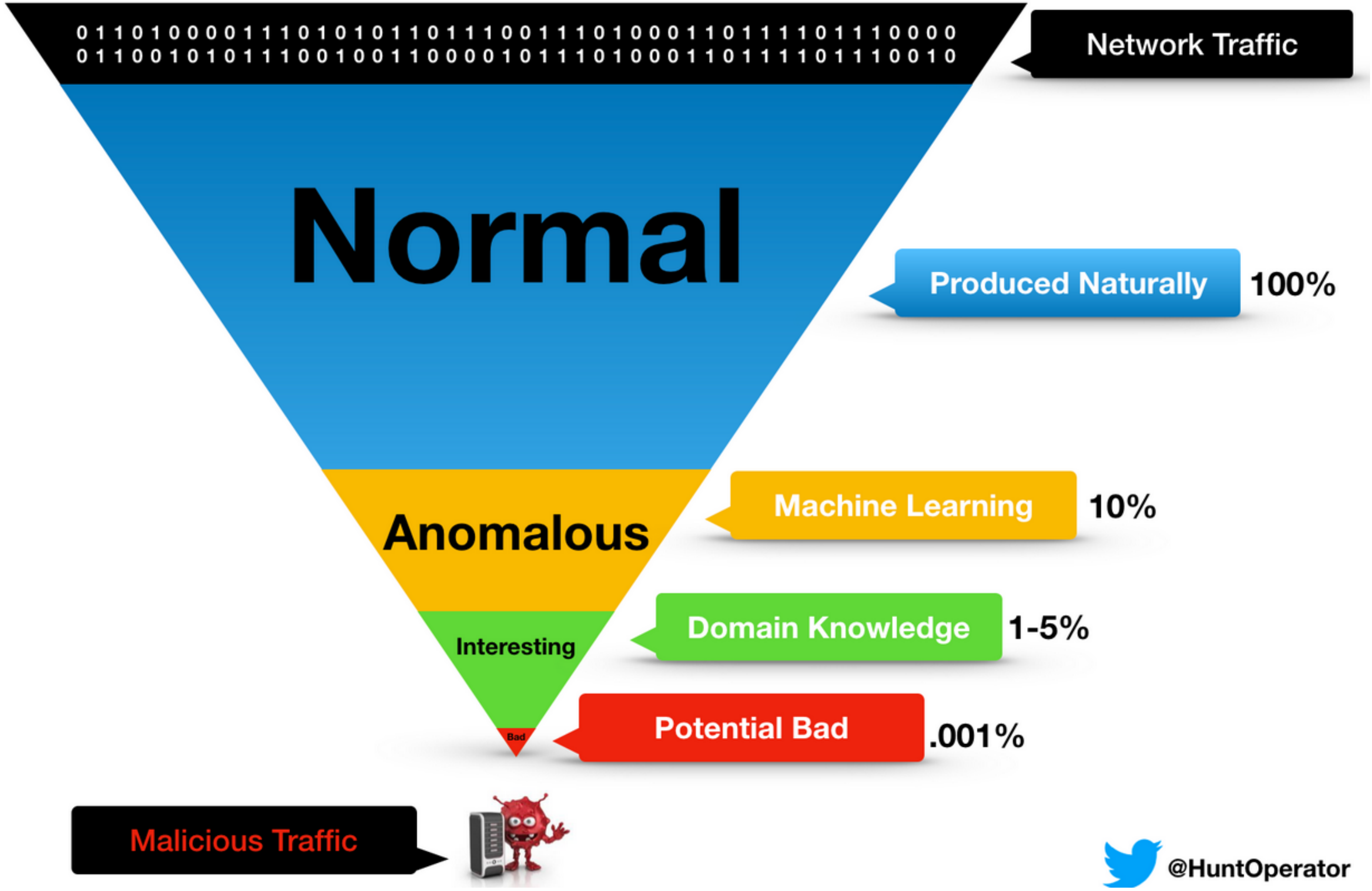
- Requires separate capture system
- Installation and tuning at scale can be difficult

Network Log Sources

Which source is right for you?



You have data. Now what?



Introducing...

Threat Hunting Toolkit (THT)

- One toolkit for many text log sources
- Consistent environment
- Easy installation

https://github.com/ethack/tht

Threat Hunter Toolkit

Install

```
sudo curl -o /usr/local/bin/tht \
  https://raw.githubusercontent.com/ethack/tht/main/tht
sudo chmod +x /usr/local/bin/tht
```

Start

```
tht
```

Use

```
root@zeek /host/opt/zeek/logs
$ filter --dns google.com | chop query | domain 3 | mfo
```

Example Hunt

Hypothesis

- There is command and control (C2) on our network.

Assumptions

- How can attackers hide?
- Content Delivery Networks (CDNs)
- Let's start by looking at CloudFront.

Background

- What does "normal" CloudFront traffic look like?
    - SSL/TLS to a subdomain of *cloudfront.net*.
    - Subdomain is a random string, such as *dko9feizeit4mi.cloudfront.net*.
    - Subdomains are not shared between CloudFront customers.

Possible Anomalies

Anomaly		Data Source
Newly observed CloudFront domain	⇒	`dns.log`, `ssl.log`
Abnormal traffic volume to CloudFront	⇒	`conn.log`, `ssl.log`
Rare JA3 hash	⇒	`ssl.log`

Newly Observed Domains

- How do we know a domain is new on our network?
  - Search through historical logs
  - Passive DNS

Have you seen this van in your network before?

Passive DNS

- Historical record of IP address and domain mappings
- First and last seen
- Count

https://github.com/JustinAzoff/bro-pdns

Passive DNS

```
$ pdns find individual example.com
+-------------+-------+-------+---------------------+-------------------
|    Value    | Which | Count |        First        |        Last
+-------------+-------+-------+---------------------+-------------------
| example.com | Q     |  4614 | 2021-06-18 17:02:56 | 2021-09-09 00:42:2
+-------------+-------+-------+---------------------+-------------------
```

```
$ pdns find tuples example.com
+-------------+------+-------------------+-------+-----+----------------
|    Query    | Type |      Answer       | Count | TTL |        First
+-------------+------+-------------------+-------+-----+----------------
| example.com | AAAA | 2606:2800:220:... |   590 |  84 | 2021-06-21 18:2
| example.com | A    | 93.184.216.34     |  3927 | 519 | 2021-06-18 17:0
+-------------+------+-------------------+-------+-----+----------------
```

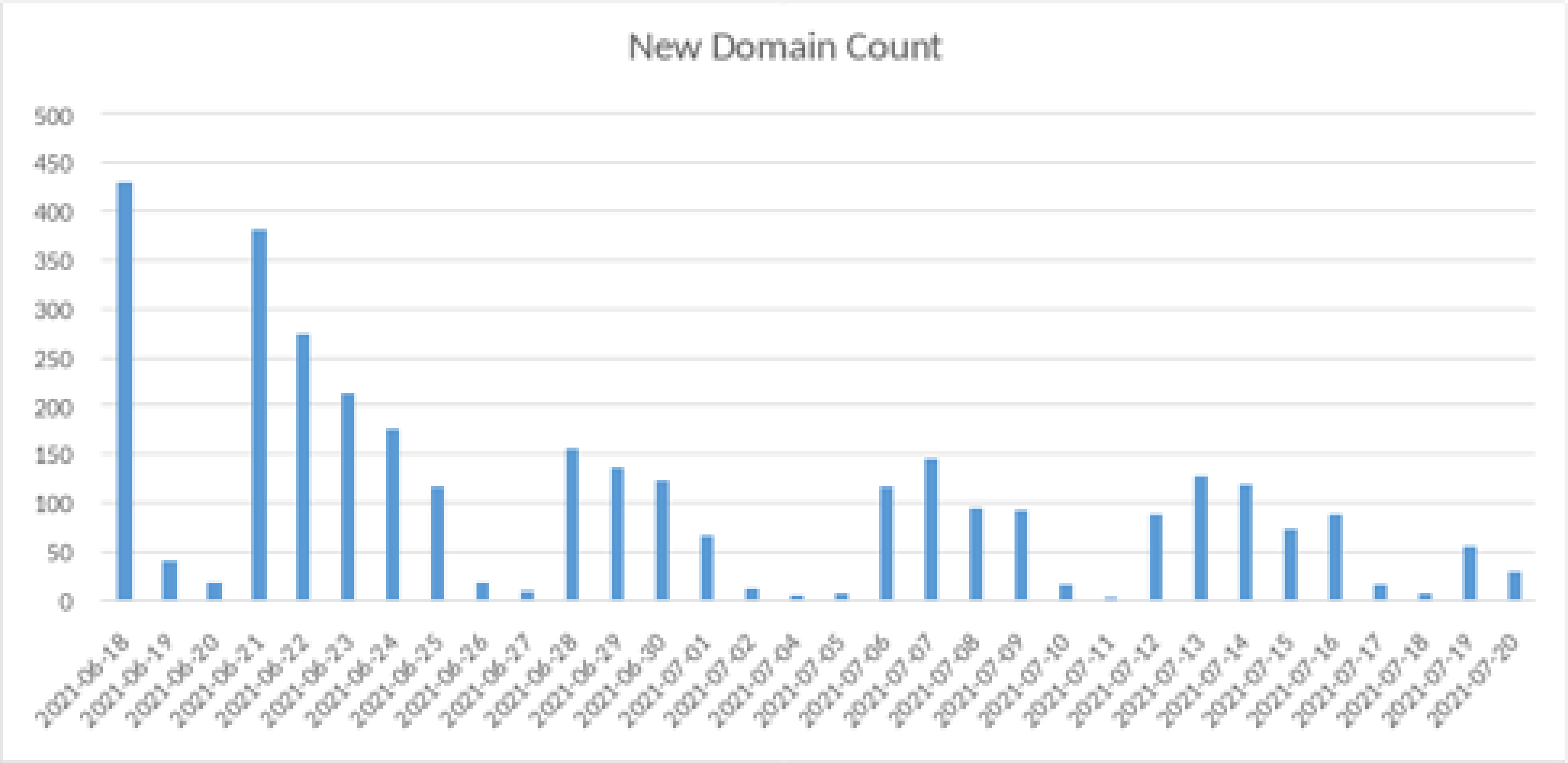
Newly Observed Domains

How many new CloudFront subdomains show up each day?

```
$ pdns like individual cloudfront.net |
  chop First | chop 1 | freq

    145 2021-07-07
     95 2021-07-08
     93 2021-07-09
     16 2021-07-10
      3 2021-07-11
     89 2021-07-12
    129 2021-07-13
    120 2021-07-14
```

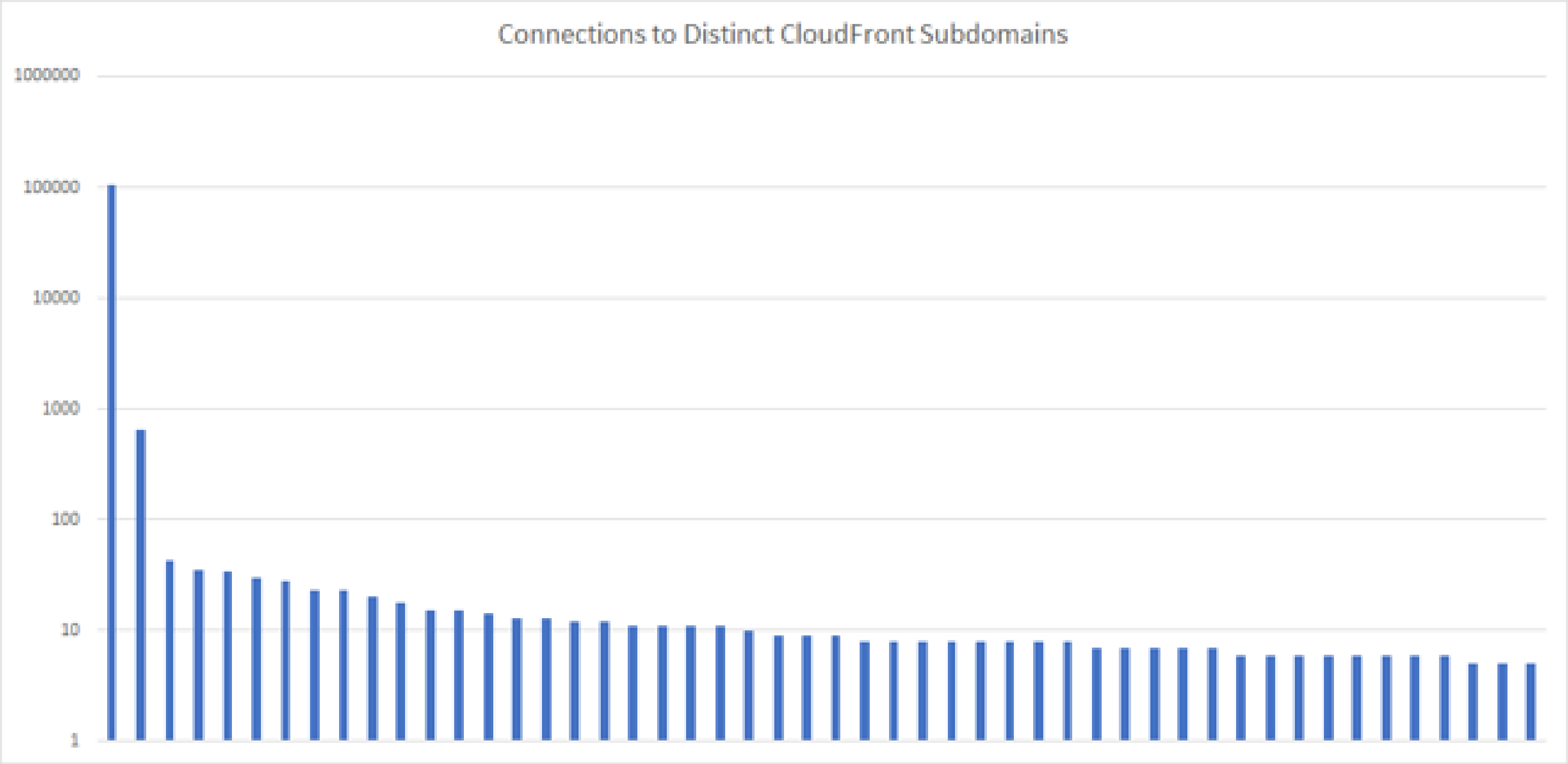
Newly Observed Domains



Abnormal Traffic Volume

```
$ filter --ssl cloudfront.net | chop server_name | mfo 5
 104132 dohshe7fai3sei.cloudfront.net
    657 dquaetheephae9.cloudfront.net
     43 dko9feizeit4mi.cloudfront.net
     35 diu3iethangeet.cloudfront.net
     34 diesh7hiegh4fo.cloudfront.net
```

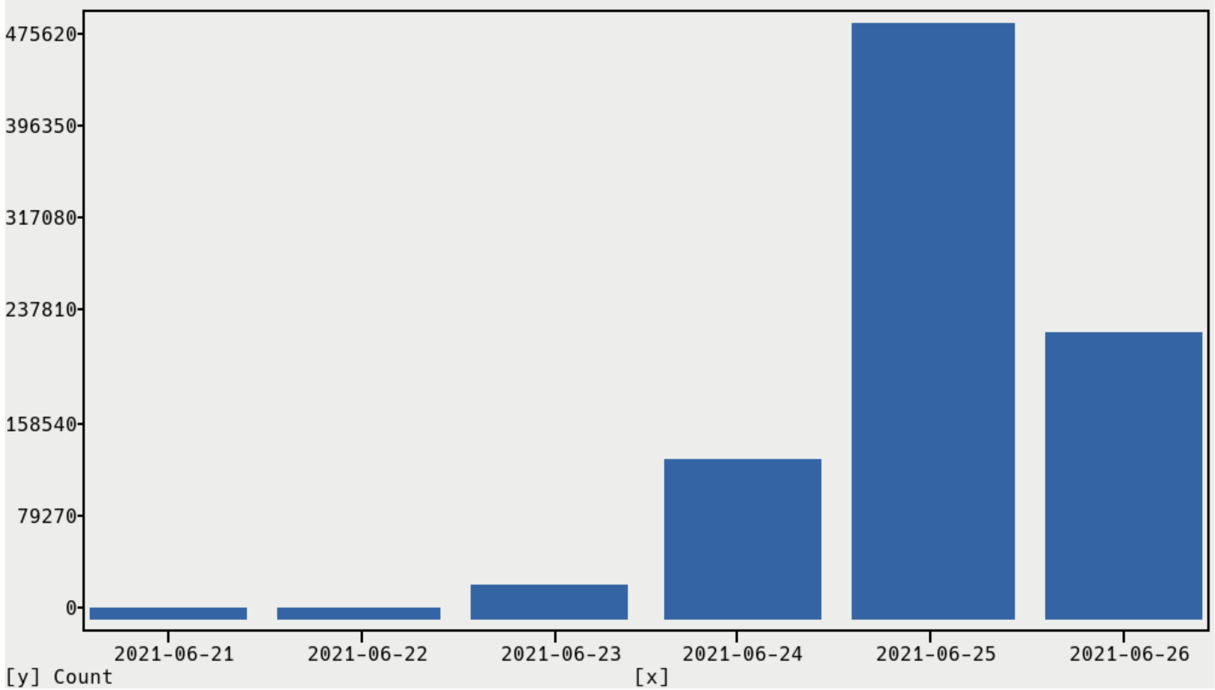
Abnormal Traffic Volume



Abnormal Traffic Volume

When did this abnormal traffic pattern start?

```
$ filter --ssl cloudfront.net | chop ts | ts2 | freq | plot-bar
```



Abnormal Traffic Volume

Now that we have an anomaly it is useful to know when it was first seen.

```
$ pdns find individual dohshe7fai3sei.cloudfront.net | chop Value First

+-------------------------------+---------------------+-----------------
|             Value             |        First        |        Last
+-------------------------------+---------------------+-----------------
| dohshe7fai3sei.cloudfront.net | 2021-06-22 21:33:55 | 2021-06-25 19:13
+-------------------------------+---------------------+-----------------
```

Abnormal Traffic Volume

Which sources were communicating with the domain?

```
$ filter --ssl dohshe7fai3sei.cloudfront.net | chop id.orig_h | distinct
192.168.2.20
192.168.2.49
192.168.2.127
192.168.3.20
```

These four systems are now our suspects.

Sidebar: Cheatsheet

Command	Purpose	Alternative
`filter`	**search** within files	`find | grep`
`chop`	**select** columns	`cut` or `zeek-cut`
`freq`	**frequency** counts	`sort | uniq -c`
`mfo`	**most frequent occurrence**	`sort | uniq -c | sort -nr`
`distinct`	**unique** elements	`sort | uniq`
`countdistinct`	**cardinality**	`sort | uniq | wc -l`
`ts2`	convert **timestamps**	
`plot-bar`	bar **graph**	

JA3 Hash

- What is a JA3 hash?
  - Semi-unique fingerprint for an SSL/TLS client.
  - Similar to User-Agent string for HTTP traffic.
  - Derived from client's choice of parameters for an SSL connection.
  - There can be different clients with the same JA3, especially if they use the same underlying SSL library.

Pivot: JA3 Hash (1)

Find the hash used to contact the suspected domain.

```
$ filter --ssl dohshe7fai3sei.cloudfront.net | chop id.orig_h ja3 | mfo

  63299 192.168.2.49   258a5a1e95b8a911872bae9081526644
  14909 192.168.2.127  258a5a1e95b8a911872bae9081526644
  25921 192.168.3.20   258a5a1e95b8a911872bae9081526644
    133 192.168.2.20   258a5a1e95b8a911872bae9081526644
```

All source IPs found so far use the same JA3 hash: `258a5a1e95b8a911872bae9081526644`

Pivot: JA3 Hash (2)

Where else has this hash been used from? Is it rare?

```
$ filter --ssl 258a5a1e95b8a911872bae9081526644 | chop id.orig_h | count

84
```

Used by 84 other sources.

Pivot: JA3 Hash (3)

Where have clients been connecting to using this hash? Do we spot any patterns or outliers?

```
$ filter --ssl 258a5a1e95b8a911872bae9081526644 | chop server_name | dom

 104132 cloudfront.net
   7336 microsoft.com
   7305 live.com
   5326 office.com
   4128 office365.com
   1544 outlook.com
   1321 sharepoint.com
    929 go-mpulse.net
    706 windows.net
    526 office.net
```

CloudFront is first by a long shot. What do the others have in common?

Pivot: JA3 Hash (4)

Which CloudFront destinations has the hash been used?

Let's limit it to our suspect IPs in `ips.txt`.

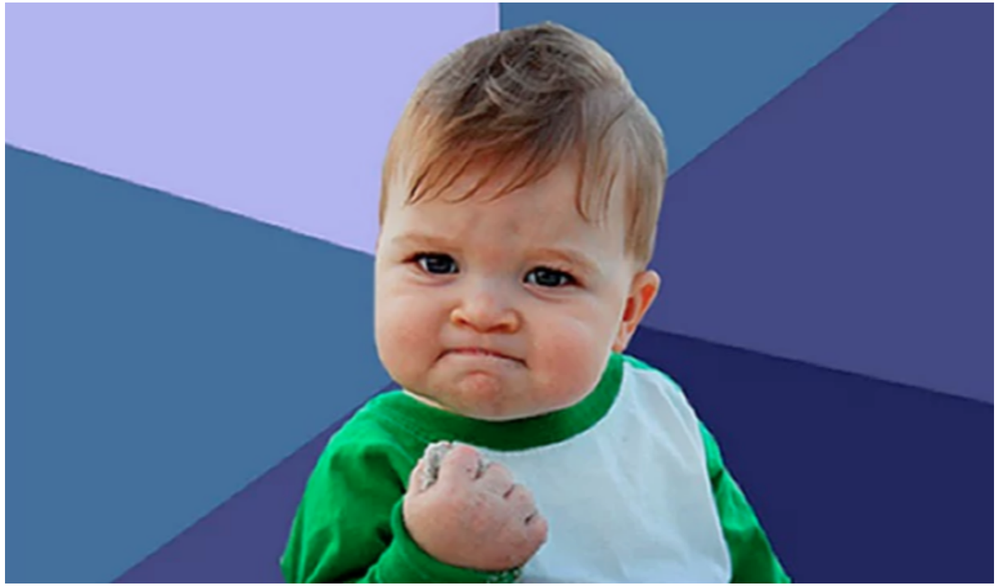
```
$ filter --ssl 258a5a1e95b8a911872bae9081526644 cloudfront.net | filter

 104132 dohshe7fai3sei.cloudfront.net
   1351 daid4aetheech4.cloudfront.net
      7 d280ht16bmiuo6.cloudfront.net
```

Two new CloudFront domains.

Conclusion

- Actual red team engagement
- Four source hosts found compromised
- Main C2 server: *dohshe7fai3sei.cloudfront.net*
- Remaining CloudFront domains were long haul / backup C2



Problems

- Heavy CloudFront usage
- TLS 1.3 with encrypted SNI
- DNS over HTTPS / TLS

References

- Bro's Before Flows - Troy Wojewoda - RVA5ec 2016
  - https://www.youtube.com/watch?v=utqsrVLM6mo
- Data Analysis, Machine Learning, Bro, and You! - Brian Wylie - BroCon 2017
  - https://www.youtube.com/watch?v=pG5lU9CLnIU
- Data Science Hunting Funnel - Austin Taylor
  - http://www.austintaylor.io/network/traffic/threat/data/science/hunting/funnel/machine/lea science-hunting-funnel/

About Us

	Derek Banks	Ethan Robish
Has kids	✓	✓
Likes to fish	✓	
Amateur photographer		✓
Incident Responder	✓	
Developer		✓
Threat Hunter	✓	✓
Writing a Course	✓	✓