Collecting, Analyzing and Responding to Enterprise Scale DNS Events

Bill Horne
Director, Security Research
HP Labs

Acknowledgements

Improving CSIRT Skills, Dynamics and Effectiveness

Parts of this material are based on research sponsored by the Department of Homeland Security (DHS) Science and Technology Directorate, Homeland Security Advanced Research Projects Agency, Cyber Security Division (DHS S&T/HSARPA/CSD), BAA 11-02 and the Air Force Research Laboratory, Information Directorate under agreement number FA8750-12-2-0258.

This is what we are dealing with...

HP IT supports 6 NGDC and 86 MCS

140+
Windows Domain Controllers
Microsoft

41K+
servers owned by HP IT

450,000
end points protected with anti-virus

1.2 million
connected devices

1,500
enterprise HPN Routers

Manage 150K+ mobile devices

15K+
HPN switches

11.5M+
Internet mails per day sent/received

2,000+ HP IT managed firewalls

2.5B
security events logged per day with ArcSight

597 IPS sensors deployed
TippingPoint

39,000,000
IP Addresses including 2 contiguous Class A's

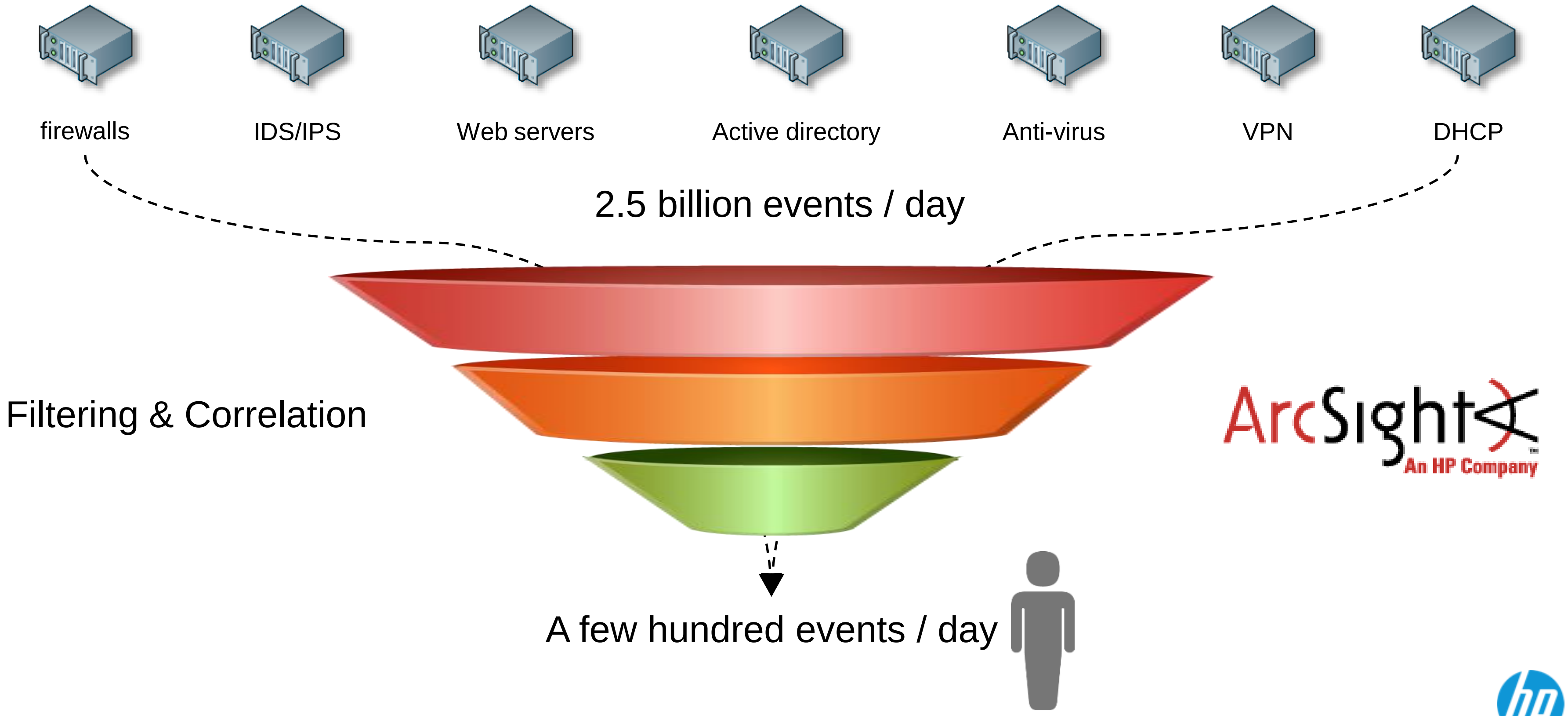
970K+
scanned devices for vulnerabilities

450,000
mailboxes managed

440K+
PCs deployed

300K+
employees + contractors

Security Information and Event Management



firewalls

IDS/IPS

Web servers

Active directory

Anti-virus

VPN

DHCP

2.5 billion events / day

Filtering & Correlation

ArcSight An HP Company

A few hundred events / day

01:36
MOSCOW
05:36
SINGAPORE
06:36
TOKYO
14:36
PALO ALTO
16:36
HOUSTON
17:37
RR
23:36
BERLIN
Mail
Calendar
People
Tasks

Challenges

Tedium

Work Force

Incentives

Laura Fletcher, Kristin M. Repchick, and Julie Steinke

Barriers and Pathways to Improving the Effectiveness of Cybersecurity Information Sharing Among the Public and Private Sectors

16:00 – 17:00 in POTSDAM I

Challenges

The Base Rate Fallacy

An intrusion detection system (IDS) performs deep packet inspection on network traffic within an organization. The system uses a signature to look for a particular type of malicious payload and fires an alert if the payload is seen. Given a payload, the IDS is quite accurate: it correctly classifies the packet as malicious or not 99.9% of the time. But, suppose that the malicious payload is rare: only 1 out of every 100,000 packets are expected to have the malicious payload. If an alert fires, what is the likelihood that the payload is malicious?

D := IDS fires an alert

M := payload is malicious

$$P(M|D) = \frac{P(D|M)P(M)}{P(D)} = \frac{P(D|M)P(M)}{P(D|M)P(M) + P(D|{\sim}M)P({\sim}M)}$$

$P(D|M) = 0.999$

$P(D|{\sim}M) = 0.001$

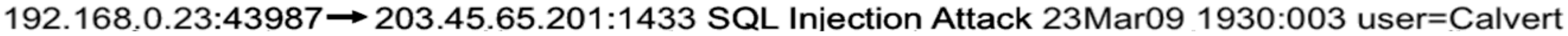
$P(M) = 0.00001$

$$P(M|D) = \frac{0.999 * 0.00001}{0.999 * 0.00001 +\ 0.001 * 0.99999} \cong 0.0098$$

$$\mathit{or\ 1\ per\ 102\ alerts}$$

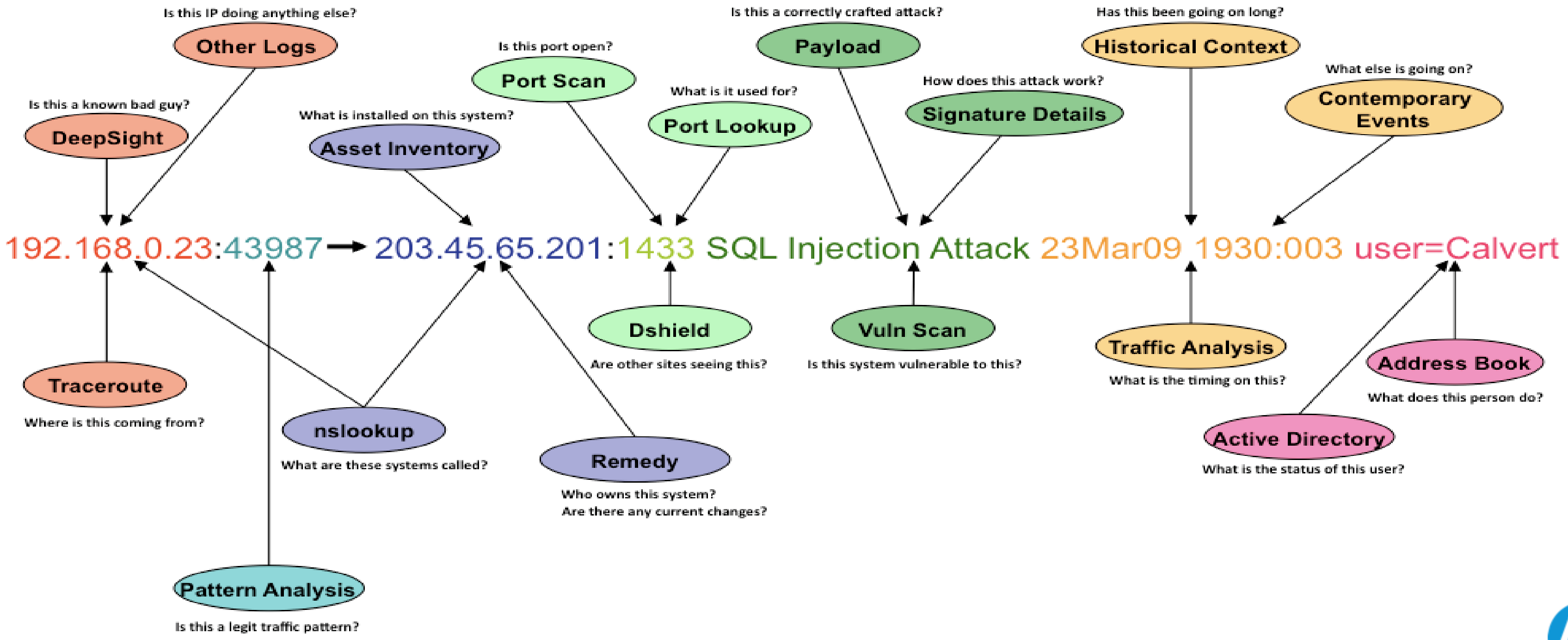
S. Axelsson, "The base-rate fallacy and the difficulty of intrusion detection," *ACM Transaction on Information System Security,* pp. 186-205, 2000.

What the analyst sees



```
192.168.0.23:43987→203.45.65.201:1433 SQL Injection Attack 23Mar09 1930:003 user=Calvert
```

What the analyst does



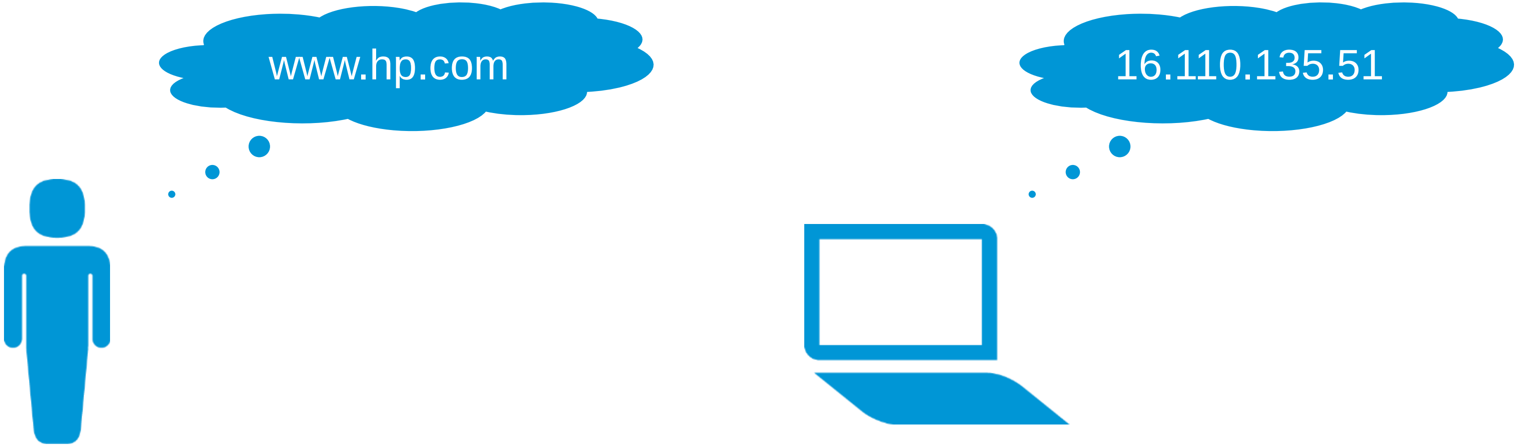
Is this IP doing anything else?
Other Logs
Is this a known bad guy?
DeepSight
What is installed on this system?
Asset Inventory
Is this port open?
Port Scan
What is it used for?
Port Lookup
Is this a correctly crafted attack?
Payload
How does this attack work?
Signature Details
Has this been going on long?
Historical Context
What else is going on?
Contemporary Events
192.168.0.23:43987 → 203.45.65.201:1433 SQL Injection Attack 23Mar09 1930:003 user=Calvert
Dshield
Are other sites seeing this?
Vuln Scan
Is this system vulnerable to this?
Traffic Analysis
What is the timing on this?
Address Book
What does this person do?
Active Directory
What is the status of this user?
Traceroute
Where is this coming from?
nslookup
What are these systems called?
Remedy
Who owns this system?
Are there any current changes?
Pattern Analysis
Is this a legit traffic pattern?

DNS

What is the Domain Name System (DNS)?

People think in terms of domain names

Computers communicate by IP addresses



www.hp.com

16.110.135.51

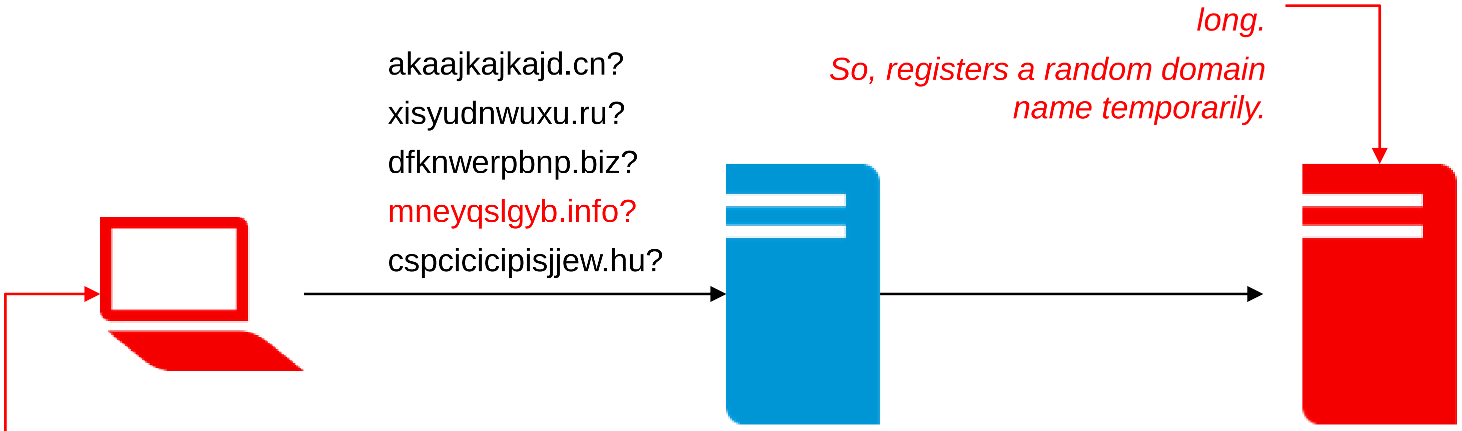
DNS maintains the mapping between domain names and IP addresses

DNS is important for security

Attacks Against DNS Servers	• Malformed Packets • Cache Poisoning
Attacks that use DNS to attack third parties	• DDoS Reflection & Amplification Attacks
Attacks that use DNS as part of their infrastructure	• Botnet Command and Control • Data Exfiltration & Tunnelling

Example

Botnet Command and Control



Attacker can’t maintain C&C server at IP address for very long.

So, registers a random domain name temporarily.

akaajkajkajd.cn?

xisyudnwuxu.ru?

dfknwerpbnp.biz?

mneyqslgyb.info?

cspcicicipisjjew.hu?

Bot

DNS server

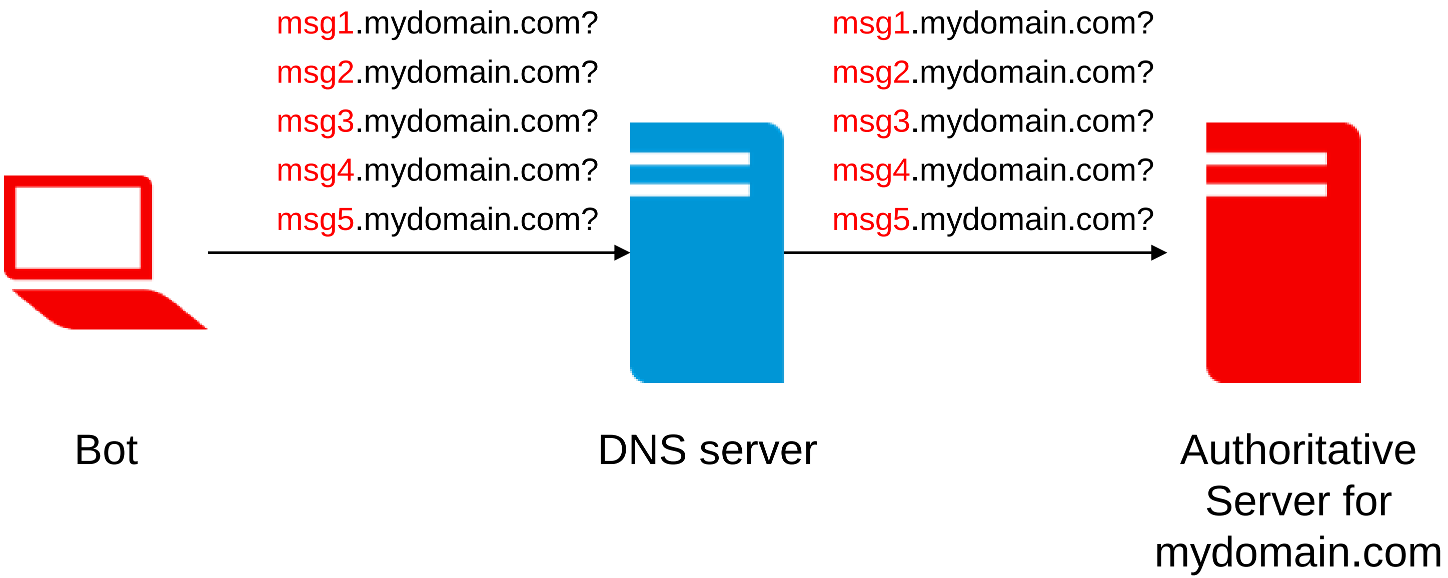
Command and Control Server

(mneyqslgyb.info
)

Bot tries a bunch of random names until it finds one that resolves.

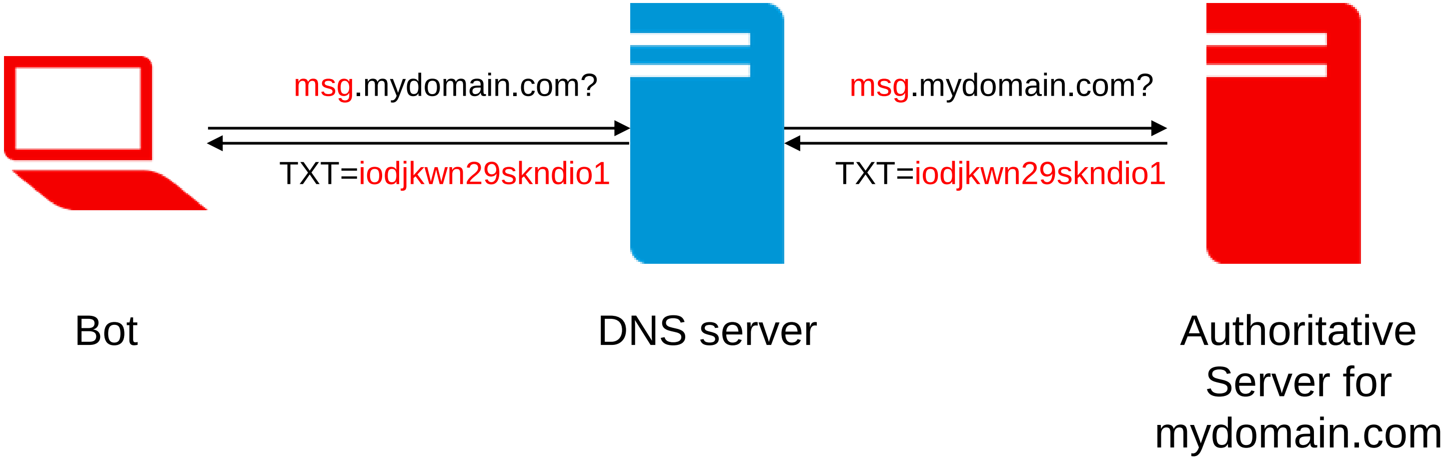
Example

Exfiltration



Example

Tunneling



Our Problem

Challenges in collecting DNS Data

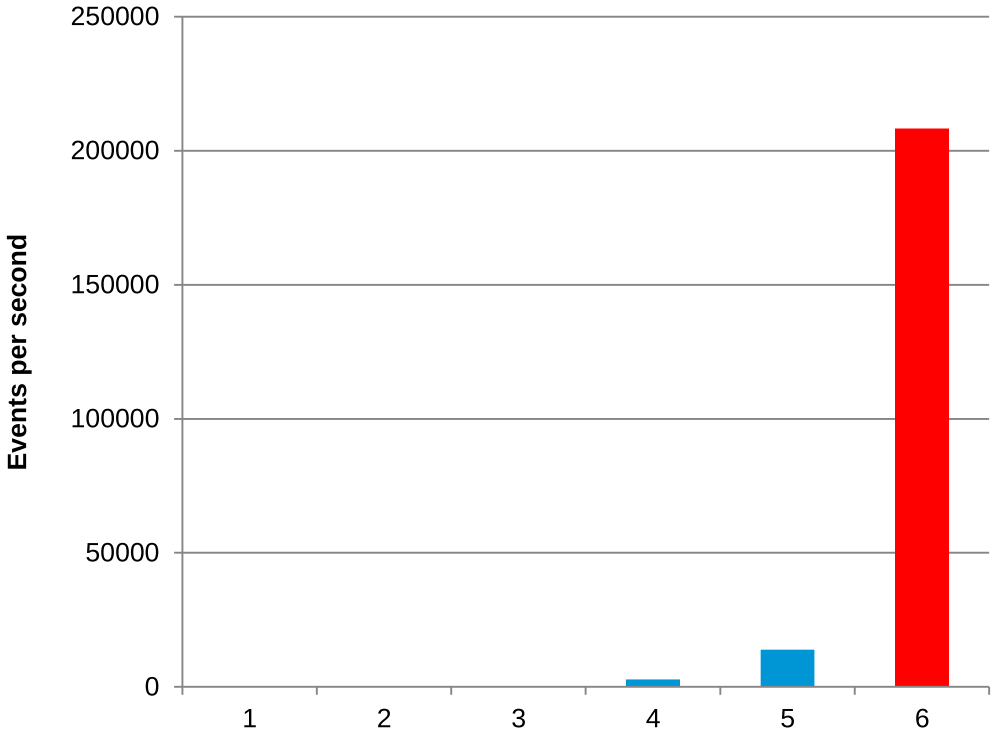
Volume and Detail

Why is this a hard problem?

18B DNS packets move through HP’s core data centers every day

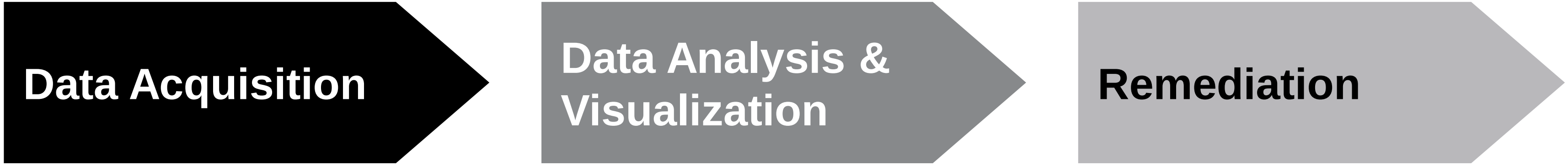
Logging severely impacts performance

The right information is not logged



Our Approach

End-to-end handling of DNS events



- Hardware Packet Sniffers
- Drop normal traffic, collect the rest
- Goal: Throw out 99% of events

- Real-time and near-time analysis
- Novel visualizations
- Integration with ArcSight SIEM workflow in SOCs

- Block traffic automatically
- Generate threat intelligence

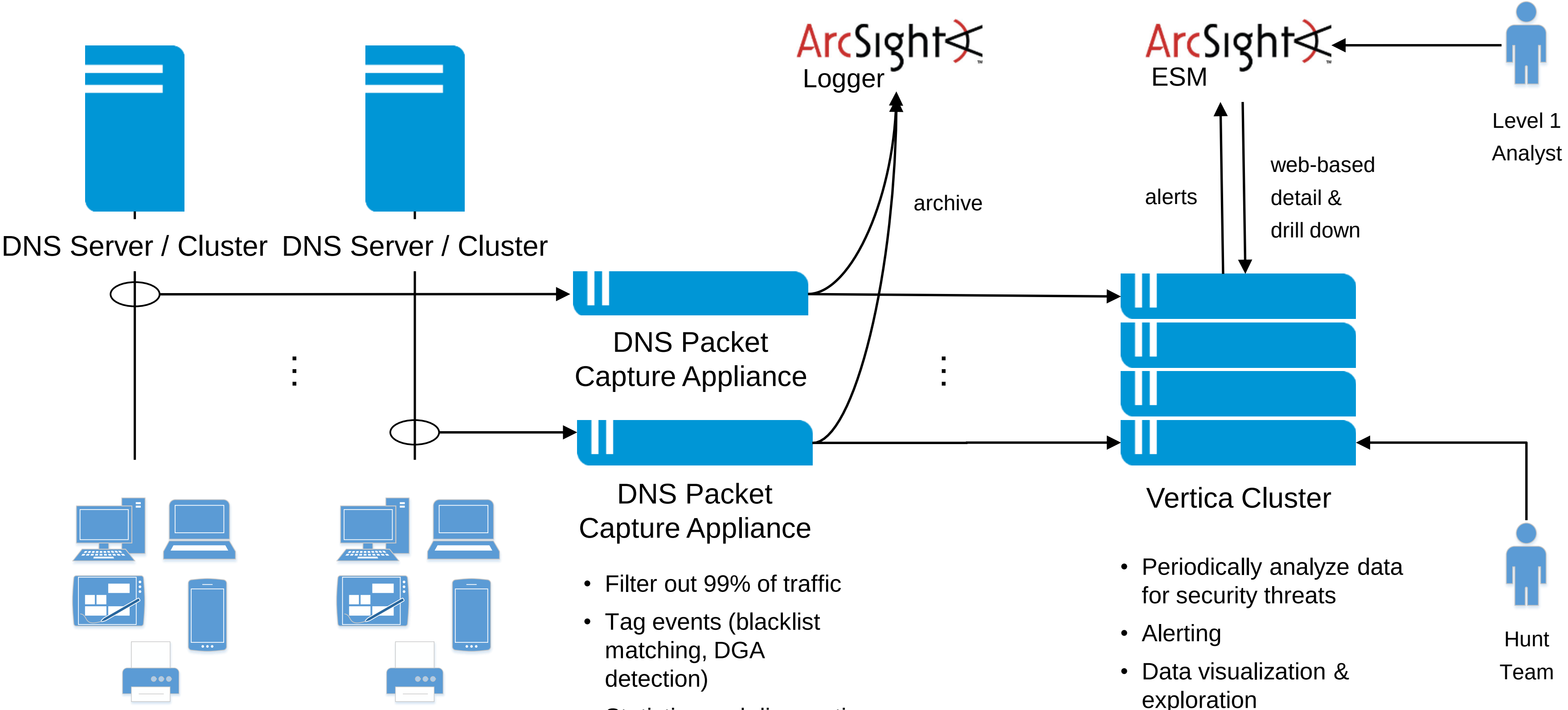
ArcSight™

VERTICA
An HP Company

REPUTATION DIGITAL VACCINE

TippingPoint

Architecture



ArcSight Logger

ArcSight ESM

Level 1 Analyst

archive

alerts

web-based detail & drill down

DNS Server / Cluster

DNS Server / Cluster

DNS Packet Capture Appliance

DNS Packet Capture Appliance

Vertica Cluster

Hunt Team

- Filter out 99% of traffic
- Tag events (blacklist matching, DGA detection)
- Statistics and diagnostics

- Periodically analyze data for security threats
- Alerting
- Data visualization & exploration

How do we filter out 99% of the traffic?

Exceptions

Unresolvable queries

- not FQDN, illegal characters, non-existent TLDs

Certain protocols

- Web Proxy Autodiscovery Protocol
- Bind version queries

Whitelisted Sources

“Aggregators”

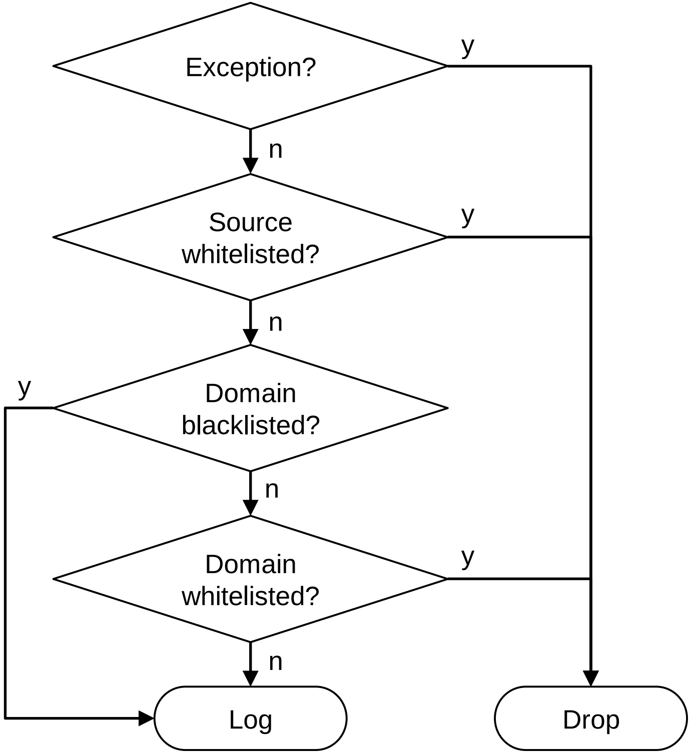
Security devices

Blacklisted Queries

Whitelisted Queries

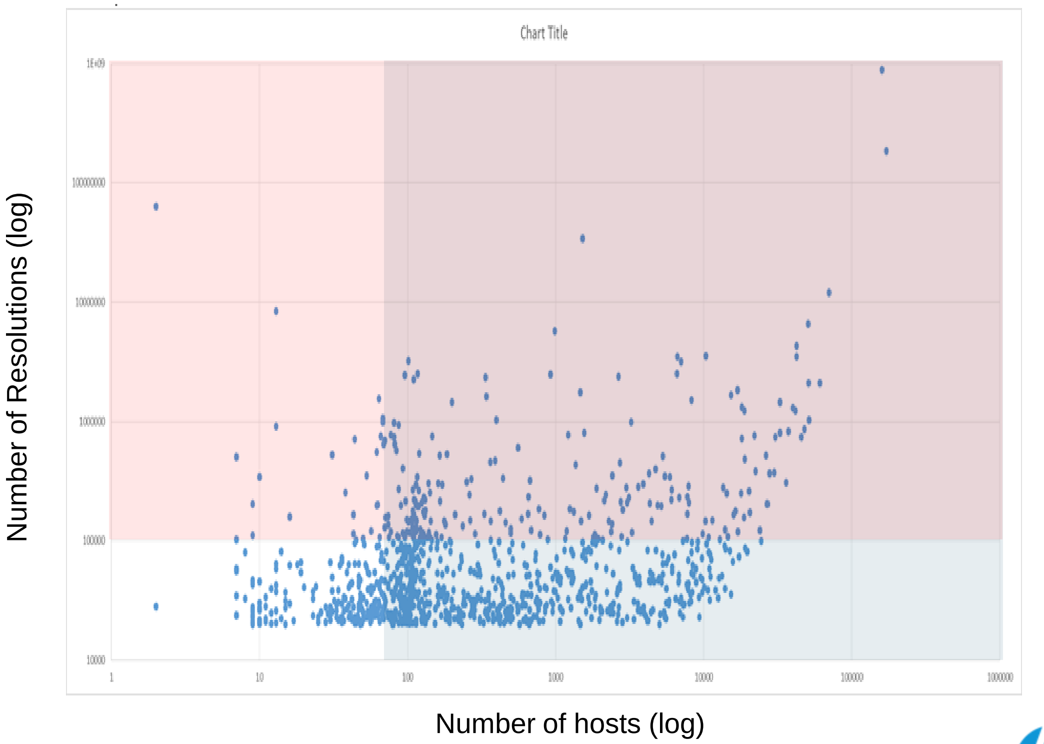
85% of queries are for HP authoritative domains

The rest we get from Alexa Top 1m



Heavy Hitters based Whitelisting

- Each dot represent one of the top 1000 most queried domains
- By choosing domains with >= 50 hosts we cover all the points in the right half-plane
- Further choosing domains resolved more than 10,000 times we cover most of these points
- Choosing the OR of these two conditions covers a large fraction of the traffic (Typically 90%)
- Observation: Very few of these heavy hitting domains are in black lists.



DGA Detection and Classification

Logistic Regression Classifiers

Labeled data from: Alexa, reversed malware, takedown/block lists, clustering real data

17 malicious DGA families, 3 suspicious, 2 unknown, and 3 benign

~1.4 million samples in dataset

K-way cross validation

Features

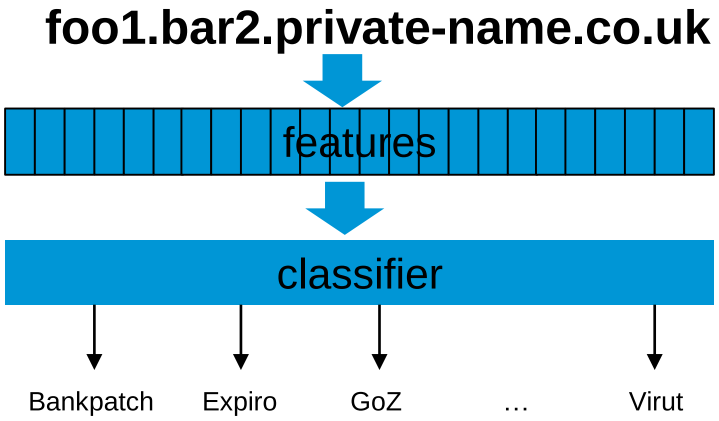
97110 features

Character groups: hex, upper, lower, digit, punctuation, etc.

Characters: 1,2,3-grams, character by position

Length of TLD, top private domain, rest

TLD



foo1.bar2.private-name.co.uk

features

classifier

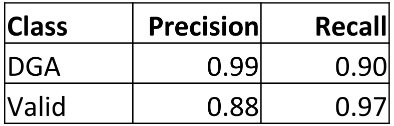
Bankpatch

Expiro

GoZ

…

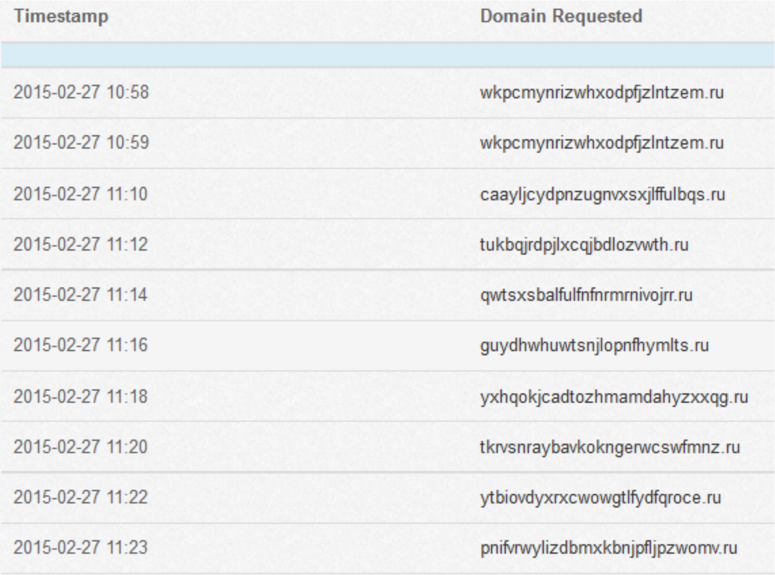
Virut



Class	Precision	Recall
DGA	0.99	0.90
Valid	0.88	0.97

Cheating the Base Rate Fallacy

- Look for machines making lots of queries to DGAs or blacklist entries in a short time period
- Assuming false positives are independent (questionable), then the machine is likely actually doing something bad (or is a security researcher!). Confirmed in practice.
- Can this be proved??



Timestamp	Domain Requested
2015-02-27 10:58	wkpcmynrizwhxodpfjzlntzem.ru
2015-02-27 10:59	wkpcmynrizwhxodpfjzlntzem.ru
2015-02-27 11:10	caayljcydpnzugnvxsxjlffulbqs.ru
2015-02-27 11:12	tukbqjrdpjlxcqjbdlozvwth.ru
2015-02-27 11:14	qwtsxsbalfulfnfnrmrnivojrr.ru
2015-02-27 11:16	guydhwhuwtsnjlopnfhymlts.ru
2015-02-27 11:18	yxhqokjcadtozhmamdahyzxxqg.ru
2015-02-27 11:20	tkrvsnraybavkokngerwcswfmnz.ru
2015-02-27 11:22	ytbiovdyxrxcwowgtlfydfqroce.ru
2015-02-27 11:23	pnifvrwylizdbmxkbnjpfljpzwomv.ru

This machine made 62 such queries in 4 hours.

Data Exfiltration and Tunneling

Queries

```
BLGCOFDAGOOOESDULBOOBOOOOOOOOOOOOOOOOOOLDOSESKGKHHF.detacsufbo.ru

EUJSFLDAGOOOESDUDBOOBOOOOOOOOOOOOOOOOOOOSSJHGHFCLFOHCHLGHSAHAHU.CHLAAFHLSGHAFGFUOOEUGDKLCSHEKLJBOCOSECHFFUGBSKGDJGGGHOJHJCGJG.KCDOELDUOEGUCUOUHJUA
KEGGGFGEKHLGFDFESJOEL.detacsufbo.ru

SHUDHFDAGOOOESDUGBOOBOOOOOOOOOOOOOOOOOOEDKDFBBHLEGGJLGUFABHCCU.DHDFFCHHKSHGHAOUBGEGEJLGFHUBDFGUGJDFFEAKFSBFFGSDACGHCSKBHLSCGHH.EHSHHJFHUAAOOGKKSDD
AHAUBBJDCCKGSHKLGJGAS.detacsufbo.ru

OHDOBHDAGOOESDUGBOOHOOOAOOOOOOOOOOOOOOO.detacsufbo.ru

HBSGGCDAGOOESDUUSOOBOOOOOOOOOOOOOOOOOOO.detacsufbo.ru

FSBLDDAGOOOESDUUSOOBOOOOOOOOOOOOOOOOOOO.detacsufbo.ru

KHFJCDAGOOOESDUGBOOHOOOAOOOOOOOOOOOOOOO.detacsufbo.ru

BSGKCDAGOOOESDULBOOBOOOOOOOOOOOOOOOOOOOLDOSESKGKHHF.detacsufbo.ru
```

Responses (TXT records)

```
LLCDGHDABOOOSSUHOOOFOOOOOOOOOOOOOOOOOOO

KJGDUDABOOOSBSUHOOOFOOOOOOOOOOOOOOOOOOO

JJDHUDABOOOSBSUHOOOFOOOOOOOOOOOOOOOOOOO

HBEAGDABOOOSBSUHOOOUOOOOOOOOOOOOOOOOOOO

KALFCSDAOOOSBSUHOOOFOOOOOOOOOOOOOOOOOOO

GHHFDDABOOOSBSUHOOOFOOOOOOOOOOOOOOOOOOO

COGOODABOOOSBSUHOOOUOOOAOOOOOOOOOOOOOOO

GHHKDGDABOOSBSUHOOOFOOOOOOOOOOOOOOOOOOO
```

Results

Since June 2014…

Processed 3.75 trillion DNS packets

Thrown 11,132 alerts for 3,840 distinct clients to our SOC

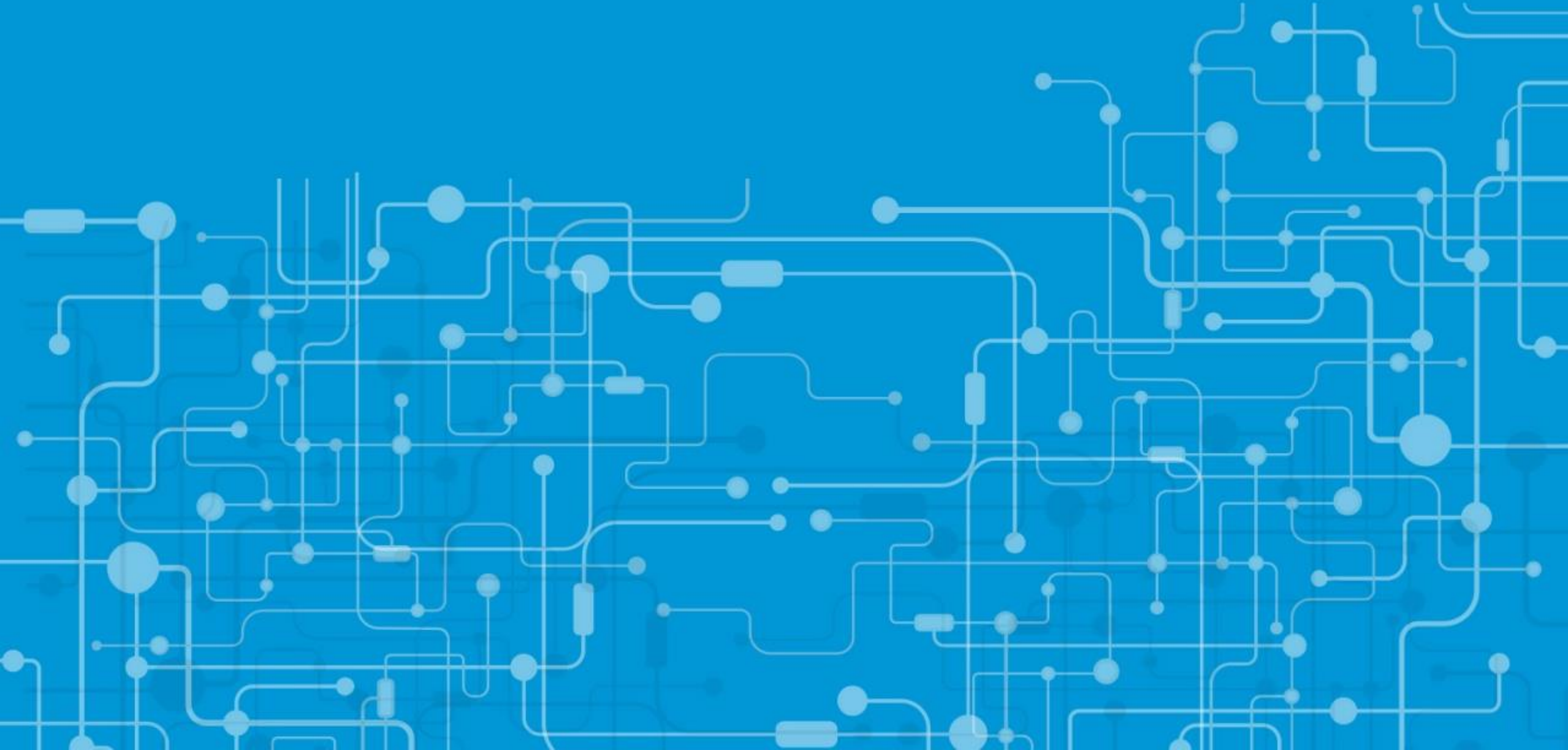
No reported false positives

Weird things we found that we weren't expecting

If there is a way to construct a malformed packet, it will appear on your network.

All sorts of machines do apparently bad things for good reasons

Demo



Lessons Learned

Solve Real Problems

Lots of interesting hard problems come up when you have to solve a real problem.

Get Good Data

If you have (lots of) good data, you can do interesting things.

Technology Isn't Everything

You have to make your technology compatible with the tools, workflow, and mandate of your users.

Thank you

Developed with HP Labs