Choose Your Battles
How To Fight The Right Wars

Eyal Paz, Security Researcher

Check Point
SOFTWARE TECHNOLOGIES LTD.

28th ANNUAL FIRST CONFERENCE
SEOUL
JUNE 12 - 17, 2016

whoami

- Security Researcher at Check Point
- B.Sc. in Software Engineering, studying towards M.Sc. in Computer Science
- Information Security lecturer
- Father

Check Point®
SOFTWARE TECHNOLOGIES LTD.

Agenda

- Research Motivation & Goals
- Under The Hood - Algorithmic Overview
  - Aggregating events to incidents
  - Differentiating incidents on host
  - In-house TI feed
  - Threat context

Motivation

Staying a Step Ahead of Threats

Make every effort to **PREVENT** attacks

Detection is not enough. The only way to avoid the cost of an attack is to prevent it altogether

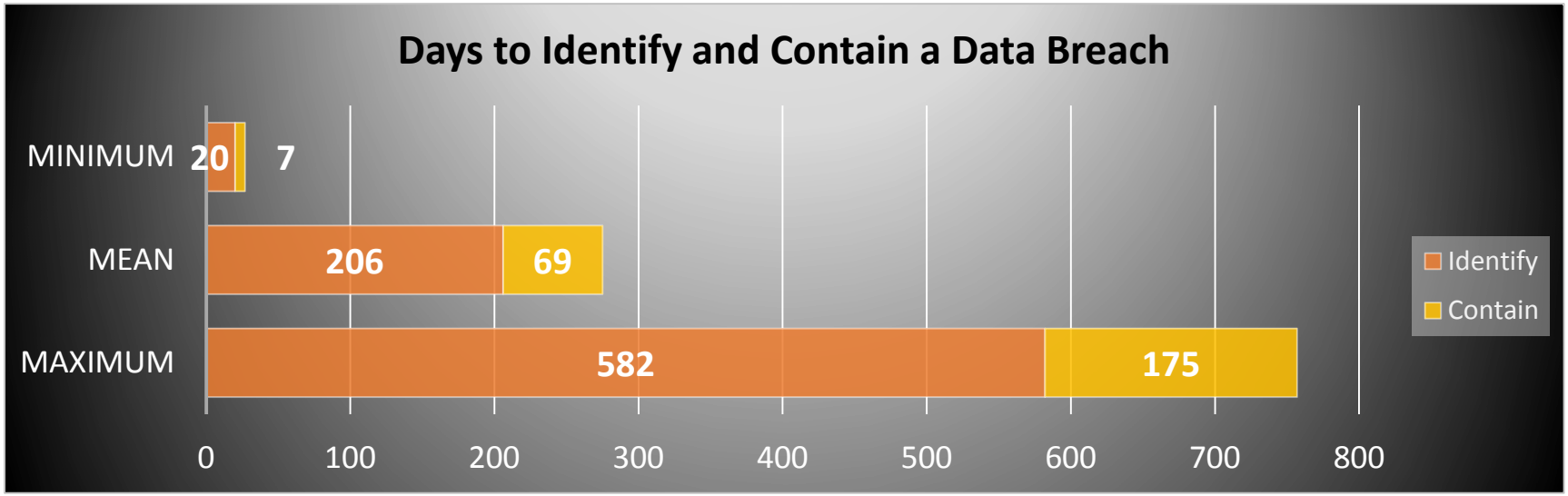
DETECT and **CONTAIN** attacks as soon as possible

Once infected, the cost of the attack just keeps on rising

Effectively **RESPOND** and **REMEDIATE**

Address the real business impact
Make sure the intrusion doesn't come back

Timing is Everything

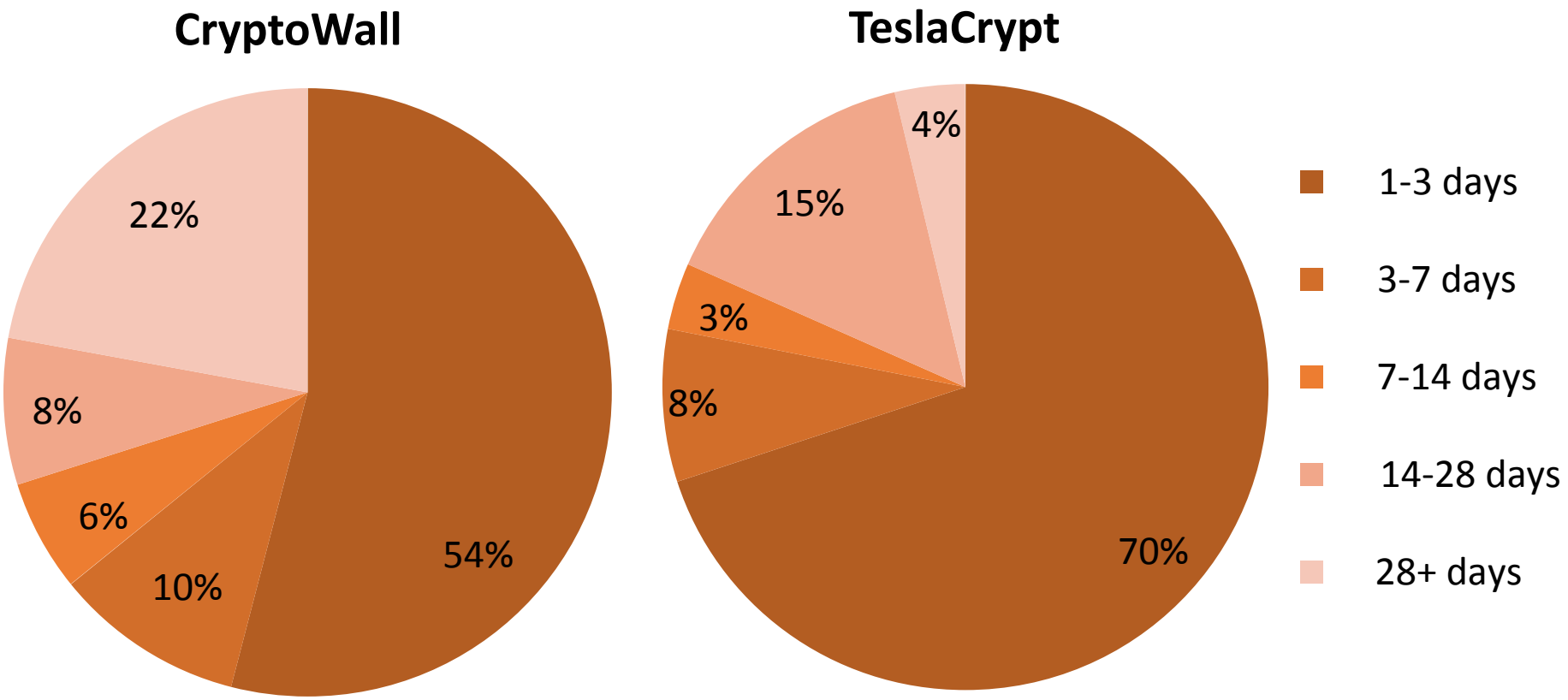


Source: *2015 cost of data breach study: global analysis, Ponemon Institute*

The Longer an attack goes **UNDETECTED**, the more time it takes to **CONTAIN** it

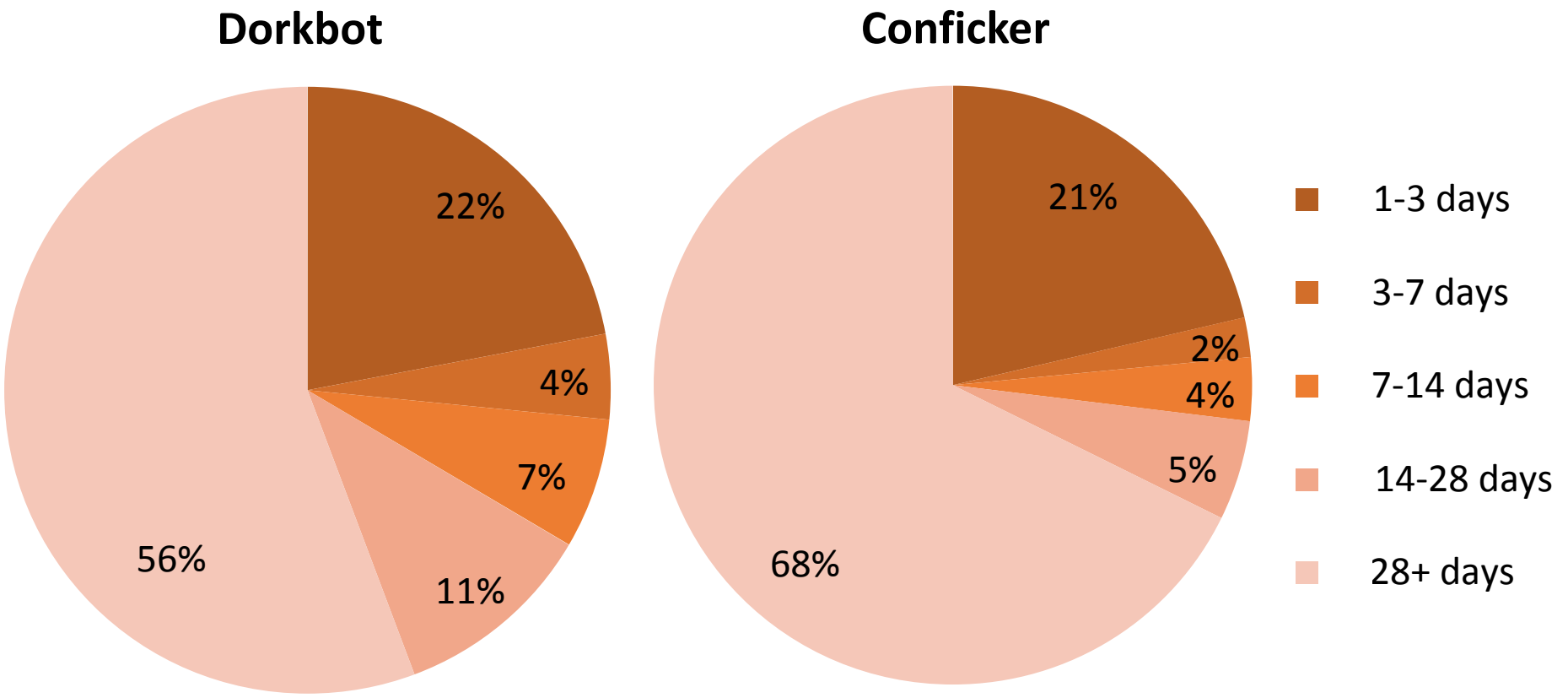
The longer it takes to **CONTAIN** it, the more it will **COST**

Loud Infection → Fast Response



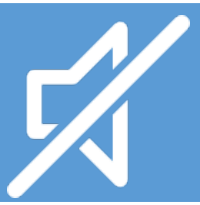
~70% of the infected machines are remediated within a week.

Silent Infection → Slow Response



~60% of the infected machines remediation takes more than a month.

Loud vs. Silent; What is More Severe?

		
Median Response Time	1-3 days	More than 28 days
Attack Vector	Mostly Phishing & Exploit kits	Mostly Phishing & Exploit kits
Attack Type	Data corruption, Denial of Service, Ransom demanding	Espionage, Banking credentials, Data breach
Discovery	Easy	Hard
Damage	Temporal	Continual

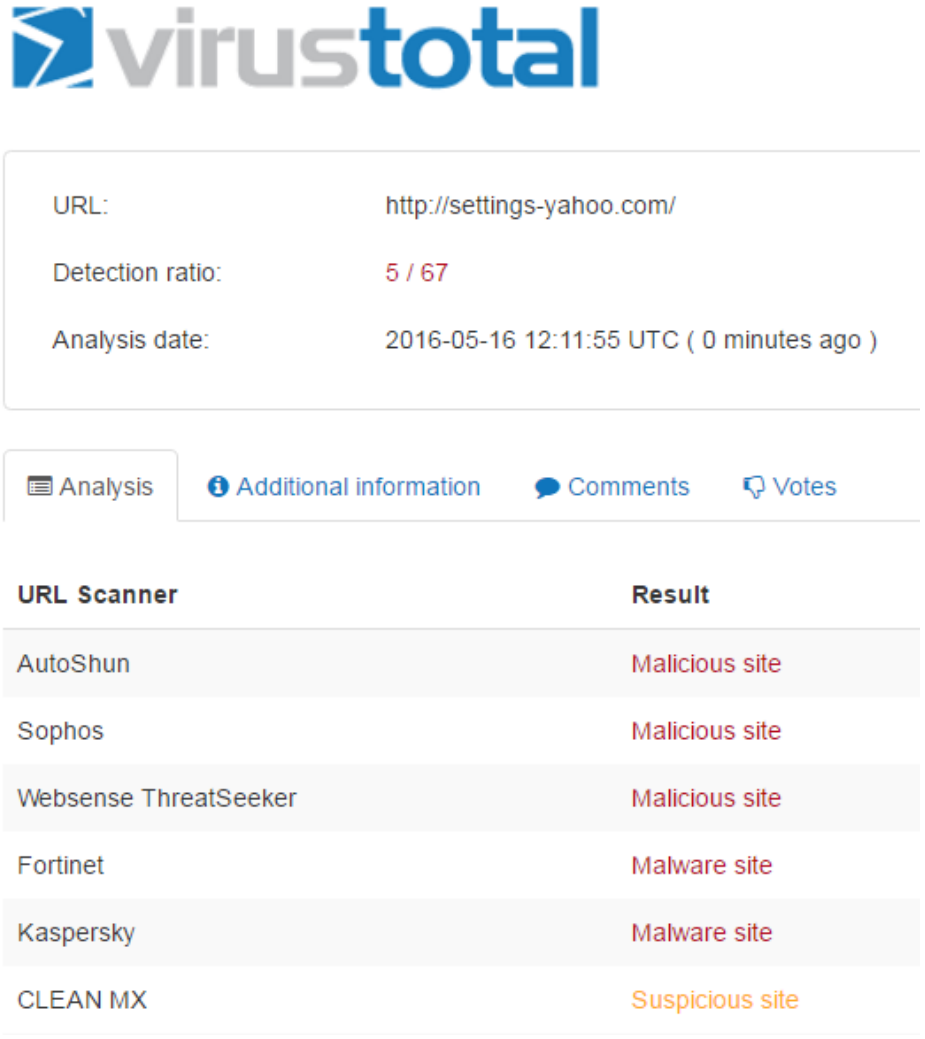
Reasons For Slow Response

- Internal bureaucracy and politics
  - Different teams with different agendas need to collaborate
- Network configuration issues
  - Difficult or impossible to track the infected host
- Understaffed security teams
  - “62% of organizations are receiving more alerts than they can feasibly investigate”

  Source: *2015 Incident Detection & Response Survey, RAPID7*

Threat Context

- Given one or more hosts access a “Malicious site”
- What should the security team do with such information?
- How should it be prioritized vs. other alerts?



Research Questions & Directions

- How to choose your battles

  Aggregate & summarize multiple alerts to a reasonable number of incidents to decrease workload

- How to fight the right war

  Adding a context layer to incidents to better prioritize their urgency

Algorithmic Overview

Aggregating Events to Incidents

- Discover similarity between compromised hosts
- Reduce overhead of security incidents
- Assist in prioritization & remediation
  - One script to clean them all

Step 1 – Pre-processing

- Get all alerts from all available sensors' events:
  - FW & IDS
  - End Point
  - Domain Controller
  - Proxy & DNS Servers

Step 2 – Feature Vector

- Create a list of all unique IoC
  - Domains
  - Destination IP for non HTTP/DNS addresses
  - Destination port
  - And any other forensics telemetry type you can get
- Not all features are equally weighted features

Step 3 – Host Matrix

- Create a matrix where the rows are for hosts and the columns are for the features
- Example:

  3 hosts – A, B, C

  4 IoCs – evil-1.com, evil-2.com, 1.2.3.4, TCP/6667

  Domain weight is 1, IP weight is 1.3, Port weight is 1.6

	evil-1.com	evil-2.com	1.2.3.4	TCP/6667
Host A	1	1	0	0
Host B	0	0	1.3	1.6
Host C	0	1	1.3	1.6

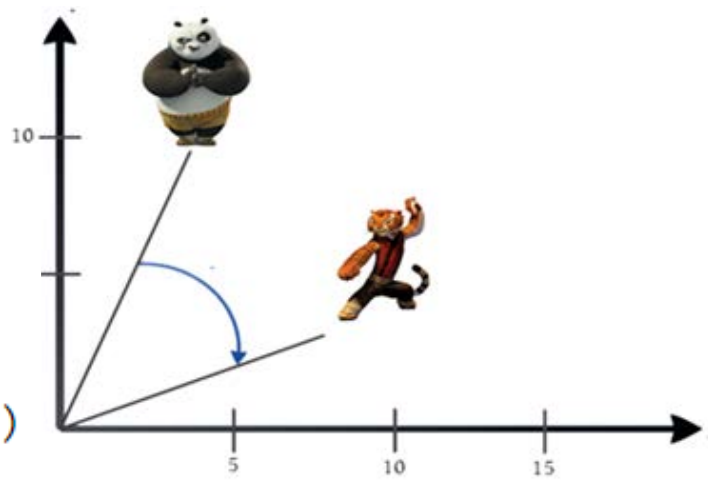
Cosine Similarity

- a measure of similarity between two vectors of an inner product space that measures the cosine of the angle between them – number in range [0,1]

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

```
def cosine_similarity(x,y):

 numerator = sum(a*b for a,b in zip(x,y))
 denominator = square_rooted(x)*square_rooted(y)
 return round(numerator/float(denominator),3)
```



Source: https://dataaspirant.com/2015/04/11/five-most-popular-similarity-measures-implementation-in-python/

Step 4 – Similarity Matrix

- Create the Cosine Similarity matrix when we are comparing every 2 hosts'
- In the below example:

  Green is for strong matches

  Yellow is for weak matches

  Red is for non-matches

	Host A	Host B	Host C
Host A	1	0	0.3
Host B	-	1	0.9
Host C	-	-	1

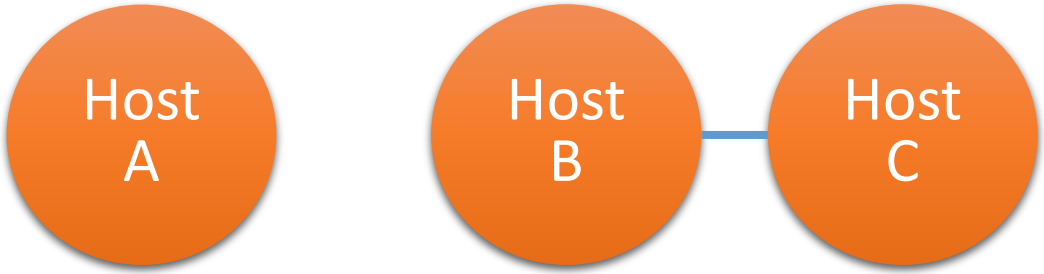
Step 5 – Noise Reduction

- Mask out weak matches for noise reduction

	Host A	Host B	Host C
Host A	1	0	0
Host B	-	1	0.9
Host C	-	-	1

Step 6 - Extract Incidents

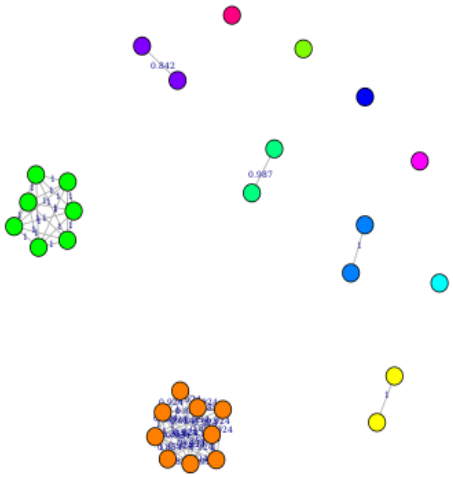
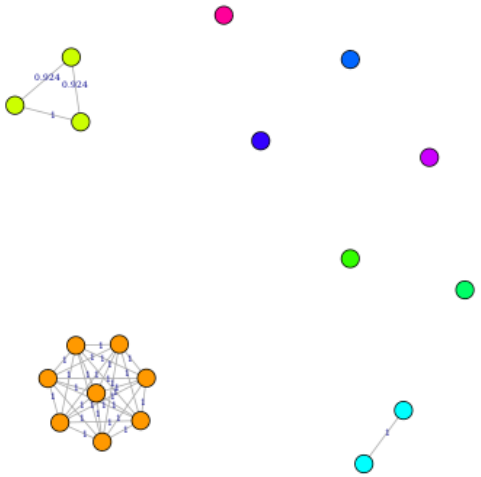
- Create a graph using the similarity matrix as a graph adjacency matrix



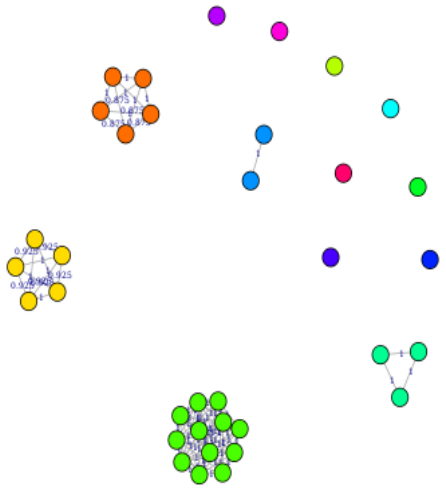
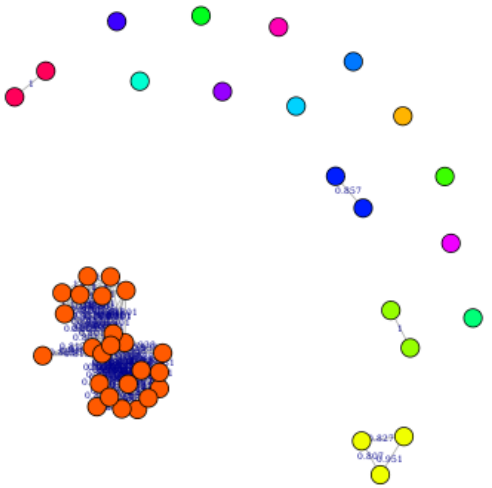
- Find the graph connected components which comprise the security incident that we looked for:

  {Host A}, {Host B, Host C}

PoC at Customer sites (24 Hours)

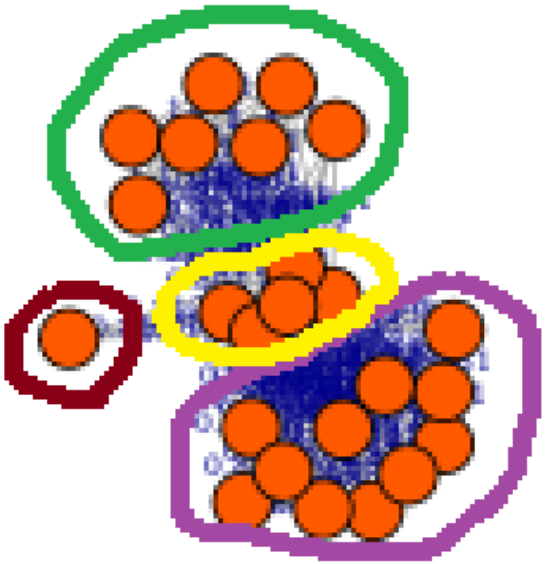
	Organization A	Organization B
Unique Indicators	177	41
Compromised Hosts	29	19
Security Incidents	11 (-62%)	9 (-52%)
Illustration		

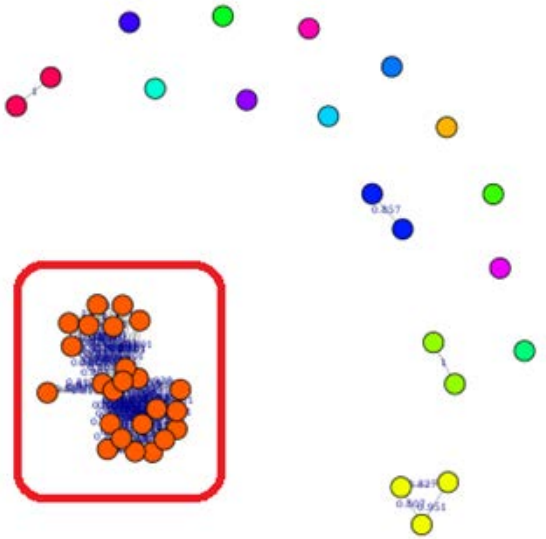
PoC at Customer sites (24 Hours)

	Organization C	Organization D
Unique Indicators	42	90
Compromised Hosts	35	44
Security Incidents	13 (-62%)	16 (-63%)
Illustration		

Model Limitation

- This model has a hidden assumption that all indicators that are found on a given host are related
- We all know that’s not always the case





Differentiate Incidents

- To differentiate the incident we need to break it down to its components – indicators
- Define similarity between indicators
- Consider recurring occurrences of the same indicators on different hosts

URL Similarity

- Equal non-zero amount of dashes
- Equal non-zero amount of digits
- Digits/Dash are on the same index
- Subdomains under same domain
- Same exact registrant
- Same anonymized registrant service
- Different anonymized registrant
- Small domain/registrant edit distance
- Same exact domain name
- Same domain name length
- Same IP resolutions amount
- Both domains had never had IP allocated
- Shared ASN
- Shared IP addresses
- Same TLD which is not .com and not local
- Close registration date
- Close first detected date
- Close language ratio
- Shared URL path exactly
- Similar URL path

CryptoWall C2 Servers

- Are the URLs below related?
  - `abelindia.com/1LaXd8.php`
  - `purposenowacademy.com/5_YQDI.php`
  - `mycampusjuice.com/z9r0qh.php`
  - `theGinGod.com/HS0ILJ.php`
  - `yahoosupportaustralia.com/8gX7hN.php`
  - `successafter60.com/iCqjno.php`
  - `alltimefacts.com/EiFSId.php`
- Other than the funny URL path pattern
  - All the above URLs were first seen on 04-Nov-2015 which indicate they belong to the same campaign

Emotet Malware DGA

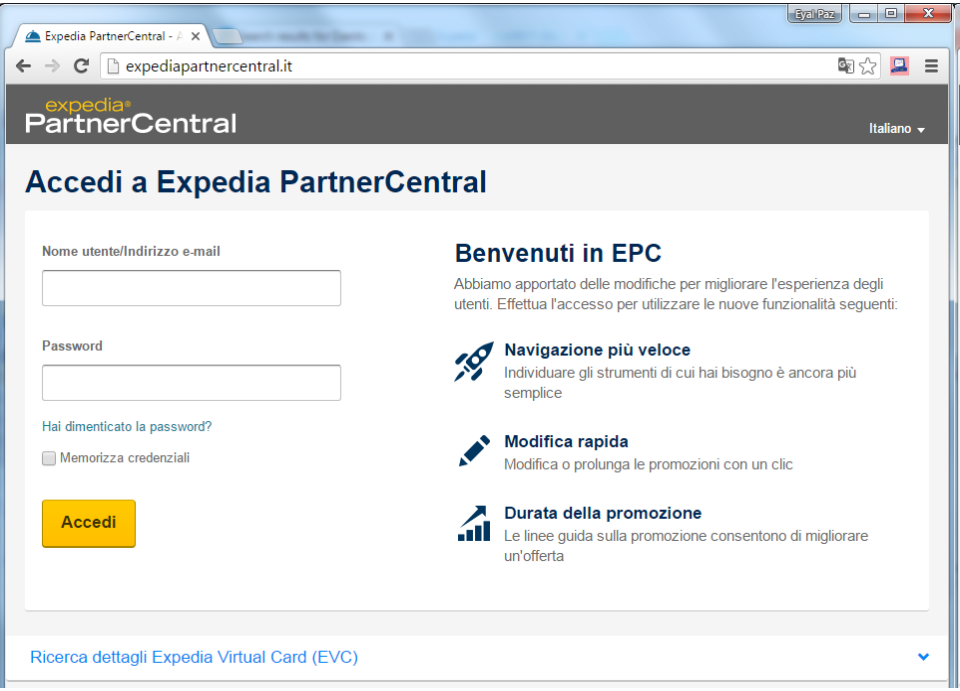
- Are the domains below similar?
  - `myjfqirgagnpboou.eu`
  - `kgpaorkwqlgrfcre.eu`
  - `pqxhqpvumylnikjh.eu`
  - `iddxbogywitoaddv.eu`
  - `clgarxlbvxcraqht.eu`
  - …
- Other than the simple pattern [a-z]{16}\.eu
  - All domains had never had an IP allocated
  - All domains were never registered
  - Close linguistic ratio
  - Same TLD which is not .com and not local

Virus Total URL - Emotet DGA

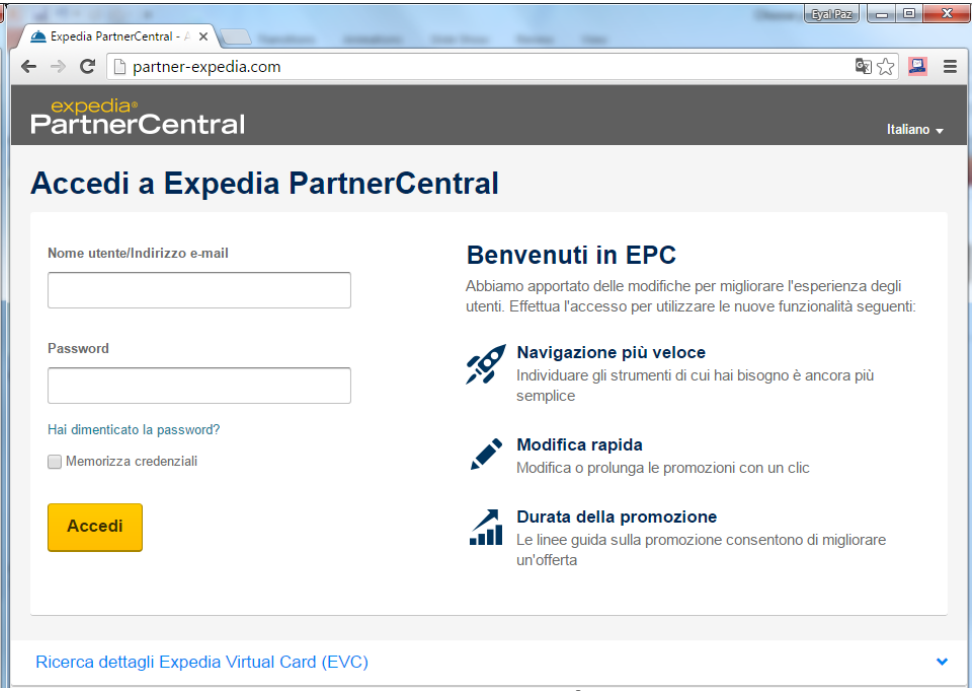
Domain	VT URL Detection*
pqxhqpvumylnikjh.eu	0/67
iddxbogywitoaddv.eu	0/67
idlueqkbfkkclcdj.eu	0/67
jjnstqfppyclvonk.eu	0/67
clgarxlbvxcraqht.eu	1/67
kgpaorkwqlgrfcre.eu	1/66

* Scanned on May-2016

Expedia Phishing Campaign



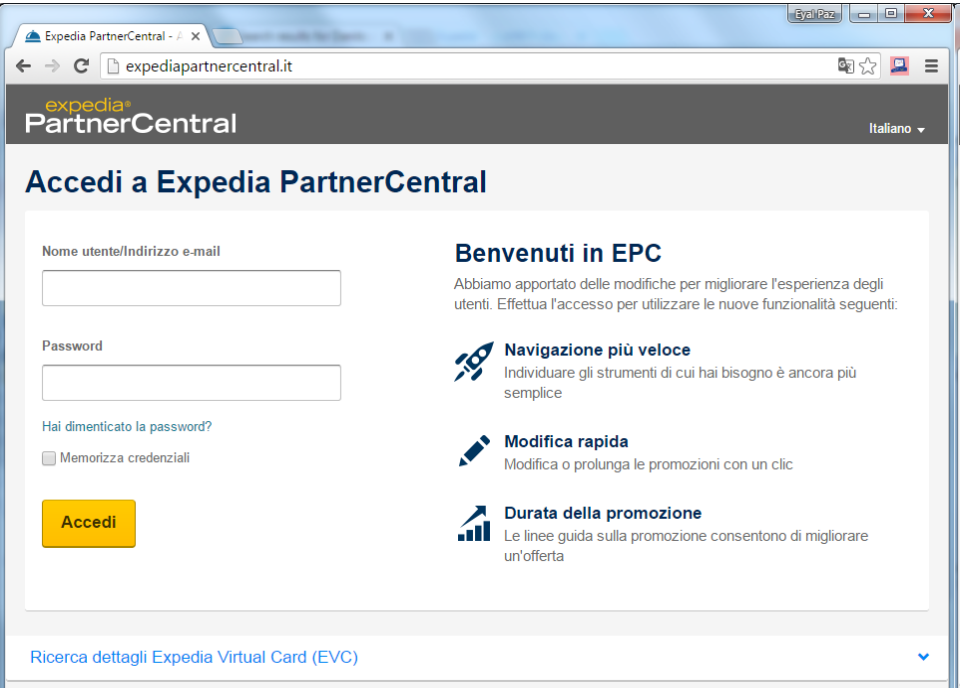
expediapartnercentral.it



partner-expedia.com

Can you spot the Phishy one?

Expedia Phishing Campaign



expediapartnercentral.it	partner-expedia.com
Legal Department- Domain Administrator	danito alex
domains@expedia.com	alexxissisi@libero.it

More Like This…

- Under the name of “danito alex” two more domains were registered on the same day
  - accessoclienti-expedia.it
  - accessoclienti-expedia.com

List of domain names registred by **Danito Alex**

Domain Name	Create Date	Registrar
partner-expedia.com	2016-04-27	pop.it
accessoclienti-expedia.com	2016-04-27	ascio.com
accessoclienti-expedia.it	2016-04-27	

Source: http://domainbigdata.com/name/danito%20alex

VT URL - Expedia Phishing Campaign

Domain	VT URL Detection*
accessoclienti-expedia.com	0/67
accessoclienti-expedia.it	2/67
partner-expedia.com	7/67

* Scanned on June-2016

Step 1 – Pre-processing

- Get all IoC from all available sensors' events:
  - FW & IDS
  - End Point
  - Domain Controller
  - Proxy & DNS Servers

Step 2 – Similarity Graph

G ← Init-Graph()

For each pair of IoC of same type, do:

    G.Add-Node(IoC-A)

    G.Add-Node(IoC-B)

    If G.Has-Path(IoC-A, IoC-B) = False
    AND IoC-A is similar to IoC-B, then:

        G.Add-Edge(IoC-A, IoC-B)

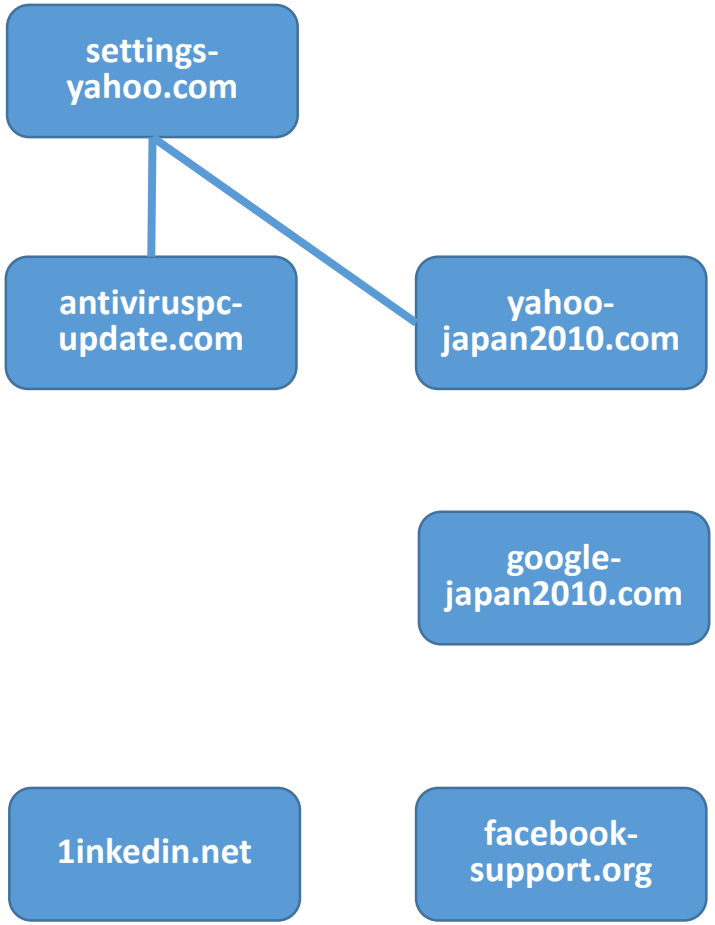
Incidents ← G.Connected-Components()

Phishing Actor

- Are the domains below similar?
  - `settings-yahoo.com`
  - `1inkedin.net`
  - `antiviruspc-update.com`
  - `google-japan2010.com`
  - `yahoo-japan2010.com`
  - `facebook-support.org`

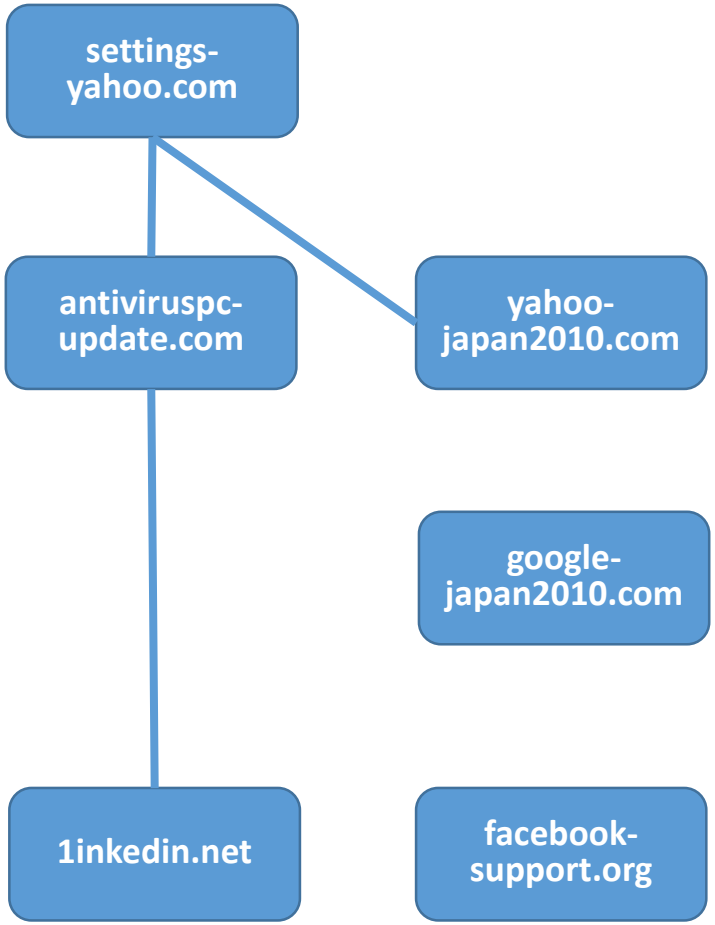
Phishing Actor

- settings-yahoo.com ⇔ 1inkedin.net
  - Same anonymized registrant service provider
- settings-yahoo.com ⇔ antiviruspc-update.com
  - Shared IP addresses
  - Same anonymized registrant service provider
  - Equal non-zero amount of dashes
  - Same IP resolutions amount
- settings-yahoo.com ⇔ google-japan2010.com
  - Same anonymized registrant service provider
  - Equal non-zero amount of dashes
  - Both contain popular domain name
- settings-yahoo.com ⇔ yahoo-japan2010.com
  - Shared IP addresses
  - Same anonymized registrant service provider
  - Equal non-zero amount of dashes
  - Both contain same popular domain name
- settings-yahoo.com ⇔ facebook-support.org
  - Shared IP addresses
  - Same IP resolutions amount
  - Equal non-zero amount of dashes
  - Both contain popular domain name



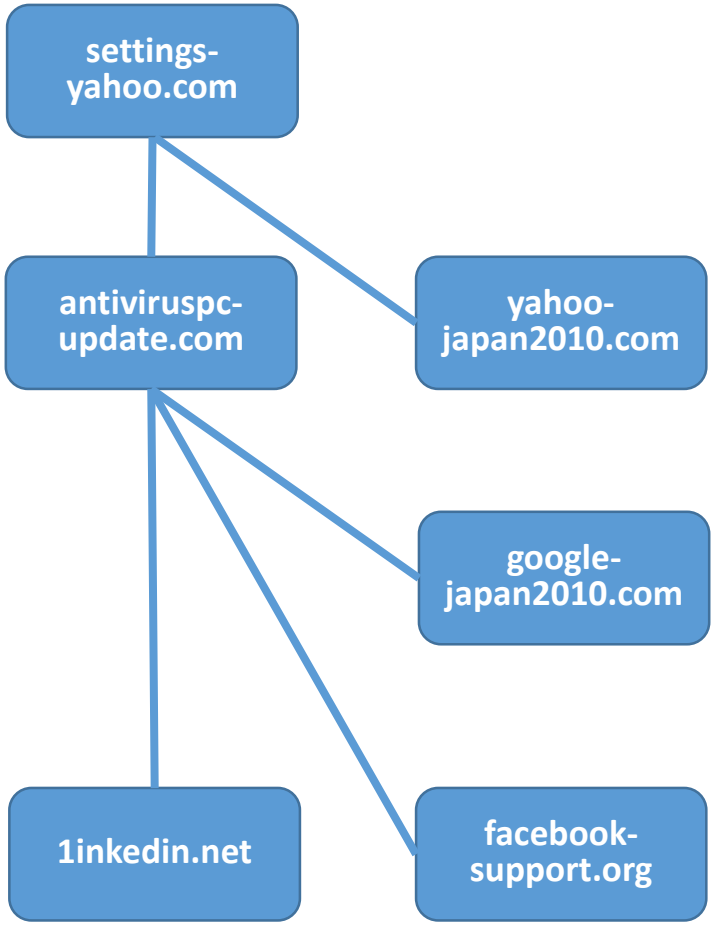
Phishing Actor

- ✓ 1inkedin.net ⇔ antiviruspc-update.com
  - Same anonymized registrant service provider
  - Close registration date
  - Shared IP addresses
- ✗ 1inkedin.net ⇔ google-japan2010.com
  - Same anonymized registrant service provider
  - Shared IP addresses
- ✗ 1inkedin.net ⇔ facebook-support.org
  - Shared IP addresses
  - Close registration date



Phishing Actor

- ✓ antiviruspc-update.com ⇔ google-japan2010.com
  - Same anonymized registrant service provider
  - Equal non-zero amount of dashes
  - Shared IP addresses
- ✓ antiviruspc-update.com ⇔ facebook-support.org
  - Close registration date
  - Same IP resolutions amount
  - Equal non-zero amount of dashes
  - Shared IP addresses



The graph is connected; therefore, all the domains are related

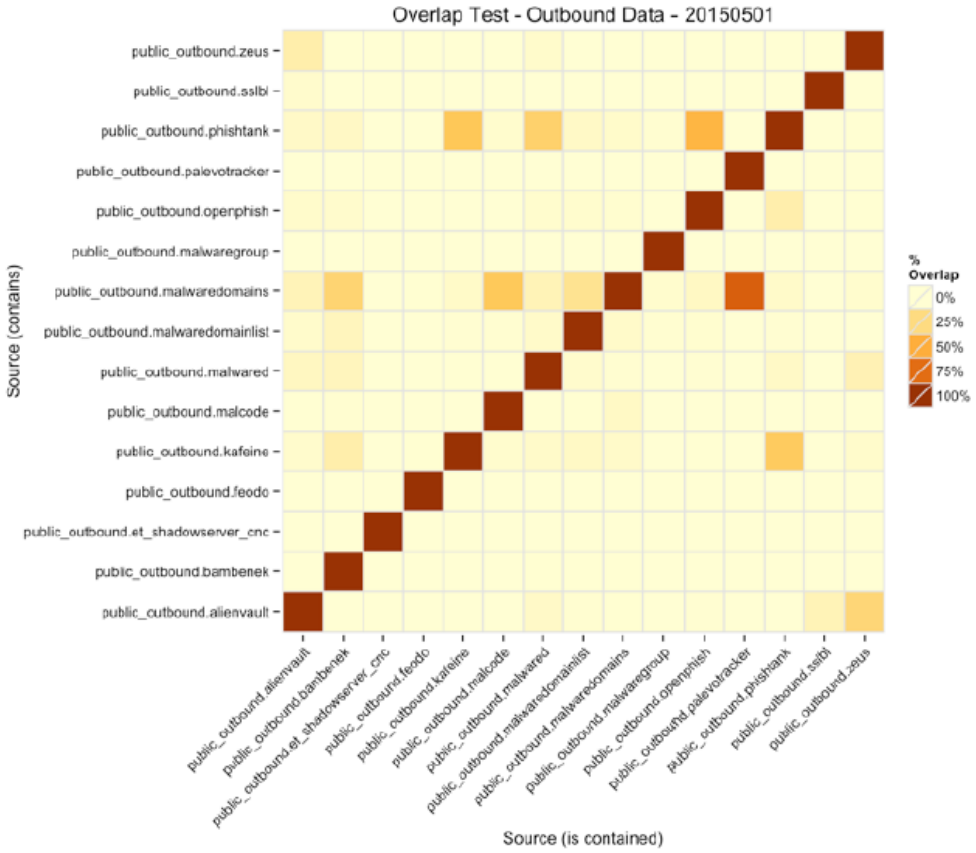
Virus Total URL - Phishing Actor

Domain	VT URL Detection*
google-japan2010.com	0/67
yahoo-japan2010.com	0/67
facebook-support.org	1/66
1inkedin.net	1/67
antiviruspc-update.com	2/67
settings-yahoo.com	5/67

* Scanned on May-2016

There's Always Room For More BL

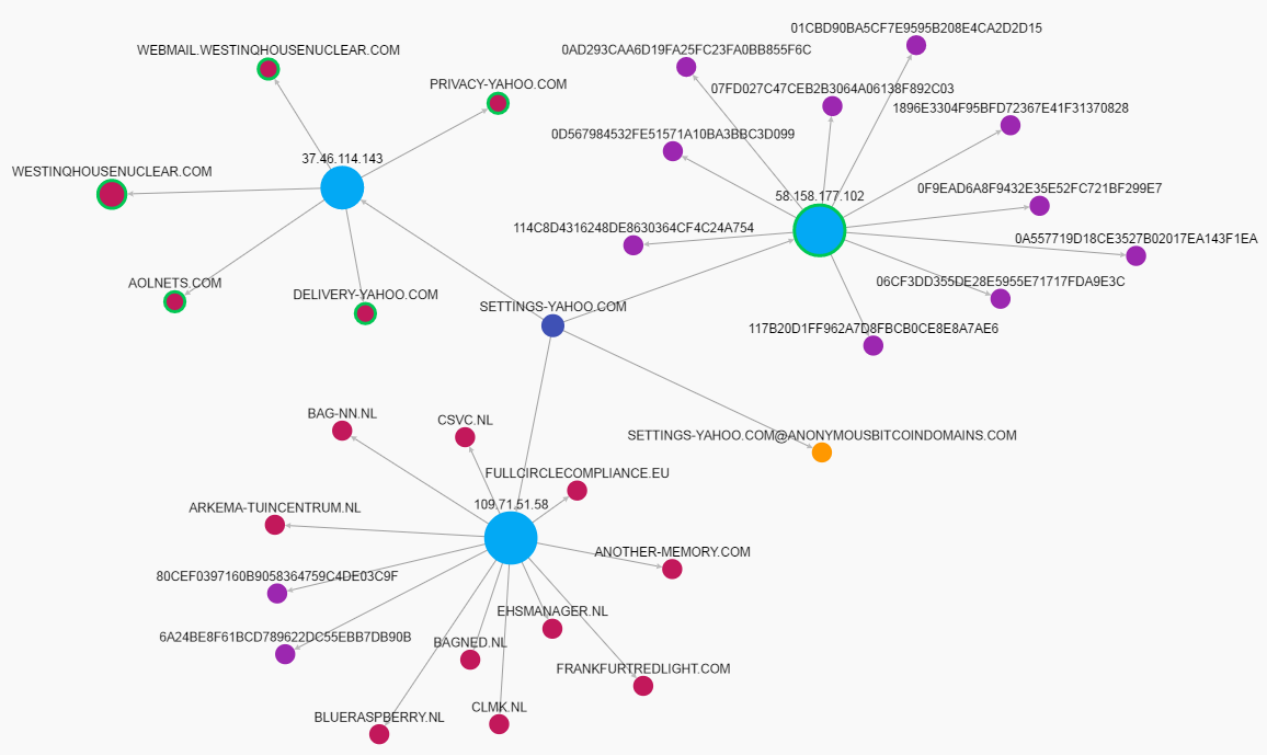
- There are many Threat Intelligence (TI) feeds out there
- The overlap between them is surprisingly low
- Putting all the vendors together still gives a partial coverage of the evilness on the internet



Source: *Data Driven Threat Intelligence: Metrics on Indicator Dissemination and Sharing, MLSec/Niddel*

IoC Similarity as a TI Feed

- The idea is to leverage existing feeds to create an in-house TI feed



Source: *https://www.threatcrowd.org/*

investigate-domain(domain)

If domain is suspicious, then:

For each domain’s ip resolution, do:

ip-investigation-queue.enqueue(ip)

For each file downloaded/communicated with the domain:

file-investigation-queue.enqueue(file)

For each registrant owned the domain:

registrant-investigation-queue.enqueue(registrant)

In-House Feed Value

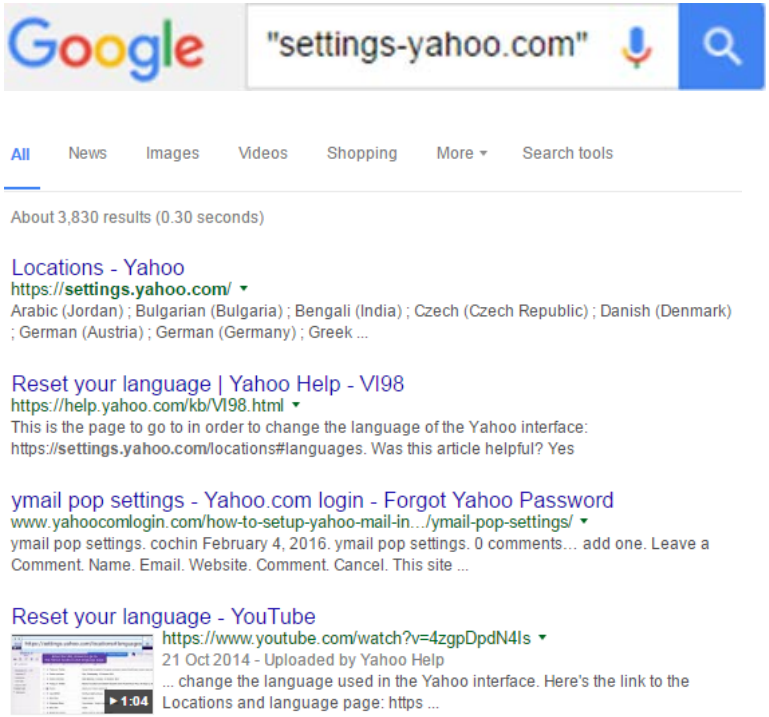
- Feed relevancy is crucial
- High hit rate of harvested indicators comparing to common TI feeds
- Proactively get as many indicators as possible of the current actor attacking the network

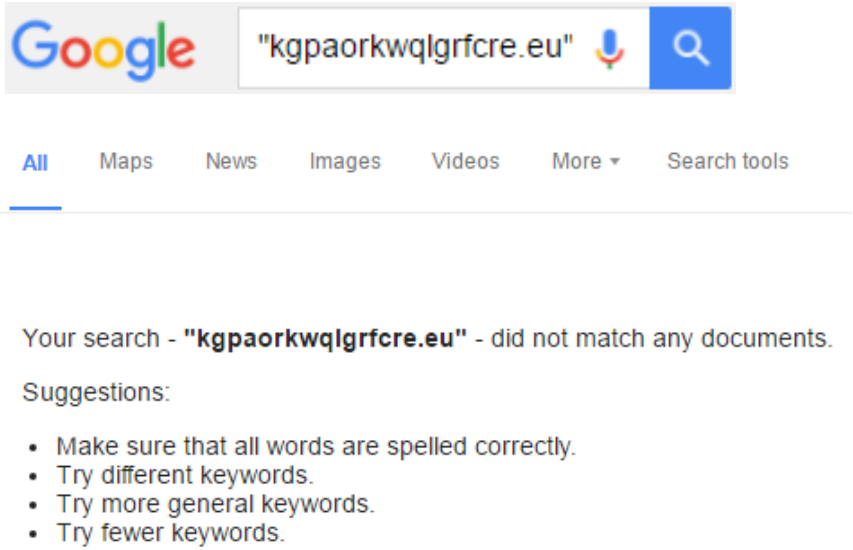
Share TI For Your Own Interest

- Organizations on same geo/industry/size are likely to get the same kind of attacks
- Sharing indicators between them could be the key differentiator between **DETECT** vs. **PREVENT**
- Actively sharing communities should be everyone's interest

Threat Context

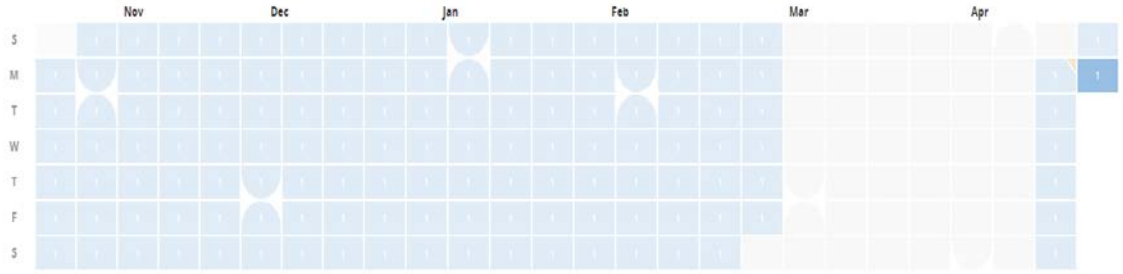
- Adding more IoC is great
- But more alerts are pointless if they are without the proper threat context





Domain Classification Analysis #1

settings-yahoo.com

Evidence	Illustration
Domain Contained popular domain string (by Alexa)	settings-yahoo.com 
Anonymized domain registrations (by who.is)	Registrant Email: whoisproxy@value-domain.com
Website going up and down (by PassiveTotal)	

Verdict: Evidence implies a phishing /infecting website – Pre-Intrusion

Domain Classification Analysis #2

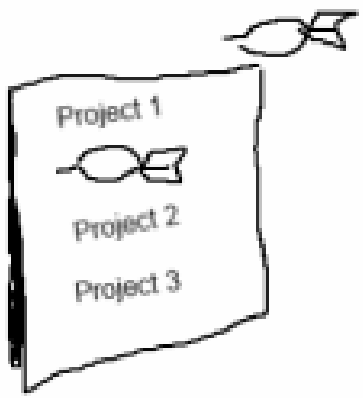
kgpaorkwqlgrfcre.eu

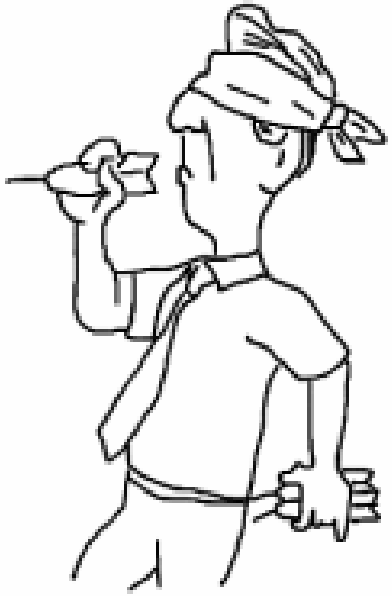
Evidence	Illustration
Domain is available for registration (by who.is)	**kgpaorkwqlgrfcre.eu** is available! $4.99 Buy it now!
Domain was never assigned to an IP (by PassiveTotal)	ATTRIBUTES First Seen N/A Last Seen N/A Resolutions 0
Domain was seen with which many like him within several minutes	myjfqirgagnpboou.eu, pqxhqpvumylnikjh.eu, iddxbogywitoaddv.eu, clgarxlbvxcraqht.eu, jjnstqfppyclvonk.eu, idlueqkbfkkclcdj.eu

Verdict: Evidence implies a CnC server – Post Intrusion

Alerts Prioritization

- Host resolving a phishing/infecting domain indicates an infection attempt
- Host resolving a *CnC server* domain indicates an on-going infection





Staying a Step Ahead of Threats

Events to Incidents → Faster Remediation

In-House TI Feed → Faster Intrusion Containment

Sharing TI → Moving From **Detect** To **Prevent**

Choose Your Battles
How To Fight The Right Wars

Eyal Paz, Security Researcher

Check Point
SOFTWARE TECHNOLOGIES LTD.

Thank You!

References

- 2015 Incident Detection & Response Survey, RAPID7
- 2015 Cost of data breach study: global analysis, Ponemon Institute
- Data Driven Threat Intelligence: Metrics on Indicator Dissemination and Sharing, MLSec/Niddel
- Similarity measures in Python, dataaspirant.com
- Cosine similarity, Wikipedia

References cont.

- http://blogs.checkpoint.com/
- https://threatcrowd.org/
- https://passivetotal.org/
- http://alexa.com/
- https://virustotal.com/
- http://who.is/
- https://google.com/
- http://malwarefor.me/
- http://domainbigdata.com/