

PHOENIX & CERBERUS

We haz botnets!

#Honeynet2014

Stefano Schiavoni, Edoardo Colombo

Federico Maggi

Lorenzo Cavallaro

Stefano Zanero

Politecnico Di Milano & Royal Holloway, University of London



POLITECNICO
DI MILANO



BOTNETS



About ▾

Community ▾

Events ▾

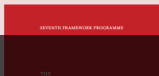
Scholarships ▾

Publications

Publicity ▾

News

The SysSec Red Book



Our updated **Roadmap for Systems Security Research** has been released! This effort has been coordinated by the SysSec consortium, with young researchers in the area playing a leading role in shaping the Roadmap and the consultation of the SysSec community and Associate Members.

Our Research Roadmap, labeled "**The Red Book**" was published on **September 1st** and will also be printed in hard copies as a book. Click on the book cover or go to red-book.eu to get it.

@syssecproject

SysSec Project: <http://syssec-project.eu>

Stay updated

Social media

Tweets

Follow

SysSec Project @syssecproject 2 May
#CFP: International #Workshop on Building Analysis Datasets & Gathering Experience Returns for #Security #BADGERS2014 necoma-project.eu/badgers14
Expand

SysSec Project @syssecproject 2 May
#CFP: 3rd International #Workshop on Quantitative Aspects in #Security Assurance #QASA2014. lit.cnr.it/qasa2014/ #poland #ESORICS2014
Expand

SysSec Project @syssecproject 10 Apr
Call for Participation: 1st A4Cloud Summer School #Accountability #Security #Cloud #SummerSchool a4cloud.eu/first_summer_s...

Tweet to @syssecproject

Managing Threats and Vulnerabilities in the Future Internet

Defending ICT infrastructures

Over the past decade we are witnessing an ever-increasing amount of cyberattacks on the Internet. Prolific, ingenious, and ranging in style from large-scale worms to "below the radar" phishing attempts, cyberattacks have evolved to unprecedented levels of sophistication.

These attacks pose a significant threat to the security of our society and the economy. In the past, the security of ICT infrastructures was often based on a reactive approach, where security measures were only implemented after the attack had occurred. This is no longer sufficient in the current environment, where the threat is constantly evolving and the attack surface is constantly expanding. Therefore, a proactive approach is needed, where security measures are implemented before the attack occurs. This is the focus of the SysSec project, which aims to develop new security measures and tools to defend ICT infrastructures against cyberattacks.

The bot **encrypts** files on the victim's computer and asks for a **ransom** to recover them.

¹<http://www.cybersec.kent.ac.uk/Survey2.pdf>

²<http://www.zdnet.com/cryptolockers-crimewave-a-trail-of-millions-in-laundered-bitcoin-7000024579/>

³<http://www.theguardian.com/technology/2013/nov/21/us-police-force-pay-bitcoin-ransom-in-cryptolocker-malware-scam>

The bot **encrypts** files on the victim's computer and asks for a **ransom** to recover them.

- First appeared in early September 2013

¹<http://www.cybersec.kent.ac.uk/Survey2.pdf>

²<http://www.zdnet.com/>

cryptolockers-crimewave-a-trail-of-millions-in-laundered-bitcoin-7000024579/

³<http://www.theguardian.com/technology/2013/nov/21/us-police-force-pay-bitcoin-ransom-in-cryptolocker-malware-scam>

The bot **encrypts** files on the victim's computer and asks for a **ransom** to recover them.

- ▶ First appeared in early September 2013
- ▶ In the UK 41% victims paid the ransom¹

¹<http://www.cybersec.kent.ac.uk/Survey2.pdf>

²<http://www.zdnet.com/>

[cryptolockers-crimewave-a-trail-of-millions-in-laundered-bitcoin-7000024579/](http://www.zdnet.com/cryptolockers-crimewave-a-trail-of-millions-in-laundered-bitcoin-7000024579/)

³<http://www.theguardian.com/technology/2013/nov/21/us-police-force-pay-bitcoin-ransom-in-cryptolocker-malware-scam>

The bot **encrypts** files on the victim's computer and asks for a **ransom** to recover them.

- ▶ First appeared in early September 2013
- ▶ In the UK 41% victims paid the ransom¹
- ▶ Earnings estimated at 27 million USD on Dec 18 2013²

¹<http://www.cybersec.kent.ac.uk/Survey2.pdf>

²<http://www.zdnet.com/>

cryptolockers-crimewave-a-trail-of-millions-in-laundered-bitcoin-7000024579/

³<http://www.theguardian.com/technology/2013/nov/21/us-police-force-pay-bitcoin-ransom-in-cryptolocker-malware-scam>

The bot **encrypts** files on the victim's computer and asks for a **ransom** to recover them.

- ▶ First appeared in early September 2013
- ▶ In the UK 41% victims paid the ransom¹
- ▶ Earnings estimated at 27 million USD on Dec 18 2013²
- ▶ Massachusetts police have admitted to paying a ransom³

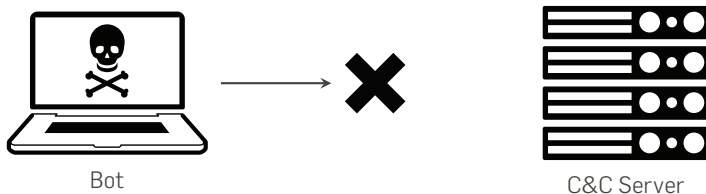
¹<http://www.cybersec.kent.ac.uk/Survey2.pdf>

²<http://www.zdnet.com/>

cryptolockers-crimewave-a-trail-of-millions-in-laundered-bitcoin-7000024579/

³<http://www.theguardian.com/technology/2013/nov/21/us-police-force-pay-bitcoin-ransom-in-cryptolocker-malware-scam>

CENTRALIZED BOTNETS > MITIGATION



- ▶ **C&C channel:** single point of failure.
- ▶ **Rallying Mechanisms:** the countermeasure.

BOTNETS > DOMAIN GENERATION ALGORITHMS

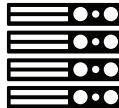
C&C Server, sjq.info



Bot



DNS Resolver



BOTNETS > DOMAIN GENERATION ALGORITHMS

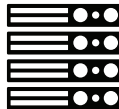
C&C Server, **sjq.info**



Bot



DNS Resolver



DNS query: **ahj.info**



BOTNETS > DOMAIN GENERATION ALGORITHMS

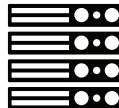
C&C Server, **sjq.info**



Bot



DNS Resolver



DNS query: **ahj.info**

DNS reply: **NXDOMAIN**

BOTNETS > DOMAIN GENERATION ALGORITHMS

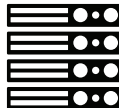
C&C Server, **sjq.info**



Bot



DNS Resolver



DNS query: **ahj.info**

DNS reply: **NXDOMAIN**

DNS query: **sjq.info**

BOTNETS > DOMAIN GENERATION ALGORITHMS

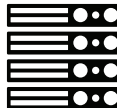
C&C Server, **sjq.info**



Bot



DNS Resolver



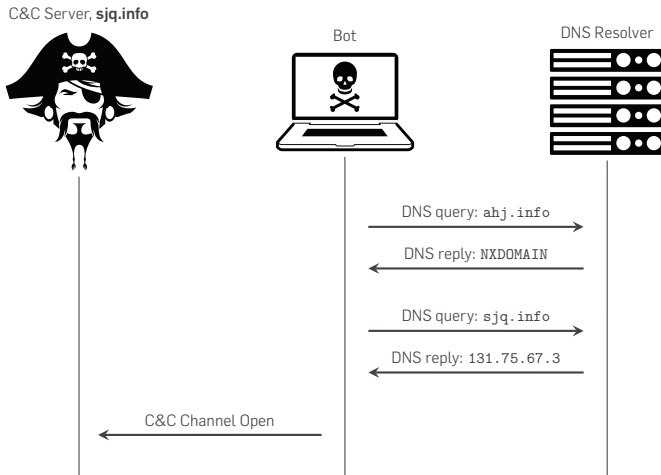
DNS query: **ahj.info**

DNS reply: NXDOMAIN

DNS query: **sjq.info**

DNS reply: **131.75.67.3**

BOTNETS > DOMAIN GENERATION ALGORITHMS



- ▶ **Asymmetry** Botmasters Vs Defenders
 - Thousands of domain names,
 - only one is the right one.
- ▶ **Blacklists** do not work well

Limitations of current **research approaches**:

Limitations of current **research approaches**:

- ▶ **Supervised:** require labeled data

Limitations of current **research approaches**:

- ▶ **Supervised:** require labeled data
 - "That domain name is known to be DGA generated",
 - "That other domain is not".

Limitations of current **research approaches**:

- ▶ **Supervised:** require labeled data
 - "That domain name is known to be DGA generated",
 - "That other domain is not".
- ▶ Work at the **lower levels** of the **DNS hierarchy**:
 - not so easy to deploy,
 - privacy (visibility of the hosts' IP addresses).

PHOENIX



Phoenix clusters

DGA-generated domains from a list of of **domains known to be used by botnets**.

The core of Phoenix is its ability to **separate DGA from non-DGA** domains, using **linguistic features**.

(in a few slides)



Sources of malicious domains:

- ▶ **EXPOSURE** <http://exposure.iseclab.org>
- ▶ **MLD** <http://www.malwaredomainlist.com>
- ▶ ...and of course some **reversing** :-)

Meaningful Word Ratio (English dict)

$d = \text{facebook.com}$

$$R(d) = \frac{|\text{face}| + |\text{book}|}{|\text{facebook}|} = 1$$

likely **non-DGA** generated

$d = \text{pub03str.info}$

$$R(d) = \frac{|\text{pub}|}{|\text{pub03str}|} = 0.375.$$

likely **DGA** generated

N-gram Popularity (English dict)

$d = \text{facebook.com}$

fa	ac	ce	eb	bo	oo	ok
109	343	438	29	118	114	45

mean: $S_2 = 170.8$

likely **non-DGA** geneated

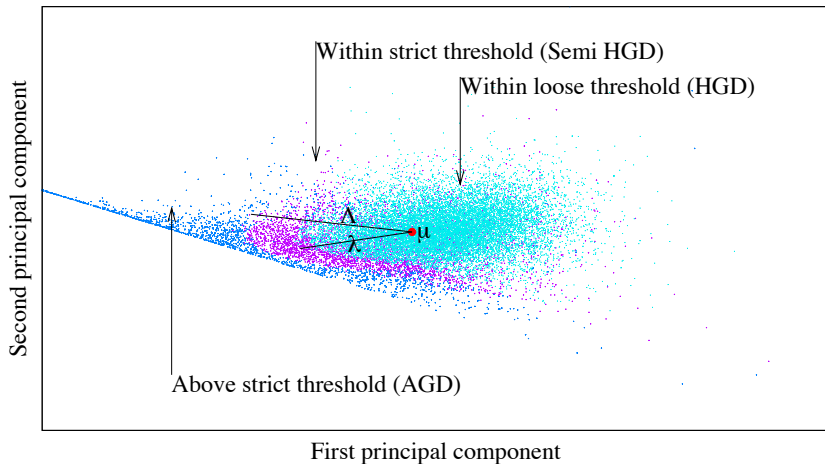
$d = \text{aawrqv.com}$

aa	aw	wr	rq	qv
4	45	17	0	0

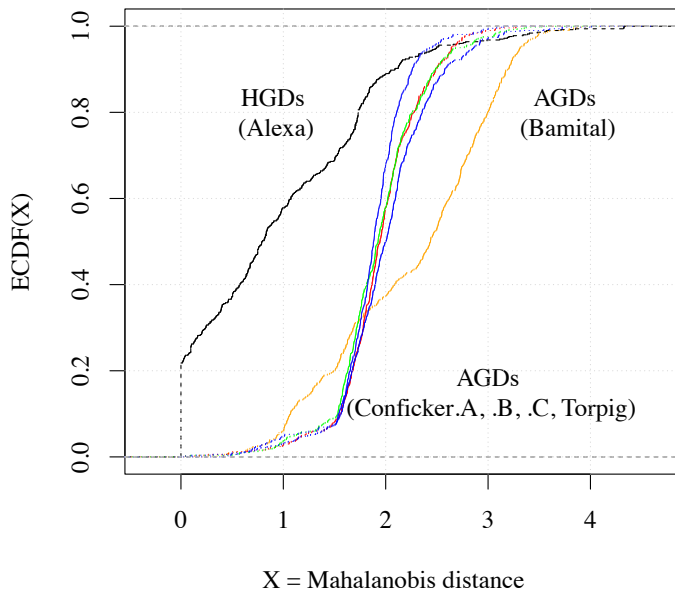
mean: $S_2 = 13.2$

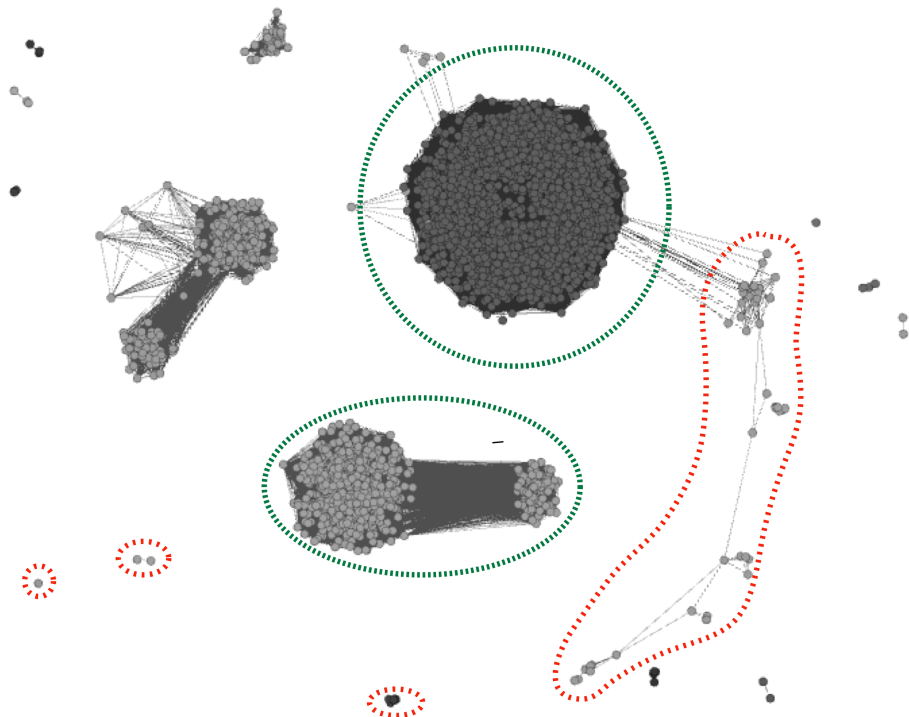
likely **DGA** generated

PHOENIX > DGA VS NON-DGA



PHOENIX > BOTNETS





PHOENIX > RESULTS (1 WEEK)

Cluster f105c

IPs: 176.74.176.175
208.87.35.107

Domains: cvq.com
epu.org
bwn.org

(Botnet: Palevo)

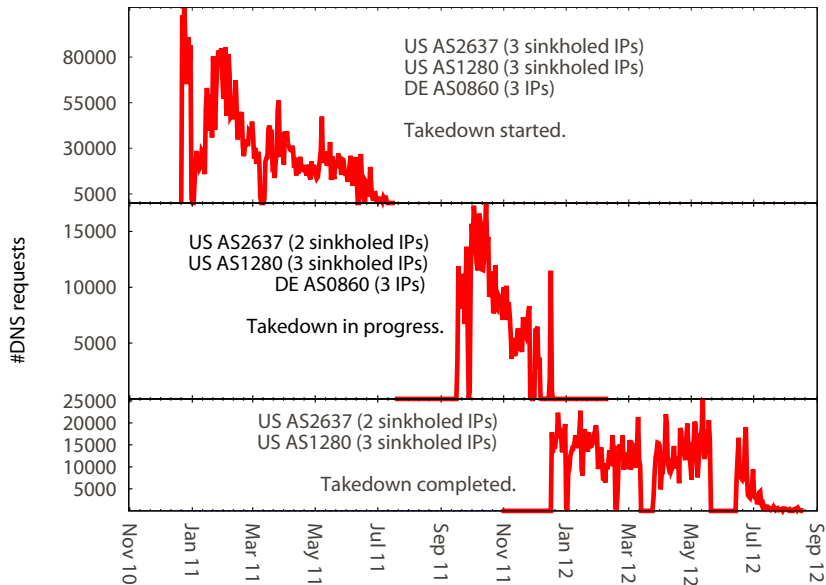
Cluster 0f468

IPs: 217.119.57.22
91.215.158.57
178.162.164.24
94.103.151.195

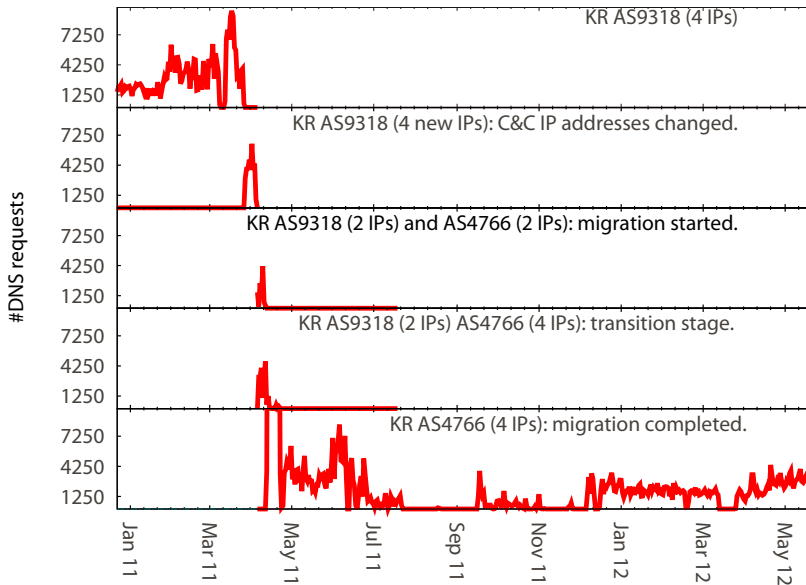
Domains: jhhfghf7.tk
faukiijjj25.tk
pvgvy.tk

(Botnet: Sality)

PHOENIX > TRACKING MIGRATIONS



PHOENIX > TRACKING MIGRATIONS



A group of people dressed in black robes, resembling Hogwarts students, are standing on a large outdoor chessboard. The chessboard is painted on a light-colored path with red and white squares, and is set on a green lawn. In the background is a grand, ornate red brick building with a prominent clock tower and many windows. The sky is clear and blue.

Meet you at Hogwarts Royal Holloway in July.

DIMVA 2014 <http://dimva2014.isg.rhul.ac.uk/>

S. Schiavoni, F. Maggi, L. Cavallaro, S. Zanero.

Phoenix: DGA-Based Botnet Tracking and Intelligence.

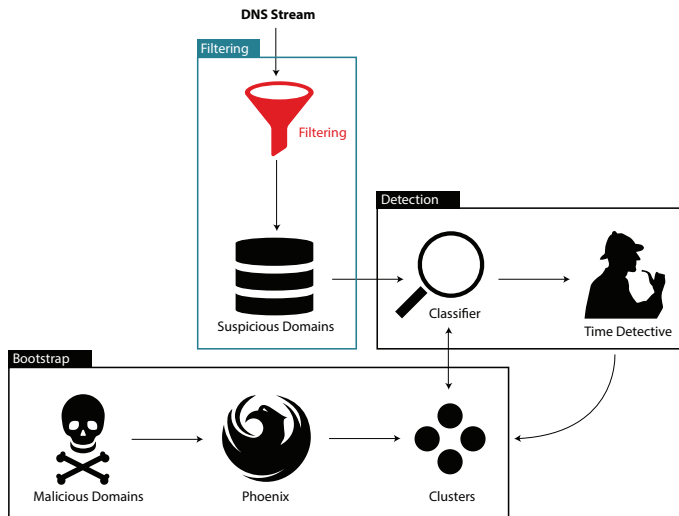
Proceedings of DIMVA2014.

Leverages historical DNS data:

- ▶ **Unable** to deal with **new DGAs**
- ▶ Unseen "domain→IP" mapping are simply **discarded**.

CERBERUS

CERBERUS > FILTERING



Insight a malicious domain automatically generated will not become popular.

Alexa Top 1M Whitelist

We whitelist the domains that appear in the Alexa Top 1M.

Insight a malicious domain automatically generated will not belong to a CDN `r4---sn-a5m7lnes.example.com`.

CDN Whitelist

We whitelist the domains that belong to the most popular CDN networks (e.g., YouTube, Google, etc.) and advertisement services.

Insight an attacker will register a domain with a TLD that does not require clearance.

TLD Whitelist

We whitelist the domains featuring a Top Level Domain that requires authorization by a third party authority before registration (e.g. .gov, .edu, .mil).

Insight How fast is fast?

- ▶ 2-3 years ago: $TTL < 100$.
- ▶ Nowadays: $80 < TTL < 300$ seconds.

Why? To save money :-) See BH-US 2013 talk⁴.

TTL

We filter out all those domains featuring a Time To Live outside these bounds.

⁴<https://media.blackhat.com/us-13/US-13-Xu-New-Trends-in-FastFlux-Networks-Slides.pdf>

Insight we are looking for DGA-generated domains.

Phoenix's DGA Filter

We filter out domains likely to be generated by humans.

Insight the attacker will register the domain just a few days before the communication will take place.

Whois

We query the Whois server and discard the domains that were registered more than Δ days before the DNS query.

RECAP ON FILTERING

Starting with 50,000 domains:

20,000 **TTL > 300** seconds;

19,000 **not** in the **Alexa Top 1M** list;

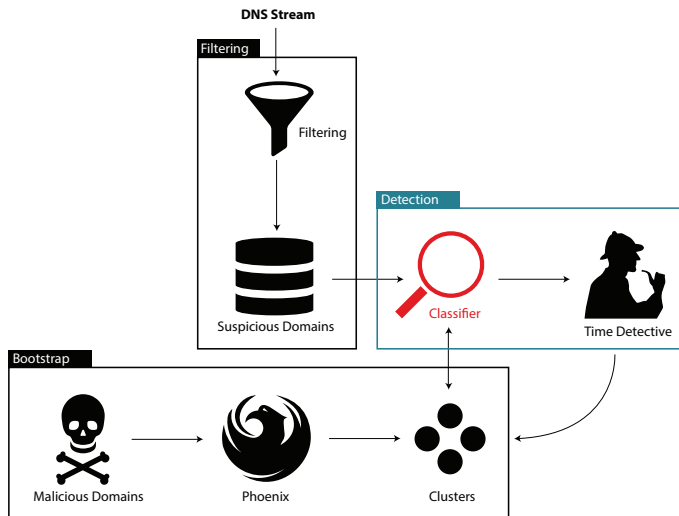
15,000 **not** in the most popular **CDNs**;

800 **likely** to be **DGA generated**;

700 **no** previous **authorization**;

300 **younger than** Δ days $\leftarrow$ suspicious.

CERBERUS > FILTERING



CLASSIFIER > CLASSIFICATION

Cluster A

69.43.161.180

379.ns4000wip.com

418.ns4000wip.com

285.ns4000wip.com

Cluster B

69.43.161.180

391.wap517.net

251.wap517.net

340.wap517.net

Cluster C

...

576.wap517.net

69.43.161.180

CLASSIFIER > CLASSIFICATION

Cluster A

69.43.161.180
379.ns4000wip.com
418.ns4000wip.com
285.ns4000wip.com

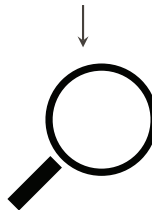
Cluster B

69.43.161.180
391.wap517.net
251.wap517.net
340.wap517.net

Cluster C

...

576.wap517.net
69.43.161.180



CLASSIFIER > CLASSIFICATION

Cluster A

69.43.161.180
379.ns4000wip.com
418.ns4000wip.com
285.ns4000wip.com

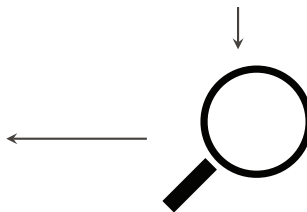
Cluster B

69.43.161.180
391.wap517.net
251.wap517.net
340.wap517.net

Cluster C

...

576.wap517.net
69.43.161.180



CLASSIFIER > CLASSIFICATION

Cluster A

69.43.161.180

379.ns4000wip.com

418.ns4000wip.com

285.ns4000wip.com

Cluster B

69.43.161.180

391.wap517.net

251.wap517.net

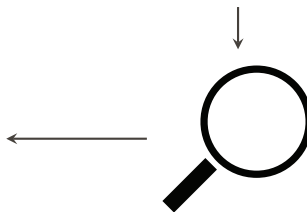
340.wap517.net

Cluster C

...

576.wap517.net

69.43.161.180



CLASSIFIER > CLASSIFICATION

Cluster A

69.43.161.180

379.ns4000wip.com

418.ns4000wip.com

285.ns4000wip.com

Cluster B

69.43.161.180

391.wap517.net

251.wap517.net

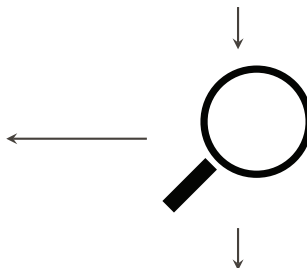
340.wap517.net

Cluster C

...

576.wap517.net

69.43.161.180



Train the Classifier on A, B

CLASSIFIER > CLASSIFICATION

Cluster A

69.43.161.180

379.ns4000wip.com

418.ns4000wip.com

285.ns4000wip.com

Cluster B

69.43.161.180

391.wap517.net

251.wap517.net

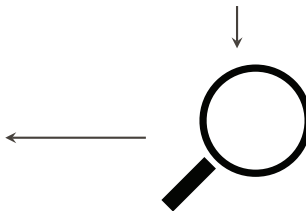
340.wap517.net

Cluster C

...

576.wap517.net

69.43.161.180



Train the Classifier on A, B

Assign 576.wap517.net to B

CLASSIFIER > SUBSEQUENCE STRING KERNEL

Developed at Royal Holloway in 2002, by Lodhi et al.

	c-a	c-t	a-t	c-r	a-r
$\phi(\text{cat})$	λ^2	λ^3	λ^2	0	0
$\phi(\text{car})$	λ^2	0	0	λ^3	λ^2

How many substrings of size $k = 2$?

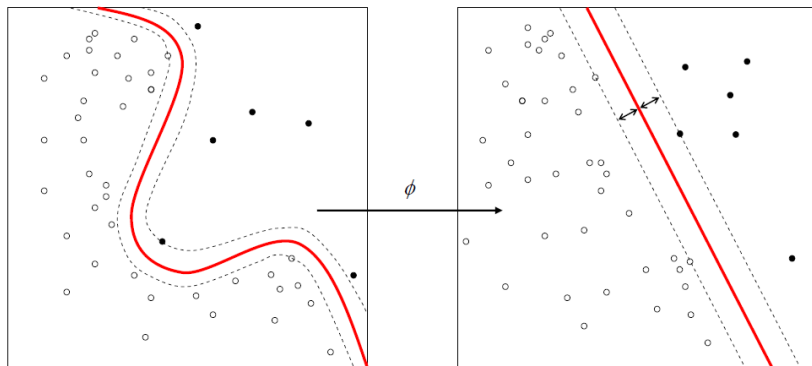
$$\ker(\text{car}, \text{cat}) = \lambda^4$$

$$\ker(\text{car}, \text{car}) = \ker(\text{cat}, \text{cat}) = 2\lambda^4 + \lambda^6$$

$$\ker_n(\text{car}, \text{cat}) = \frac{\lambda^4}{(2\lambda^4 + \lambda^6)} = \frac{1}{(2 + \lambda^2)} \in [0, 1]$$

CLASSIFIER > SUPPORT VECTOR MACHINES

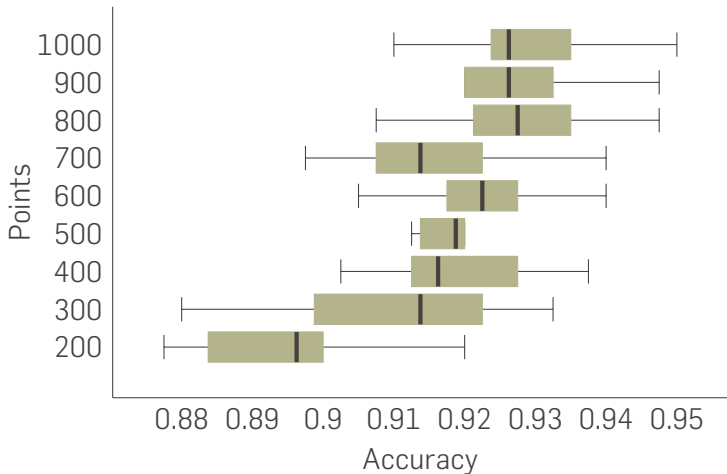
SVM: find one hyperplane or a set of them that has the largest distance to the nearest training data point of any class



RESULTS

on passive DNS data from
<https://farsightsecurity.com/Services/SIE/>

RESULTS > CLASSIFIER



CLASSIFICATION > RESULTS

Training 1000, Testing 100

Overall Accuracy $\simeq$ 0.95

	a	b	c	d
a	100	0	0	0
b	1	92	6	1
c	2	0	98	0
d	3	0	6	91

a

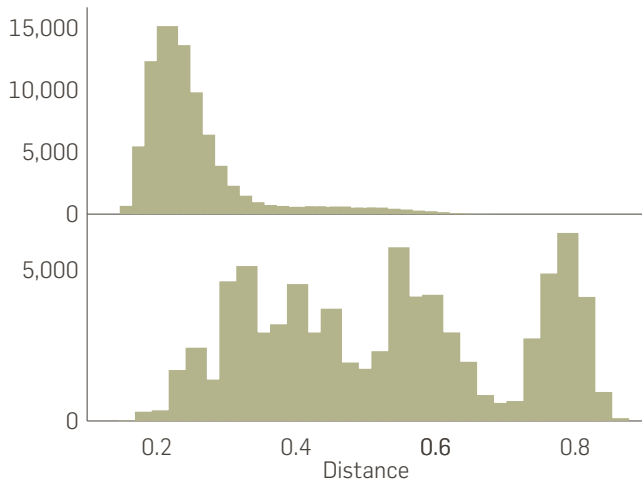
caaa89e...d4ca925b3e2.co.cc
f1e01ac...51b64079d86.co.cc
b

kdnvfyc.biz
wapzzwvpwq.info
c

jhhfghf7.tk
faukiijjj25.tk
d

cvq.com
epu.org

CLASSIFICATION > PAIRWISE DISTANCES





The **Time Detective** discovers new botnets.

Every Δ the bots **contact** the C&C Server, on a **new domain**.



Bot

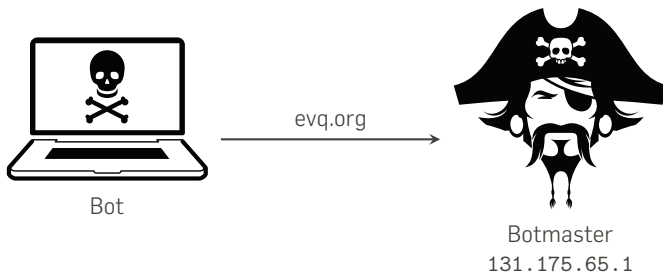


Botmaster
131.175.65.1

131.175.65.1: { }

TIME DETECTIVE > PASSIVE DNS TRAFFIC

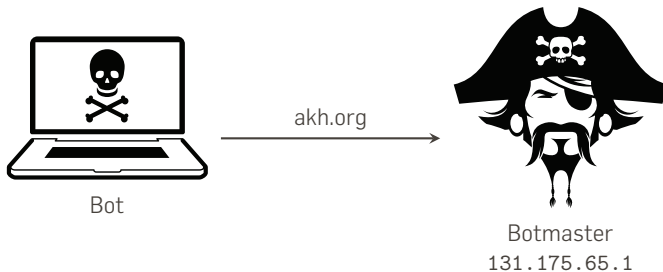
Every Δ the bots **contact** the C&C Server, on a **new domain**.



131.175.65.1: { evq.org }

TIME DETECTIVE > PASSIVE DNS TRAFFIC

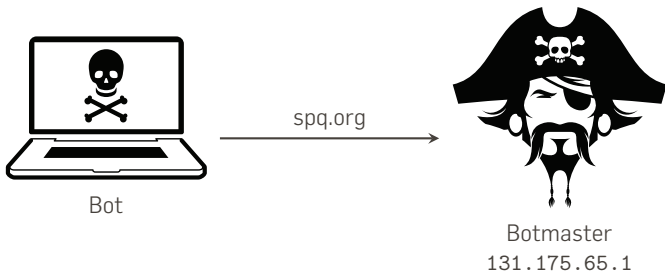
Every Δ the bots **contact** the C&C Server, on a **new domain**.



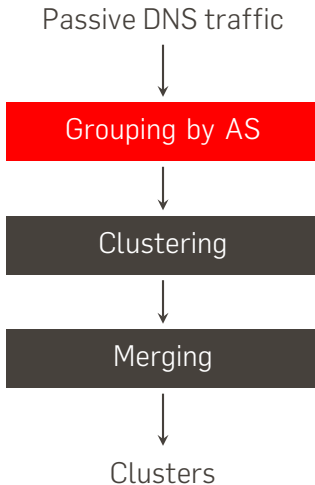
131.175.65.1: [evq.org , akh.org]

TIME DETECTIVE > PASSIVE DNS TRAFFIC

Every Δ the bots **contact** the C&C Server, on a **new domain**.



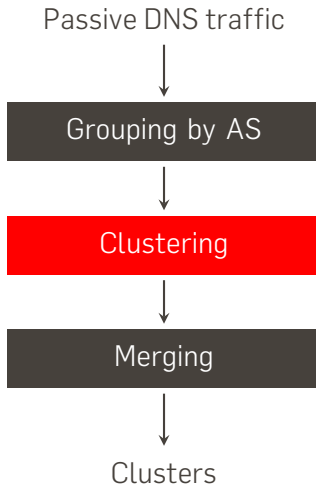
131.175.65.1: [evq.org , akh.org , spq.org]



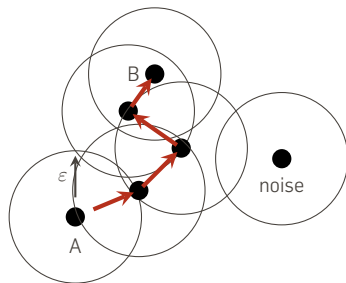


We assume a **lazy attacker** behavior: If (s)he finds an obliging AS, (s)he will buy a few IPs in there.

We group together the domains that point to IPs within the **same AS**.



DBSCAN



SSK as the distance

automatic tuning:

- ▶ *minPts* domains per cluster,
- ▶ ϵ distance threshold.

$minPts = 7$ domains per cluster

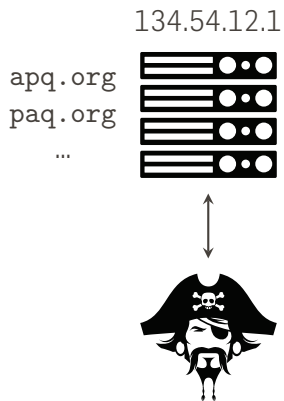
Observation period in days.

Rationale: the bots will contact the C&C server at least **once a day**.

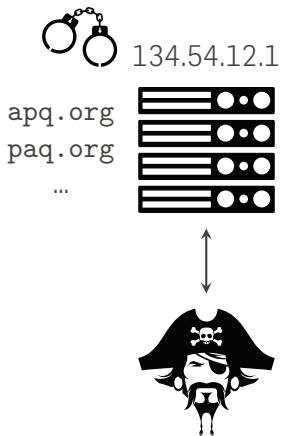
$$\frac{\text{intra-cluster distances}}{\text{inter-cluster distances}} \rightarrow 0 \text{ (minimize)}$$

What if a new cluster is actually a **known botnet** that **migrated** the C&C server somewhere else?

TIME DETECTIVE > MERGING



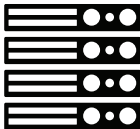
TIME DETECTIVE > MERGING



TIME DETECTIVE > MERGING



134.54.12.1

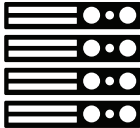


What t' h3ck!

TIME DETECTIVE > MERGING



134.54.12.1



Migration

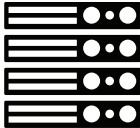


What t' h3ck!

TIME DETECTIVE > MERGING



134.54.12.1

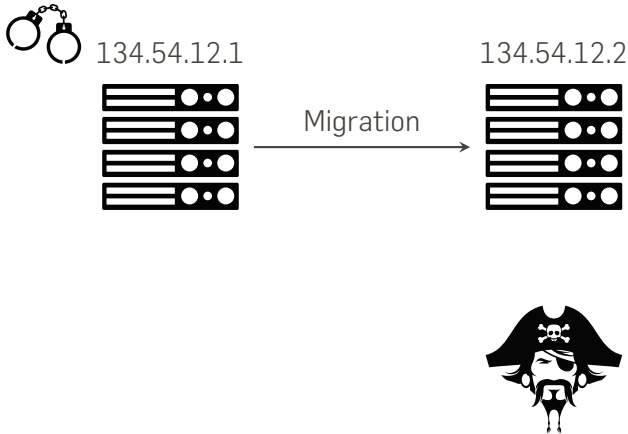


Migration

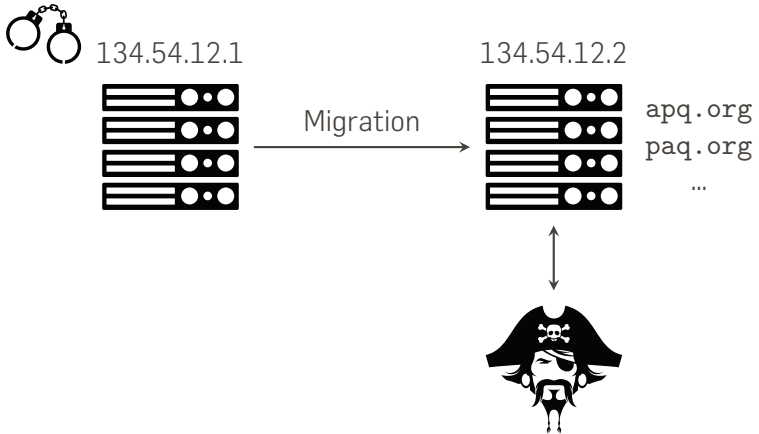


What t' h3ck!

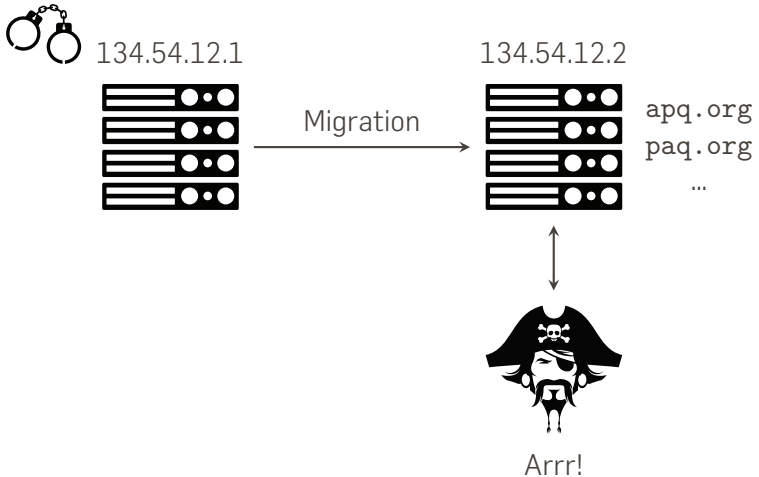
TIME DETECTIVE > MERGING

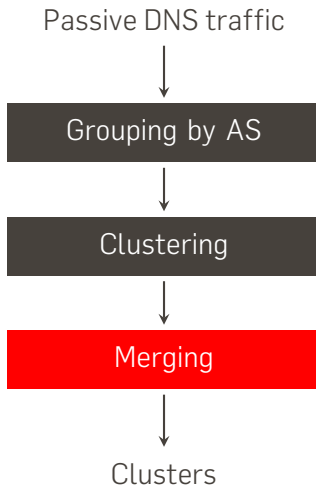


TIME DETECTIVE > MERGING



TIME DETECTIVE > MERGING





Suppose you have cluster A and B.

Suppose you have cluster A and B.

$$A = \begin{matrix} & \text{dom}_1 & \cdots & \text{dom}_m \\ \text{dom}_1 & d_{1,1} & \cdots & d_{1,m} \\ \text{dom}_2 & d_{2,1} & \cdots & d_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \text{dom}_m & d_{m,1} & \cdots & d_{m,m} \end{matrix} \left(\begin{array}{ccc} d_{1,1} & \cdots & d_{1,m} \\ d_{2,1} & \cdots & d_{2,m} \\ \vdots & \ddots & \vdots \\ d_{m,1} & \cdots & d_{m,m} \end{array} \right)$$

Suppose you have cluster A and B.

$$A = \begin{array}{c} \text{dom}_1 \\ \text{dom}_2 \\ \vdots \\ \text{dom}_m \end{array} \begin{pmatrix} \text{dom}_1 & \cdots & \text{dom}_m \\ d_{1,1} & \cdots & d_{1,m} \\ d_{2,1} & \cdots & d_{2,m} \\ \vdots & \ddots & \vdots \\ d_{m,1} & \cdots & d_{m,m} \end{pmatrix} \quad B = \begin{array}{c} \text{dom}_1 \\ \text{dom}_2 \\ \vdots \\ \text{dom}_n \end{array} \begin{pmatrix} \text{dom}_1 & \cdots & \text{dom}_n \\ d_{1,1} & \cdots & d_{1,n} \\ d_{2,1} & \cdots & d_{2,n} \\ \vdots & \ddots & \vdots \\ d_{n,1} & \cdots & d_{n,n} \end{pmatrix}$$

Suppose you have cluster A and B.

$$A = \begin{matrix} & \text{dom}_1 & \cdots & \text{dom}_m \\ \text{dom}_1 & d_{1,1} & \cdots & d_{1,m} \\ \text{dom}_2 & d_{2,1} & \cdots & d_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \text{dom}_m & d_{m,1} & \cdots & d_{m,m} \end{matrix} \quad B = \begin{matrix} & \text{dom}_1 & \cdots & \text{dom}_n \\ \text{dom}_1 & d_{1,1} & \cdots & d_{1,n} \\ \text{dom}_2 & d_{2,1} & \cdots & d_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \text{dom}_n & d_{n,1} & \cdots & d_{n,n} \end{matrix}$$

$$A \sim B = \begin{matrix} & \text{dom}_1 & \text{dom}_2 & \cdots & \text{dom}_n \\ \text{dom}_1 & d_{1,1} & d_{1,2} & \cdots & d_{1,n} \\ \text{dom}_2 & d_{2,1} & d_{2,2} & \cdots & d_{2,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \text{dom}_m & d_{m,1} & d_{m,2} & \cdots & d_{m,n} \end{matrix}$$

Stats to the rescue!

TIME DETECTIVE > WELCH TEST

Stats to the rescue!

$$A = \begin{matrix} & \text{dom}_1 & \cdots & \text{dom}_m \\ \text{dom}_1 & d_{1,1} & \cdots & d_{1,m} \\ \text{dom}_2 & d_{2,1} & \cdots & d_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \text{dom}_m & d_{m,1} & \cdots & d_{m,m} \end{matrix} \quad A \sim B = \begin{matrix} & \text{dom}_1 & \text{dom}_2 & \cdots & \text{dom}_n \\ \text{dom}_1 & d_{1,1} & d_{1,2} & \cdots & d_{1,n} \\ \text{dom}_2 & d_{2,1} & d_{2,2} & \cdots & d_{2,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \text{dom}_m & d_{m,1} & d_{m,2} & \cdots & d_{m,n} \end{matrix}$$

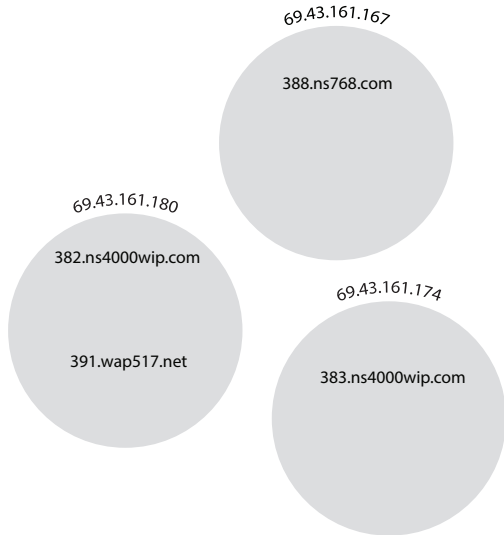
Stats to the rescue!

$$A = \begin{matrix} & \text{dom}_1 & \cdots & \text{dom}_m \\ \text{dom}_1 & d_{1,1} & \cdots & d_{1,m} \\ \text{dom}_2 & d_{2,1} & \cdots & d_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \text{dom}_m & d_{m,1} & \cdots & d_{m,m} \end{matrix} \quad A \sim B = \begin{matrix} & \text{dom}_1 & \text{dom}_2 & \cdots & \text{dom}_n \\ \text{dom}_1 & d_{1,1} & d_{1,2} & \cdots & d_{1,n} \\ \text{dom}_2 & d_{2,1} & d_{2,2} & \cdots & d_{2,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \text{dom}_m & d_{m,1} & d_{m,2} & \cdots & d_{m,n} \end{matrix}$$

Welch test: do A and $A \sim B$ have different intra-cluster distance distributions?

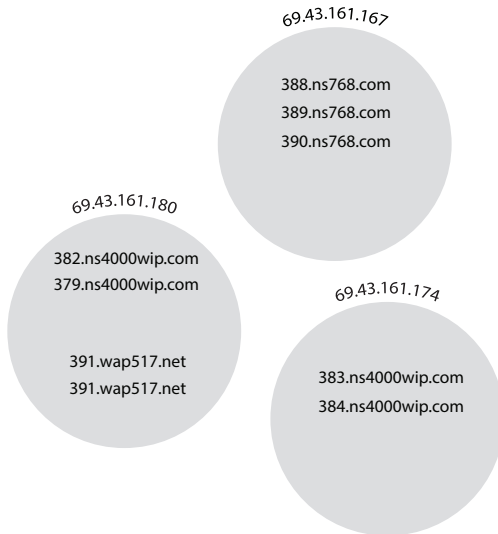
TIME DETECTIVE > EXAMPLE

Day 1



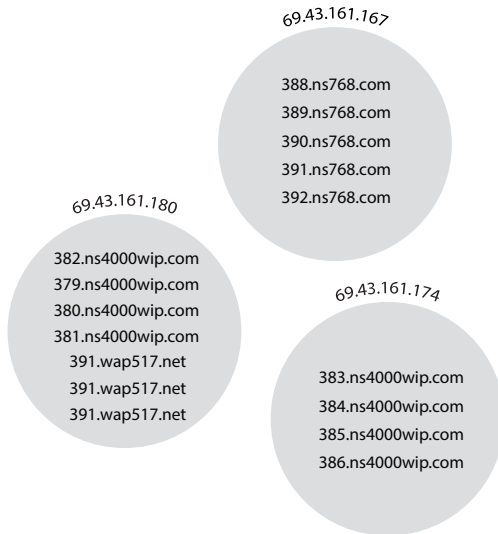
TIME DETECTIVE > EXAMPLE

Day 2



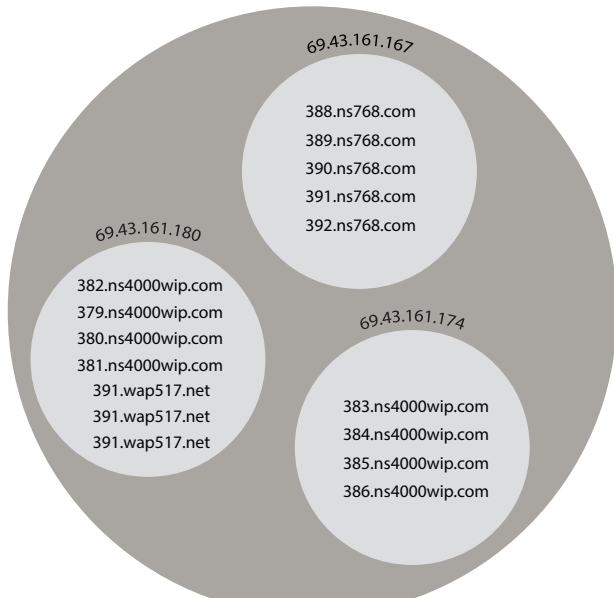
TIME DETECTIVE > EXAMPLE

Day 7



TIME DETECTIVE > EXAMPLE

AS 22489



Merge

382.ns4000wip.com
379.ns4000wip.com
380.ns4000wip.com
381.ns4000wip.com
391.wap517.net
391.wap517.net
391.wap517.net

388.ns768.com
389.ns768.com
390.ns768.com
391.ns768.com
392.ns768.com

383.ns4000wip.com
384.ns4000wip.com
385.ns4000wip.com
386.ns4000wip.com

Cluster

388.ns768.com

389.ns768.com

390.ns768.com

391.ns768.com

392.ns768.com

382.ns4000wip.com

379.ns4000wip.com

380.ns4000wip.com

381.ns4000wip.com

383.ns4000wip.com

384.ns4000wip.com

385.ns4000wip.com

386.ns4000wip.com

391.wap517.net

391.wap517.net

391.wap517.net

New clusters
produced

Cluster 1

388.ns768.com
389.ns768.com
390.ns768.com
391.ns768.com
392.ns768.com

Cluster 2

382.ns4000wip.com
379.ns4000wip.com
380.ns4000wip.com
381.ns4000wip.com

383.ns4000wip.com
384.ns4000wip.com
385.ns4000wip.com
386.ns4000wip.com

Cluster 3

391.wap517.net
391.wap517.net
391.wap517.net

RESULTS

on passive DNS data from
<https://farsightsecurity.com/Services/SIE/>

187 domains classified as malicious and **labeled**.

Labeled 07e21

Botnet: Conficker

Domains: hhdboqazof.biz
poxqmrj.fj.biz
hcsddszzzc.ws
tnoucgrje.biz
gwizoxej.biz
jnmuoiki.biz

3,576 domains were considered **suspicious** by Cerberus and **stored**, together with their IP address.

Then we ran the clustering routine to **discover new botnets**.

TIME DETECTIVE > CLUSTERING

Botnet	AS	IPs	Size
Sality	15456	62.116.181.25	26
Palevo	53665	199.59.243.118	40
Jadtre*	22489	69.43.161.180	173
		69.43.161.174	
Jadtre**	22489	69.43.161.180	37
Jadtre***	22489	69.43.161.167	47
Hiloti	22489	69.43.161.167	24
Palevo	47846	82.98.86.171	142
		82.98.86.176	
		82.98.86.175	
Jusabli	30069	69.58.188.49	73
Generic Trojan	12306	82.98.86.169	57
		82.98.86.162	
		82.98.86.178	
		82.98.86.163	

Cluster	IP	Sample Domains
Jadtres*	69.43.161.180	379.ns4000wip.com
	69.43.161.174	418.ns4000wip.com
		285.ns4000wip.com
Jadtres**	69.43.161.180	391.wap517.net
		251.wap517.net
		340.wap517.net
Jadtres***	69.43.161.167	388.ns768.com
		353.ns768.com
		296.ns768.com

TIME DETECTIVE > MERGING

Cluster a (Old)

IPs: 176.74.76.175
 208.87.35.107

Domains cvq.com
 epu.org
 bwn.org
 lxx.net

Cluster b (New)

IPs: 82.98.86.171
 82.98.86.176
 82.98.86.175
 82.98.86.167
 82.98.86.168
 82.98.86.165

Domains knw.info
 rrg.info
 nhv.org
 ydt.info

Cluster a (Old)		Cluster b (New)	
IPs:	176.74.76.175	IPs:	82.98.86.171
	208.87.35.107		82.98.86.176
			82.98.86.175
			82.98.86.167
			82.98.86.168
			82.98.86.165
Domains	cvq.com	Domains	knw.info
	epu.org		rrg.info
	bwn.org		nhy.org
	lxx.net		ydt.info

Both belonging to the **Palevo botnet**.

- ▶ **187** malicious domains **detected and labeled**

- ▶ **187** malicious domains **detected and labeled**
- ▶ **3,576 suspicious** domains collected

- ▶ **187** malicious domains **detected and labeled**
- ▶ **3,576 suspicious** domains collected
- ▶ **47 clusters** of DGA-generated domains **discovered**

- ▶ **187** malicious domains **detected and labeled**
- ▶ **3,576 suspicious** domains collected
- ▶ **47 clusters** of DGA-generated domains **discovered**
- ▶ **319** new domains **detected in the next 24 hours**

CONCLUSIONS & FUTURE WORK



- ▶ discovers and characterizes unknown DGA-based activity,
- ▶ unsupervised,
- ▶ easy to deploy,
- ▶ privacy preserving.

FUTURE WORK

this-is-an-easy-way-to-evade-the-linguistic-filter.com



Release **Cerberus** as a web service. Hopefully!

THANK YOU

`federico.maggi@polimi.it`
`@phretor`