Graph-Theoretic Characterization of Cyber-threat Infrastructures

By

Amine Boukhtouta, Djedjiga Mouheb, Mourad Debbabi, Omar Alfandi, Farkhund Iqbal and May El Barachi

Presented At

The Digital Forensic Research Conference

DFRWS 2015 USA Philadelphia, PA (Aug 9th - 13th)

DFRWS is dedicated to the sharing of knowledge and ideas about digital forensics research. Ever since it organized the first open workshop devoted to digital forensics in 2001, DFRWS continues to bring academics and practitioners together in an informal environment. As a non-profit, volunteer organization, DFRWS sponsors technical working groups, annual conferences and challenges to help drive the direction of research and development.

http:/dfrws.org

National Cyber-Forensics
and Training Alliance CANADA

GRAPH-THEORETIC CHARACTERIZATION OF CYBER-THREAT INFRASTRUCTURES

PRESENTED BY AMINE BOUKHTOUTA

Agenda

- Introduction
- Approach
- Cyber-threat infrastructures
- Statistics and insights
- Conclusion

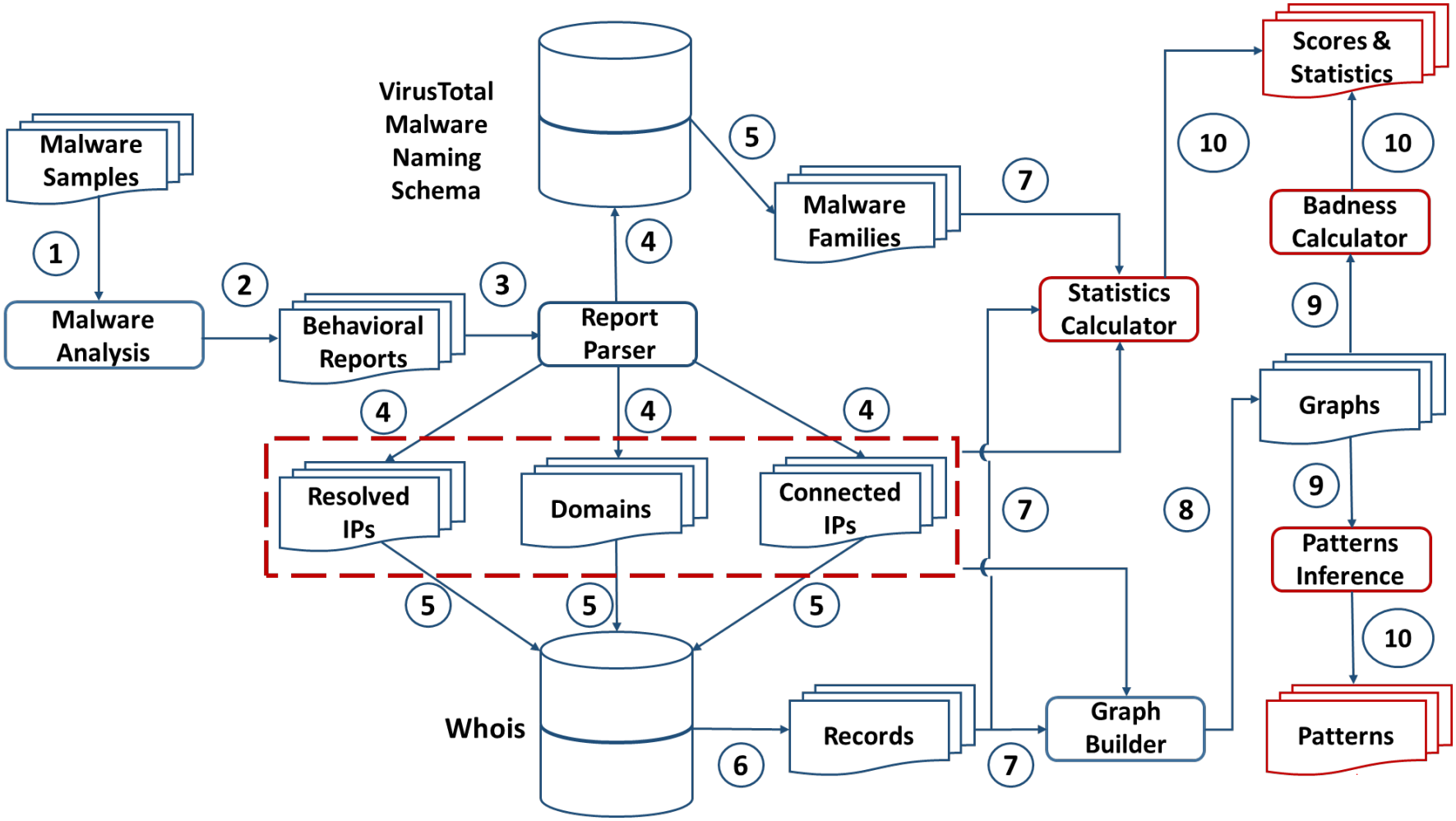
Introduction

- Cyber-space:
  - A continuous fight between hackers and security experts
  - Sophistication of threats and malware: Shellshock vulnerability, appearance of new variants of an old malware, namely, “Kaiten”. It targets unpatched shells of Linux and Mac Operating Systems
  - Events: 16th-31st of July 2015, at least 31 events related to cyber-space security were enumerated (hackmageddon.com)

Objectives

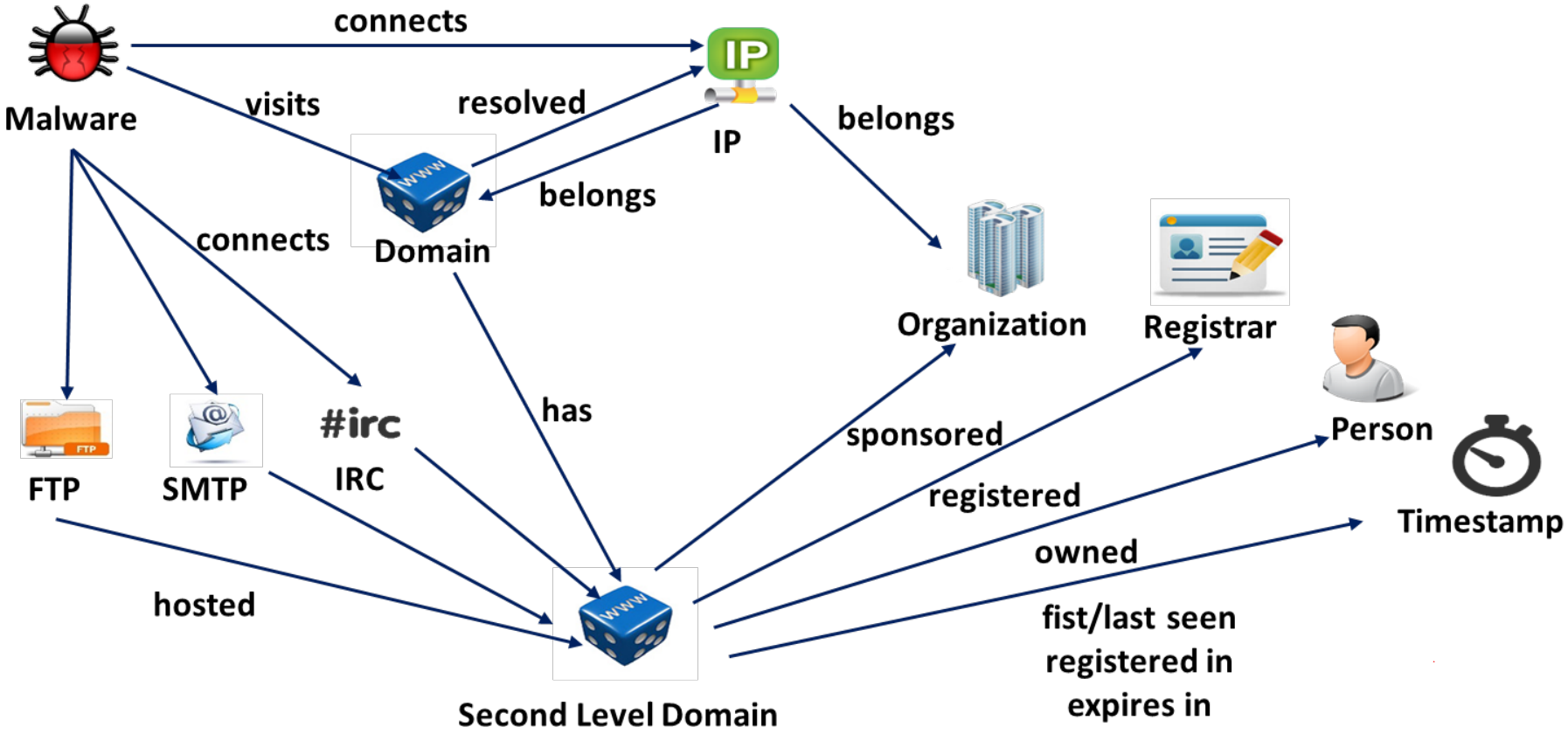
- What are the elements of a cyber-threat infrastructure and what are the relationships between them?
- What are the infrastructures used by cyber criminals to perpetrate attacks?
- What are the most important players in cyber-threat infrastructures (e.g., owners, domains, IPs)?
- How cyber-threat infrastructures evolve?

Approach


Malware Samples
1
Malware Analysis
2
Behavioral Reports
3
Report Parser
4
VirusTotal Malware Naming Schema
5
Malware Families
7
Statistics Calculator
10
Scores & Statistics
10
Badness Calculator
9
Graphs
9
Patterns Inference
10
Patterns
4
4
4
Resolved IPs
Domains
Connected IPs
5
5
5
7
8
Whois
6
Records
7
Graph Builder

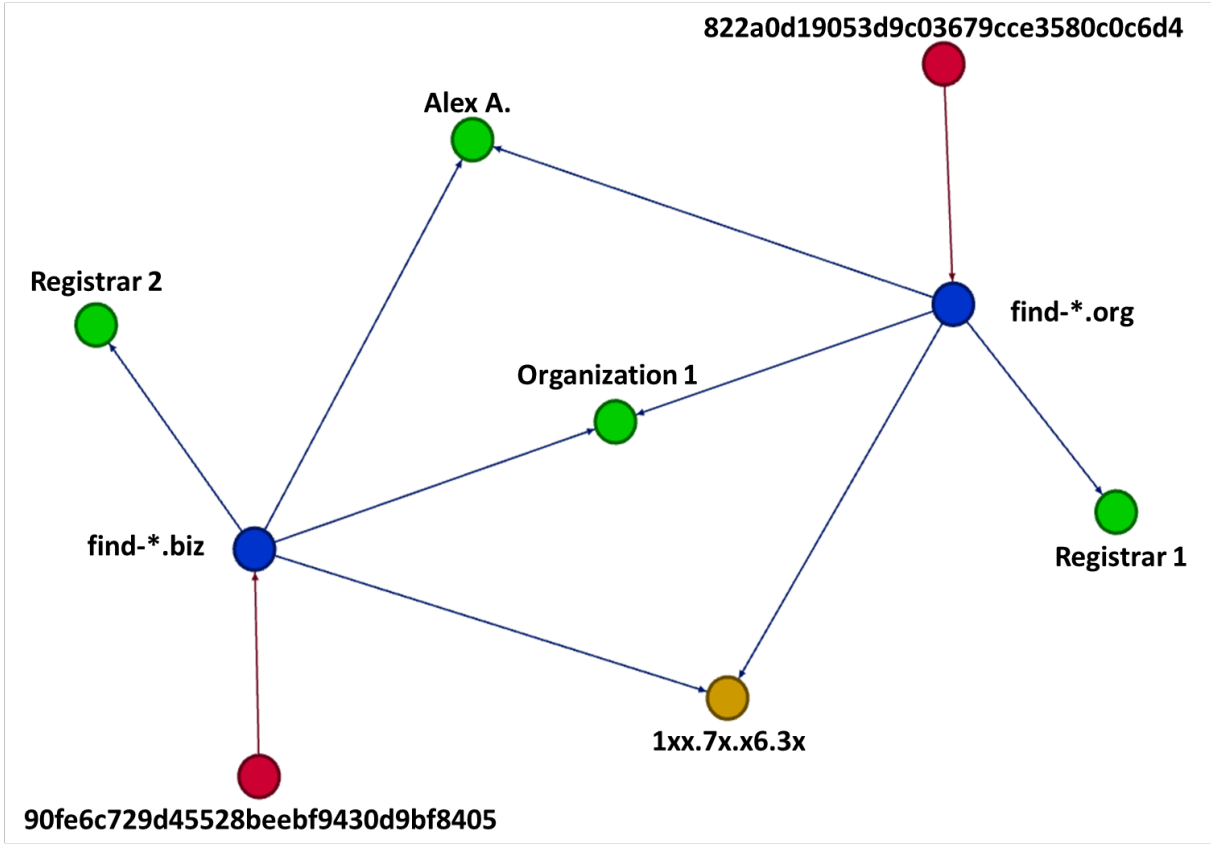
Cyber-threat Infrastructures

- What are the elements of a cyber-threat infrastructure and what are the relationships between them?



Cyber-threat Infrastructures

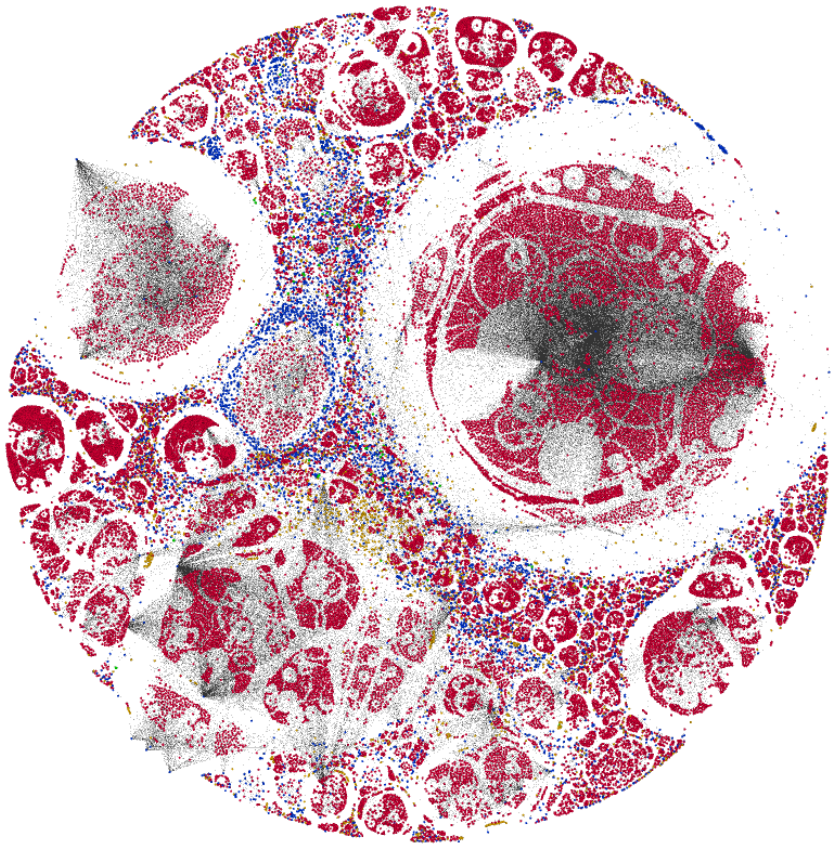
- Example of a cyber-threat infrastructure



Cyber-threat Infrastructures

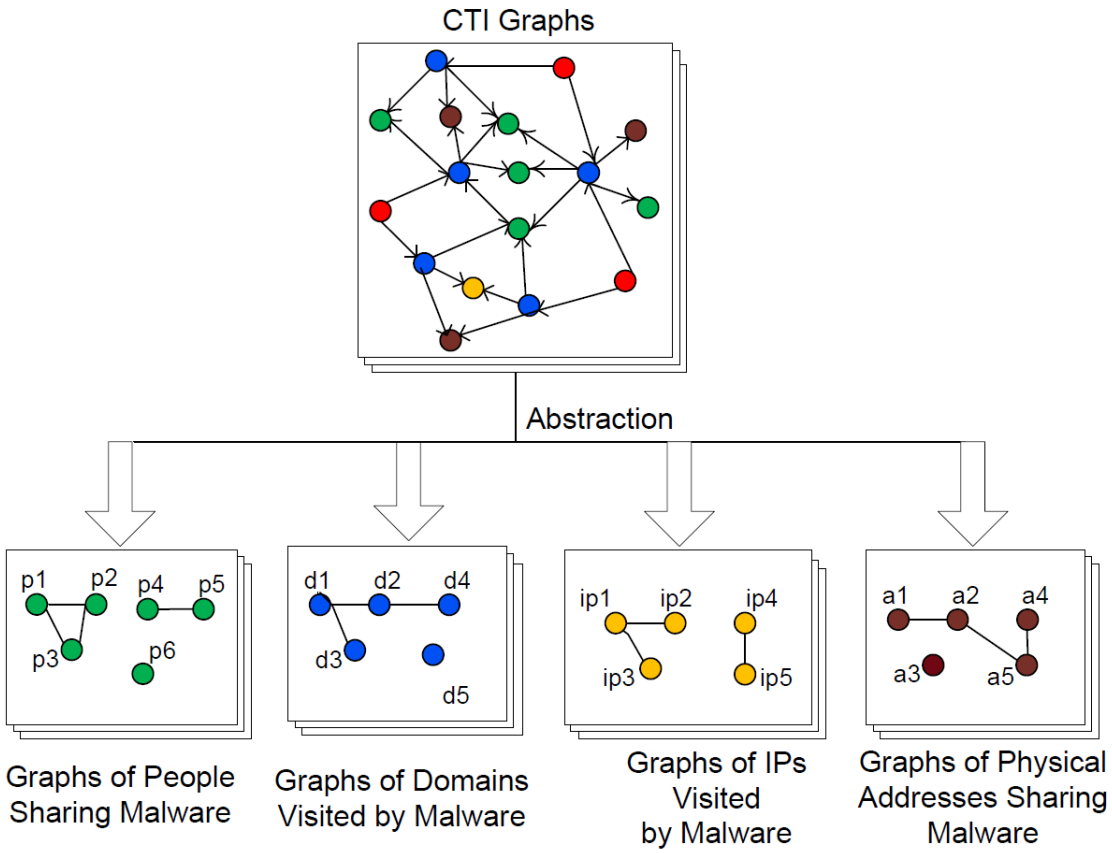
- What are the infrastructures used by cyber criminals to perpetrate attacks? Identify **connected components** in the CTI graph

5 days evolution



Cyber-threat Infrastructures

- Abstraction



Cyber-threat Infrastructures

- What are the most important players in cyber-threat infrastructures (e.g., owners, domains, IPs, organizations)?
  - Measuring the importance of different vertices through ranking
  - Centrality concepts: degree centrality, betweenness centrality, closeness centrality
  - Influence concepts: Eigenvector centrality, HITS and Google PageRank

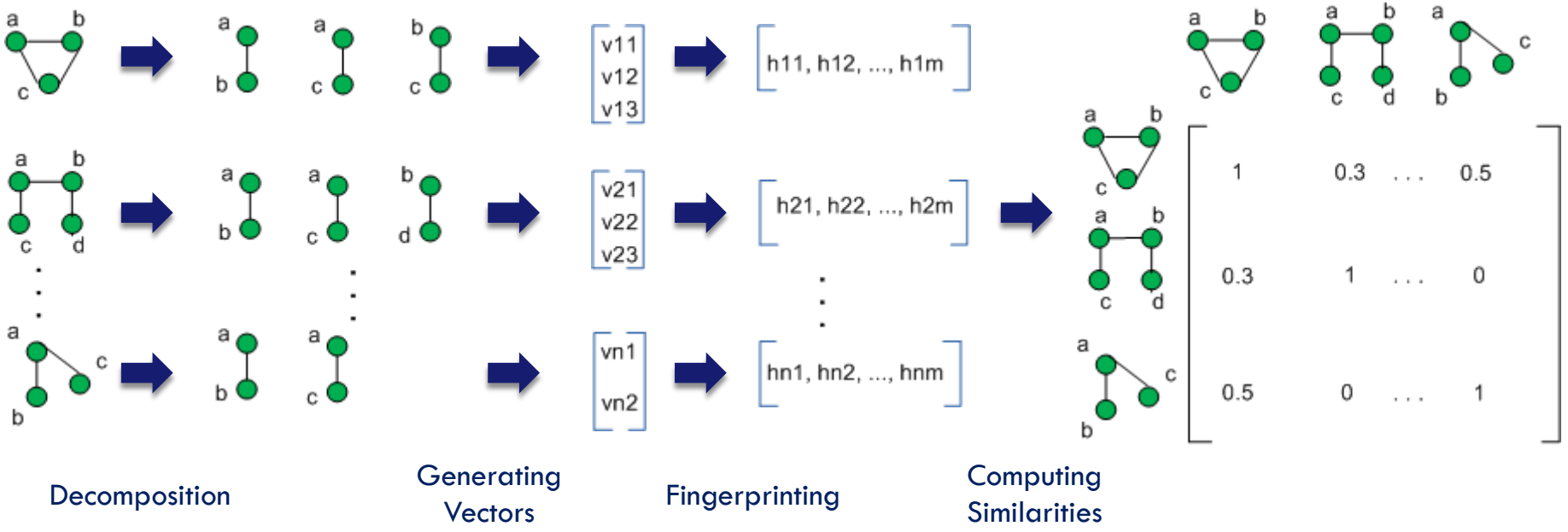
Cyber-threat Infrastructures

- Google PageRank algorithm is based on the notion of influence. A node in a graph is influent if it is connected to influent nodes.
- Page rank algorithm is used to compute the badness of IPs, domains, owners, physical addresses.
- An IP is connected to another IP if they share malware samples. Similarly, a domain is connected to a domain if they share malware samples.
- Why PageRank?
  - A stochastic approach implies randomness.
  - Web-surfer model illustrates the access of web pages through a stochastical model.
  - Analogy: the probabilistic approach reflects potential events done in cyber-threat infrastructures.

Cyber-threat Infrastructures

- How cyber-threat infrastructures evolve?
- To identify patterns of cyber-criminal activities, e.g., persistent, periodic, sporadic.
- Approach:
  - Computing the similarities between CTI graphs
  - Using the concept of **graph kernel** (mapping graph data from a non-linear to a linear space)
  - Generating graph fingerprints (signatures) using the min-hashing technique

Cyber-threat Infrastructures



Statistics and Insights

- Data Description
  - Period: 25th Aug 2013 - 25th Aug 2014
  - Number of malware: 4,717,628
  - Number of domains: 9,303,378
  - Number of second level domains: 151,757
  - Number of resolving IPs: 240,174
  - Number of connected IPs: 118,270
  - Number of domains Whois records: 110,414
  - Number of IP Whois records: 287,005

Statistics and Insights

2nd Level Domain	# Malware
*il.ru	252,358
*entre.ru	194,749
*soft.com	190,327
*update.com	166,995
*admr.com	160,123
cloud*.net	137,883
*lytics.com	119,233
*box.net	113,619
*host2.com	110,373
*tal.com	106,817

Presence of legitimate domains

Statistics and Insights

Domain	# IPs
j.nb*.com	23,021
Ip*.33*.org	10,779
f.nb*.com	10,313
i.nb*.com	7,130
g.nb*.com	5,825
*sopuli.*to.org	4,300
h.nb*.com	4,232
e.nb*.com	3,963
router.bi*.com	3,573
*lytics.com	3,342

Statistics and Insights

Resolving IPs (World)	# Domains
184.1xx.xxx.x6	171,388
199.xxx.xx.xx0	125,454
184.x7x.xxx.xx5	90,766
184.x7x.xxx.xx0	84,296
184.x7x.xxx.xx8	82,104
216.2xx.xxx.x5	21,402
216.2xx.xxx.x1	20,521
46.xx.xxx.x0	13,606
162.xxx.x.xx4	7,410

Statistics and Insights

- 184.1xx.xxx.x6
  - Canada
  - Observations:
    - Linked to dynamic generated domains
    - Pattern: [a-z|0-9]$^{+30}$. *eker.com
    - e.g. zzzhmn2ux101asbawtt8208204wzy4. *eker.com
    - Malware families: antiav, barys, esfury, graftor, injector, ramnit, sality, slugin, swisyn, symmi, vbinject, virut, zusy.
    - No registrant
    - Name Servers: ns1.b*.com, ns2.b*.com

Statistics and Insights

Many resolving IPs belong to ORG1(CHINA)

Network Name (World)	# Resolving IPs
ORG1(CHINA)-GD	12,880
ORG1(CHINA)-JS	7,245
ORG2(CHINA)-SD	5,113
ORG2(CHINA)-HA	3,745
ORG2(CHINA)-HE	3,597
ORG1(USA)--2011L	3,255
ORG1(CHINA)-SC	3,232
ORG1(CHINA)-FJ	3,001
ORG1(CHINA)-HB	2,844
ORG2(CHINA)-LN	2,815

Statistics and Insights

Connected IPs (World)	# Malware
93.xxx.xx.xx0	11,553
65.xx.xx.xx7	4,497
219.xxx.x.xx7	4,097
113.xx.xxx.xx6	3,223
95.xxx.xx.xx3	2,719
124.xxx.xxx.6	2,609
147.xxx.xxx.x7	2,498
69.xxx.xx.x0	2,429
89.xxx.xx.xx4	2,253
125.xx.xxx.x4	2,139

Statistics and Insights

- 93.xxx.xx.xx0
  - Owner: cloud hosting infrastructure (VPN anonymity service)
  - Country: Germany
  - Observations:
    - Malware families: antiav, barys, esfury, hype, injector, navipromo, pirminay, ramnit, slugin, swisyn, symmi, vbinject, virut, vundo, zbot
    - Mainly botnet trojans

Statistics and Insights

ORG1 (Malaysia) has the highest number of connected IPs, associated mainly with zlob, zbot and Proxy Trojans

Network Name (World)	# Connected IPs
ORG1(MALAYSIA)-HSDPA	10,093
ORG1(INDIA)-SouthZone	2,176
ORG1(CHINA)-GD	1,089
ORG1(KOREA)	837
ORG1(USA)-2011L	739
ORG3(CHINA)	681
ORG1(CHINA)-JS	604
ORG2(KOREA)	472
ORG4(CHINA)	358
ORG2(USA)	309

Statistics and Insights

Privacy service is widely used

One regular name

Conjar, fareit, nebular, zbot, zusy

Domains parking

Registrant (World)	# Domains
Registration Private	2,938
Whoisguard protected	744
Domain administrator	632
Domain admin	451
Whois agent	378
Perfect Privacy LLC	274
E*I Y*	187
Whois privacy protection service	184
Private Registrant	163
Oneandone Private Registration	123
Spy Eye	120
This domain for sale toll free: *-822-*	104
DNS admin	92
reactivation period	92
Domain manager	75

Statistics and Insights

708 second-level domains (Privacy Service) linked to spamming, robot calls and scam abuses

Address (World)	# Domains
P.O. box ****-**** Panama PA	708
***** Northsight blvd ****9 Scottsdale	379
***** Gran bay parkway USA	272
***** P.O Box ** Beach AUSTRALIA	228
***** Memorial Dr. office #935 USA	186
Ilyinka street ** RUSSIA	120
***** 24th street USA	115
*** Lee road suite **0 USA	108
*** Main street #*** USA	108
- Boulevard Massena FRANCE	97

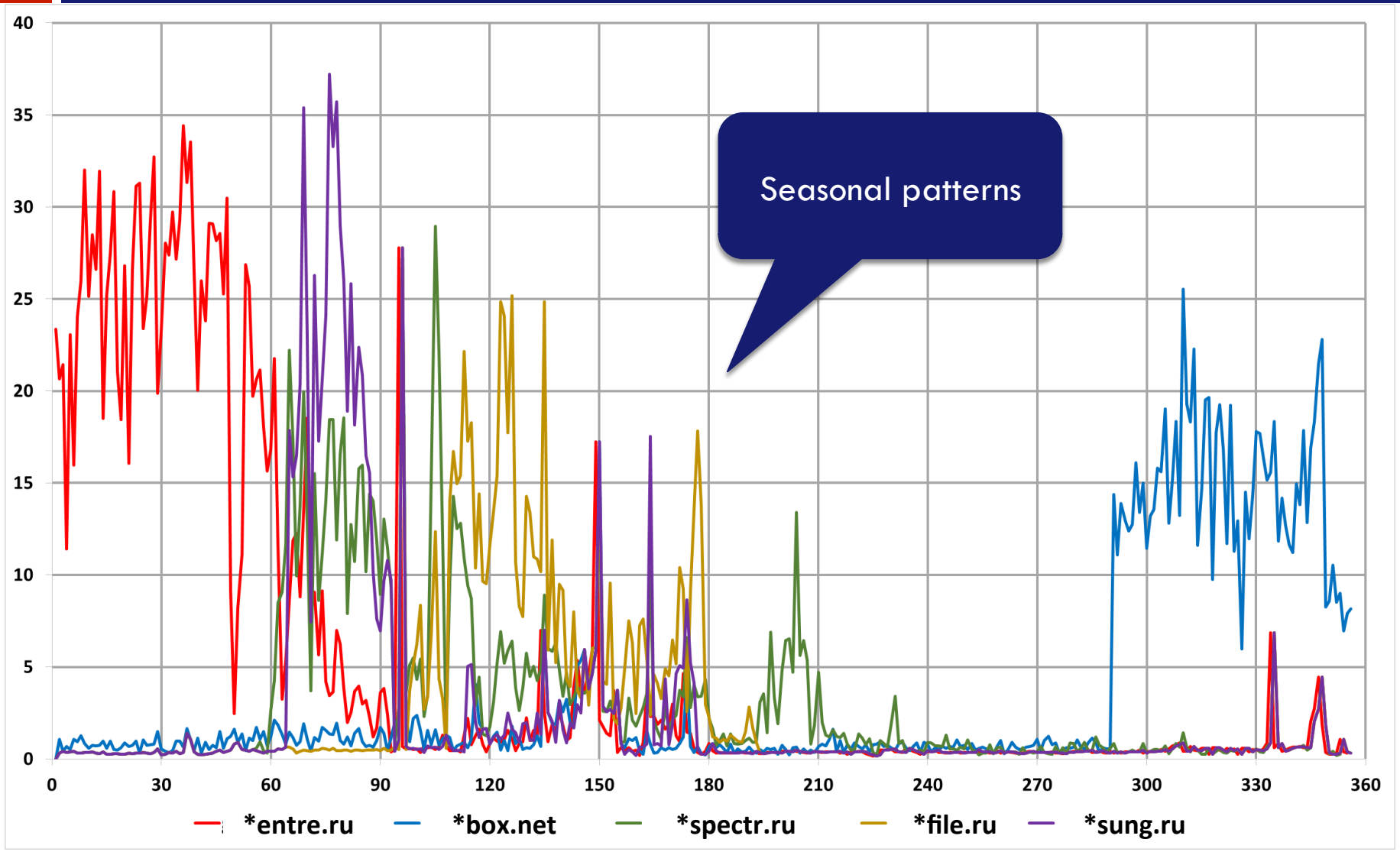
Statistics and Insights

Domains	Average Score	Owner	Malware Families
*entre.ru	5.1194617836	Private Person	bitcoinminer,generickdz,graftor, loadmoney, minggy, startpage, strictor, symmi, zusy
*box.net	3.5080075652	N. ***n	
*spectr.ru	2.8936699826	Private Person	loadmoney, kryptik, badur
*file.ru	2.3333310540	N/A	
*sung.ru	2.1706995605	Private Person	graftor, hype, loadmoney, strictor, symmi, zusy
*express.ru	2.0137619806	Private Person	graftor, loadmoney strictor, zusy
*ldr.ru	1.6519902951	N/A	
*.elb.*aws.com	1.4375291335	*on.com	
d1sx0cjuasqkw9.*ront.net	1.4113904548	*zon.com	
*pro.ru	1.4048343632	Private Person	graftor, strictor, symmi, zusy, badur

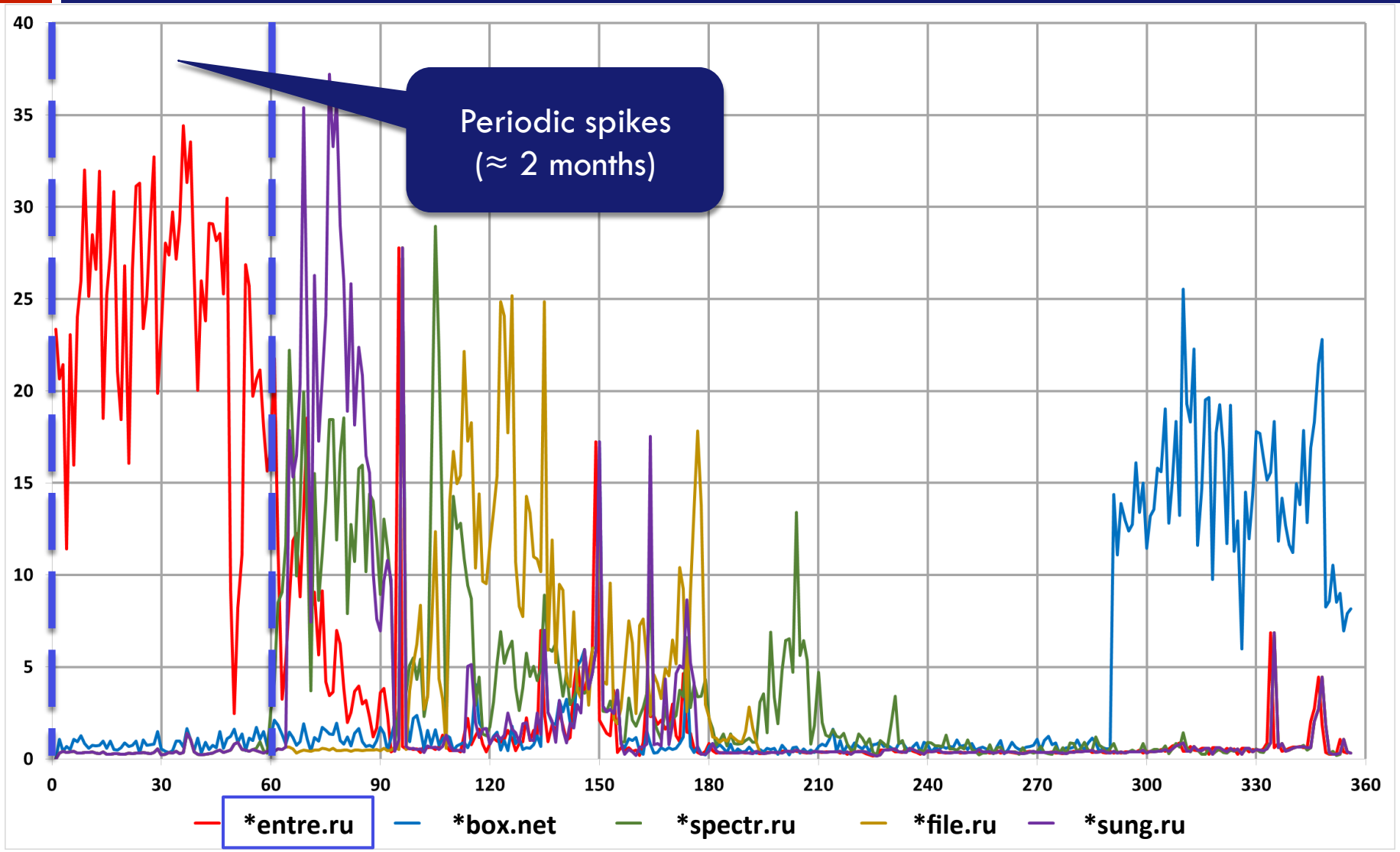
Statistics and Insights

- 5 out of top-10 domains have ".ru" extension and belong to the same registrant "Private Person"
- These domains are linked almost to the same malware families
- No physical address is associated with these domains
- Sharing same name-servers

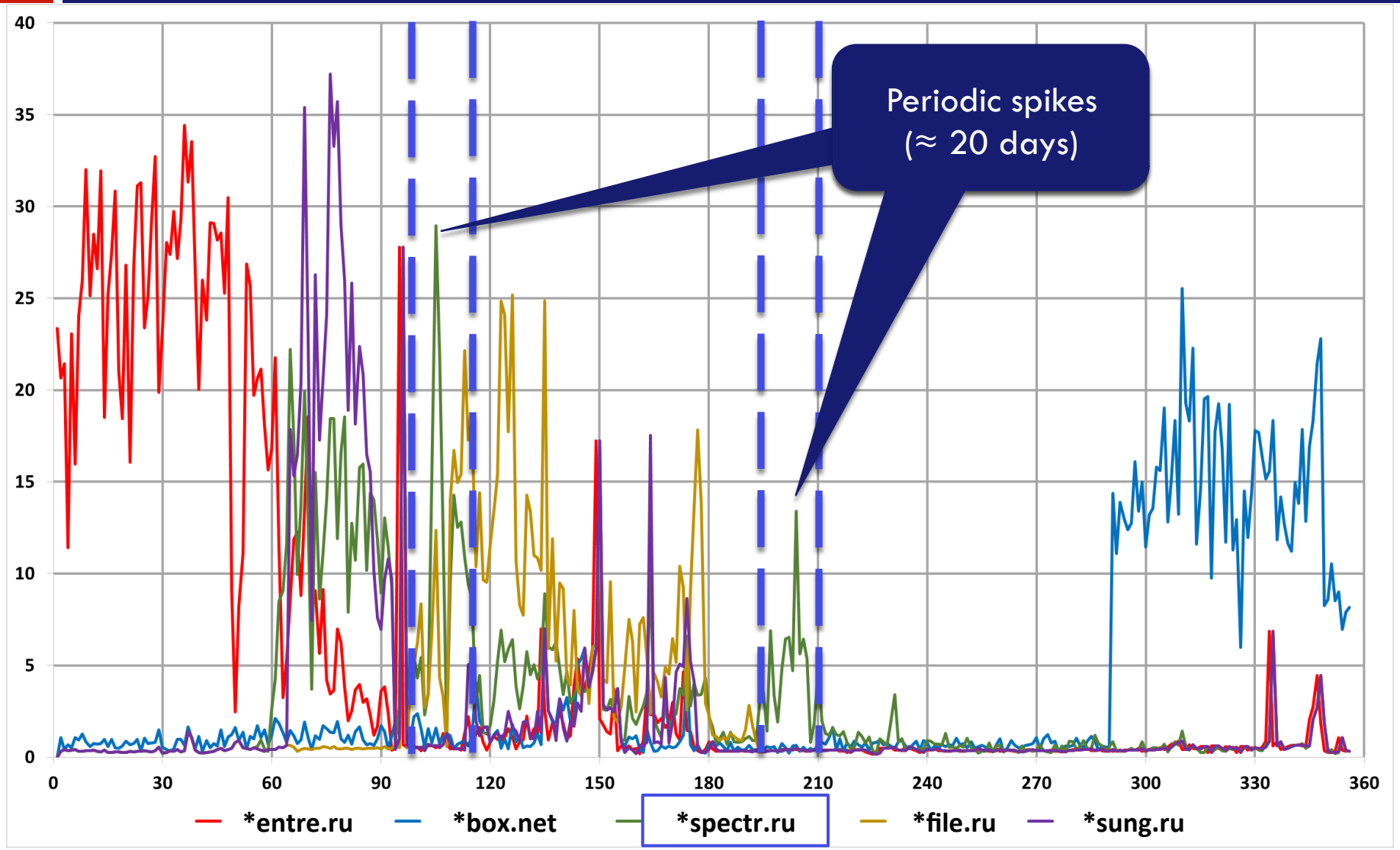
Statistics and Insights



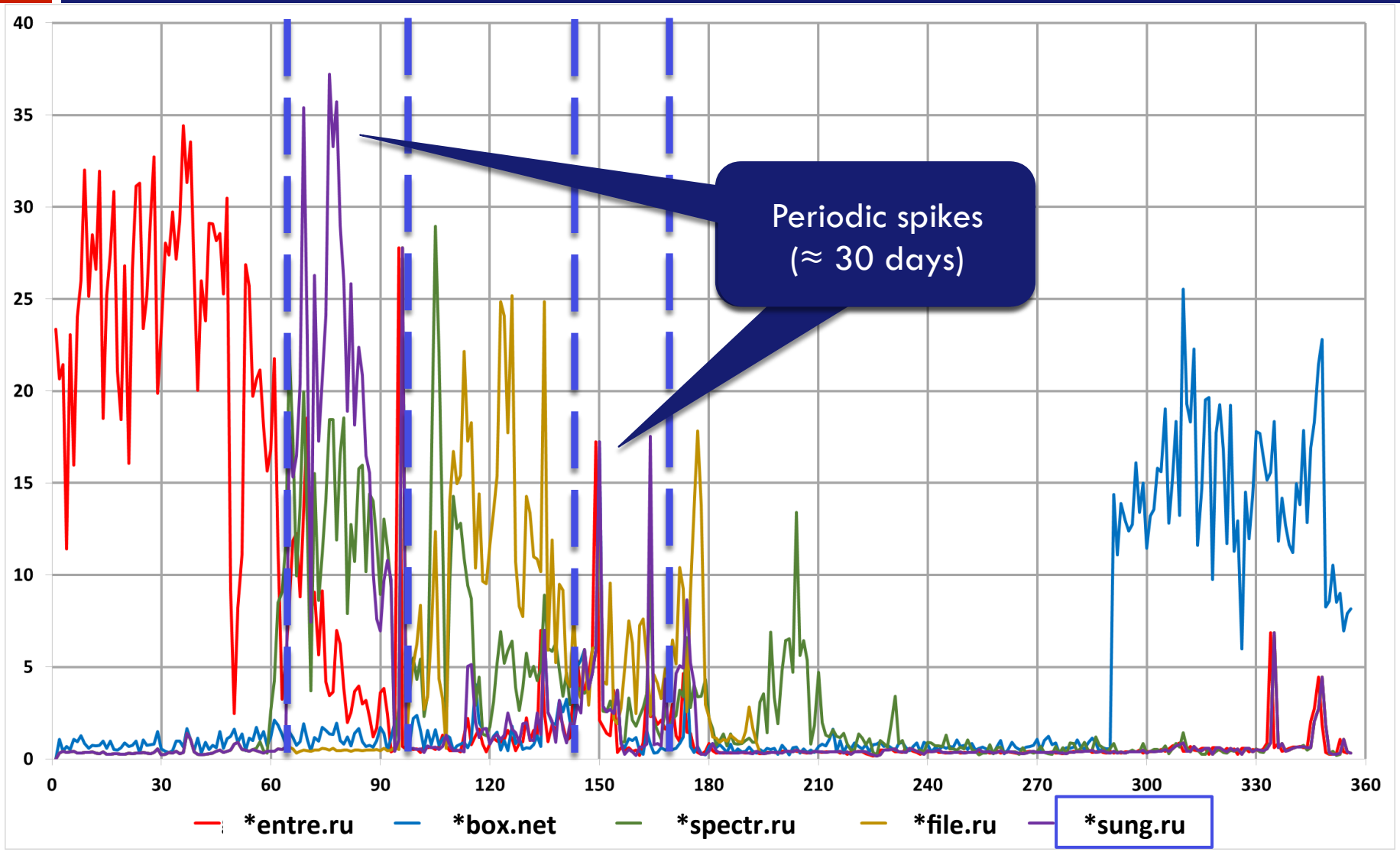
Statistics and Insights



Statistics and Insights



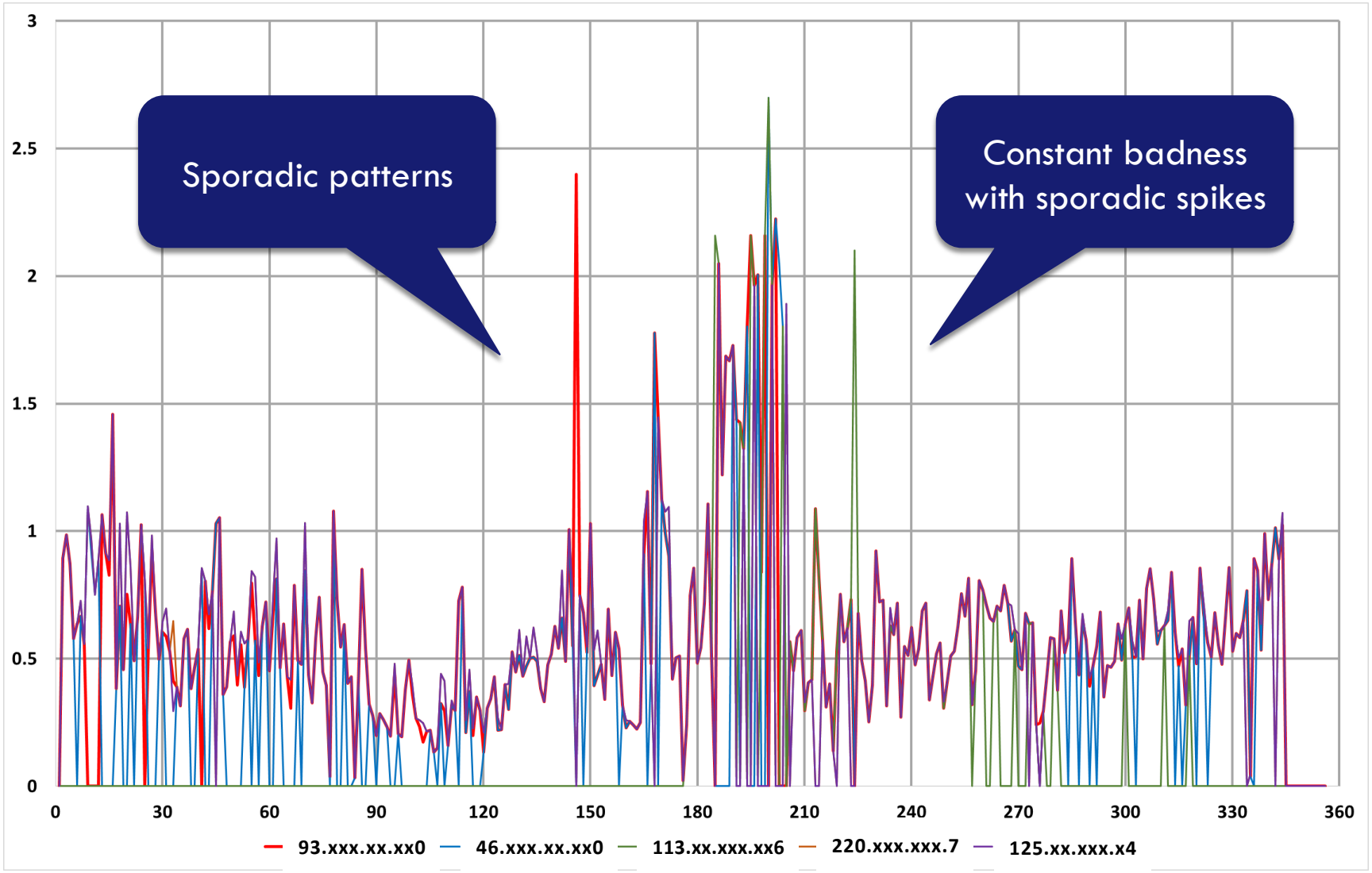
Statistics and Insights



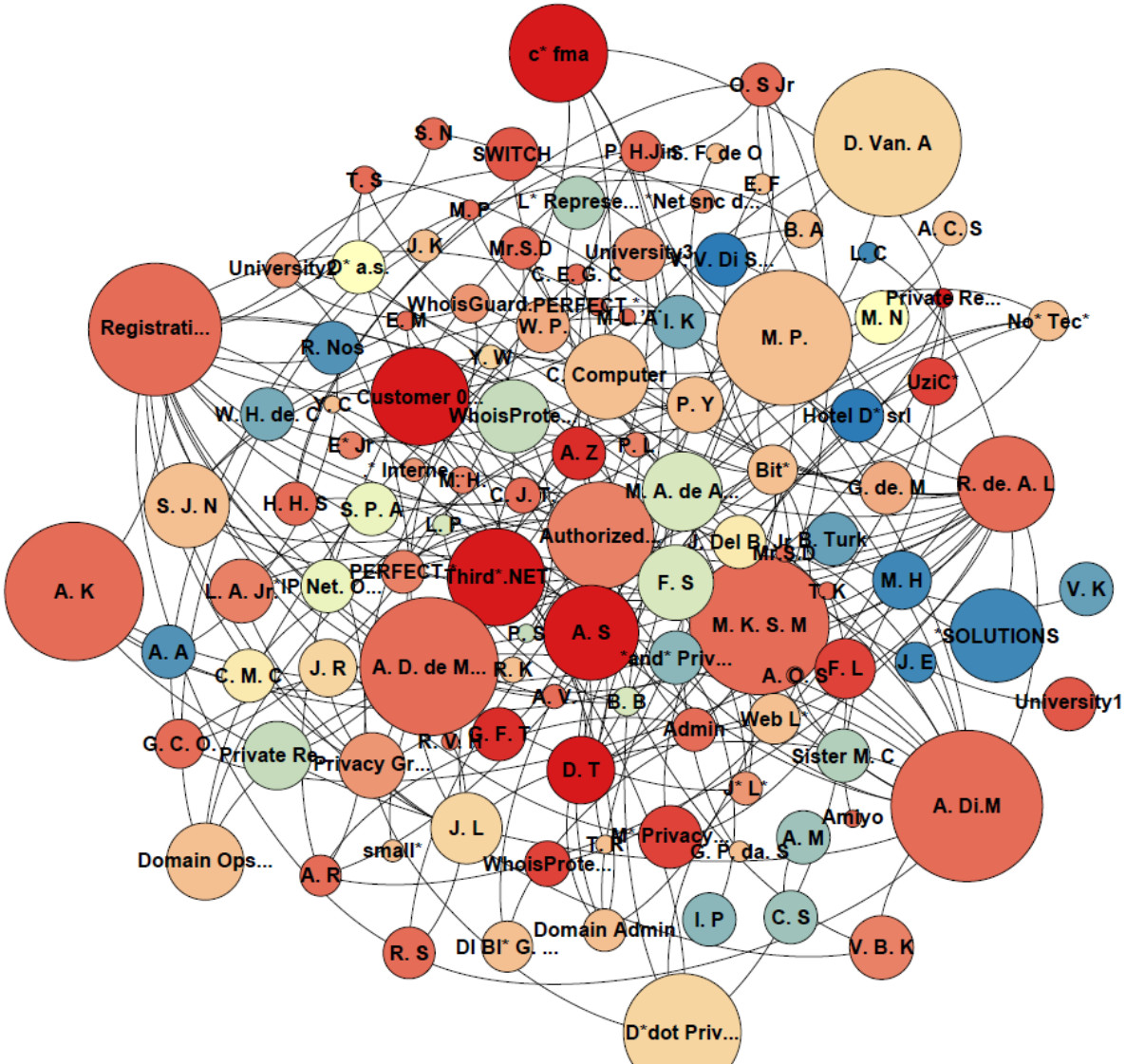
Statistics and Insights

IPs (World)	Average Score	Country
93.xxx.xx.xx0	6.90947359832	GERMANY
46.xxx.xxx.xx9	5.76183602875	FRANCE
113.xx.xxx.xx6	5.1836928519	CHINA
220.xxx.xxx.7	4.53174275775	CHINA
125.xx.xxx.x4	4.52513888983	CHINA
124.xxx.xxx.x1	4.51910871828	CHINA
221.xxx.xxx.x8	4.31768525507	CHINA
91.xxx.xx.x0	3.06311160603	UK
239.255.255.250	2.88498391036	IP multicast SSDP
89.xxx.xx.xx4	2.5103999827	ITALIA

Statistics and Insights

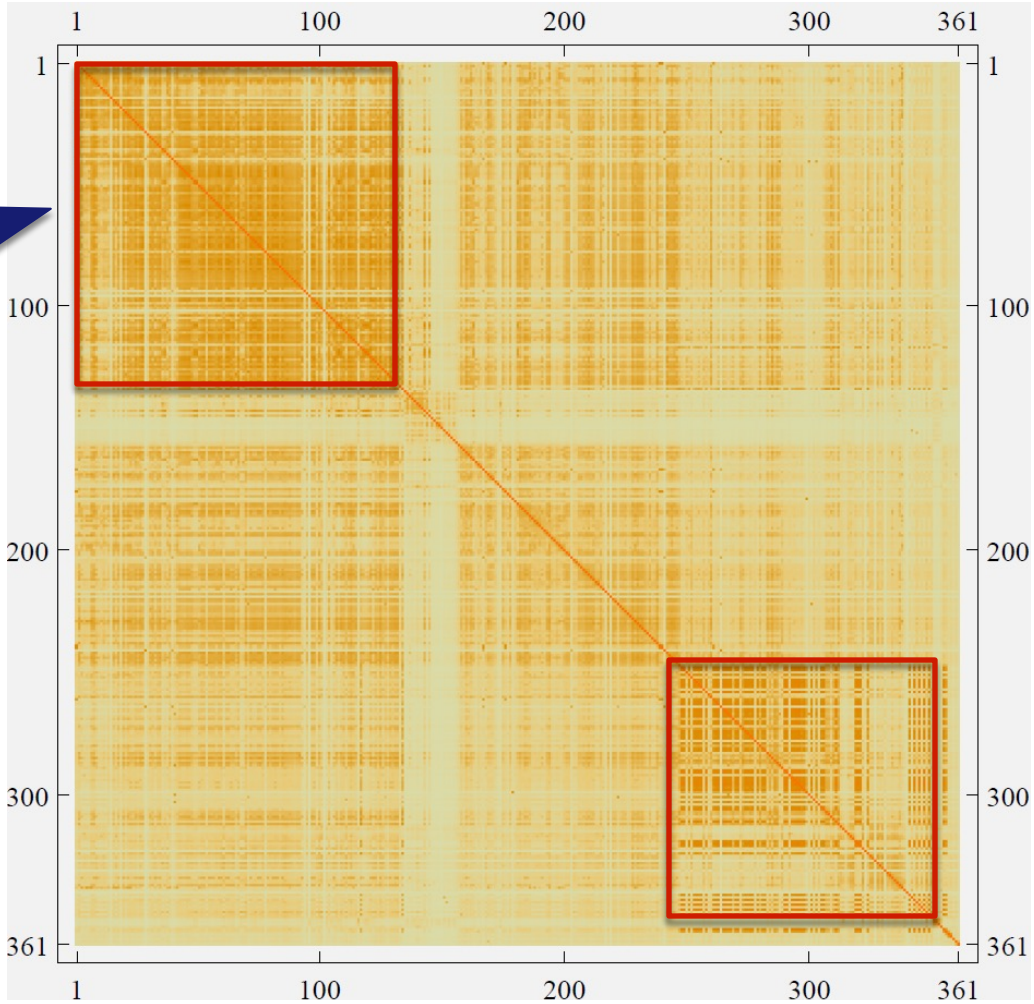


Statistics and Insights



Statistics and Insights

Sharing domains in first and last trimesters



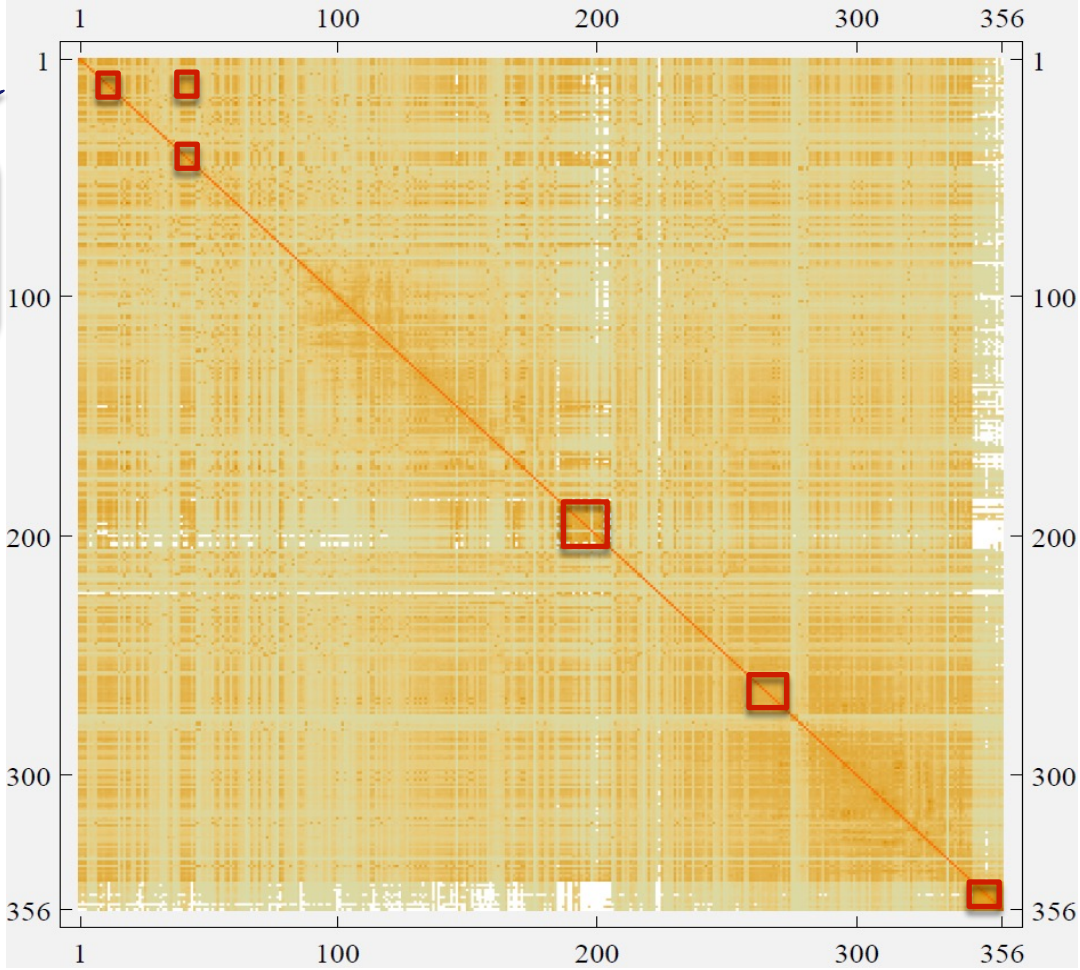
Domain Patterns Similarity Matrix

Statistics and Insights

Domain Patterns	# Days	# Malware	Owners
f*[dd]75.com;a*[dd]75.com	332	6,045	Registration Private, Domains By Proxy, LLC
f*[dd]75.com;w*[dd]88.com	329	6,046	Registration Private, Domains By Proxy, LLC
air*lrs*.com;air*.com	329	15,914	WHOISGUARD PROTECTED, INC, Panama
*uwey.net;*uwey.org	328	25,307	Undefined; Ge* Em*
*uwey.com;*uwey.org	327	25,208	Undefined; Ge* Em*
*all.org;*1532.com	323	44,182	Undefined
*1352.org;*1352.net	322	25,788	Undefined
*all.org;*all.net	318	29,398	Undefined
*od.com; *ed.net	318	6,787	*, Inc; Pre* La*
wk1888.com;af0575.com	317	5,966	Registration Private, Domains By Proxy, LLC

Statistics and Insights

Less sharing in terms of connected IPs



Short period sharing between connected IPs

IP Patterns Similarity Matrix

Statistics and Insights

IP Patterns	# Days	# Malware	Organizations
220.xxx.xxx.7;125.xx.xxx.x4	317	2,123	ORG1 Beijing; ORG4 Beijing Zhi*
220.xxx.xxx.7;124.xxx.xxx.x1	311	2,068	ORG1 Beijing; ORG1 Heibei
221.xxx.xxx.x8;125.xx.xxx.x4	289	1,938	ORG5 Company Lan*; ORG4 Beijing Zhi*
220.xxx.xxx.7;221.xxx.xxx.x8	289	1,948	ORG1 Beijing; ORG5 Company Lan*
124.xxx.xxx.x1;125.xx.xxx.x4	278	1,925	ORG1 Heibei; ORG4 Beijing Zhi*
124.xxx.xxx.x1;221.xxx.xxx.x8	123	1,836	ORG1 Heibei; ORG5 Company Lan*
91.xxx.xx.x0;239.255.255.250	97	511	ORG1(UK); SSDP multicast IP
65.xx.xx.xxx;95.xxx.xx.xx3	84	1,232	ORG3 (USA); ORG1 (Italy)
77.xx.xxx.x0;62.xx.xxx.xx2	84	204	ORG1 (Russia); ORG1(Estonia)
91.xxx.xx.x0;66.xxx.xx.xx9	79	158	ORG1(UK); ORG4 (USA)

Cluster of IP patterns indicate a collaborative malware activity

Conclusion

- We elaborate a study on one year malware data
- This study puts forward the ability to investigate cyber-threat infrastructures
- PageRank algorithm to compute maliciousness score for different entities in cyber-threat infrastructures
- Social network analysis on owners for different malicious domains to discover potential cyber-criminal groups
- Inferring patterns through min-hashing technique

Future Works

- Automation of the generation of insights and intelligence on a daily basis
- Corroborating this study by applying frequent item sets algorithm (e.g., FP-growth)
- Empirical study to find the investigation period to consider for domains badness persistence and IPs badness sporadicalness

Thanks !!

Cyber-threat Infrastructures

- PageRank Algorithm:
  - Let V be set of all vertices, Let $I(v_i)$ be the set of vertices that link to a vertex v_i and let $\deg_{out}(v_i)$ be the out-degree centrality of a vertex v_i.
  - Out-degree: Let $G=(V;E)$ be a directed graph. The out-degree of a vertex v is the total number of edges in E with source v.
  - PageRank initial formula:

$$PR(v_i)=d\left[\sum_{v_j \in I(v_i)} PR(v_j)/\deg_{out}(v_i)\right]+(1-d)1/|V| \quad (1)$$

Cyber-threat Infrastructures

- PageRank Algorithm:
  - d: damping factor (0.85), it represents random probability that a web surfer is visiting a webpage.
  - $1-d$: a web surfer stops visiting a webpage and chooses randomly another webpage.
  - $1/\deg\downarrow out\,(v\downarrow j)$: the probability that a surfer visits a webpage linked to webpage $v\downarrow i$.
  - $|V|$: cardinality of vertices set

Cyber-threat Infrastructures

- PageRank Algorithm:
  - Randomness of accessing webpages implies use of stochastic process to score webpages. Different PageRank values are through recursive computation of Equation (2).
  - W: stochastic matrix where:

$$w_{ij} = e_{ij} \times 1/\deg_{out}(v_j) \quad \text{if } v_i \text{ is linked to } v_j$$

$$w_{ij} = 0 \quad \text{if } v_i \text{ is not linked to } v_j$$

e_{ij} : edge (v_i, v_j) normalized weight value

$$PR = d[W.PR] + (1-d)1/|V|\,1 \qquad (2)$$

Cyber-threat Infrastructures

- PageRank Algorithm:
  - Stop criterion parameter: for each step t a new vector (PR,t) is computed. The algorithms stops when $|(PR,t)-(PR,t-1)|<\varepsilon$
- Why PageRank?
  - A stochastic approach implies randomness.
  - Web-surfer model illustrates the access of web pages through a stochastical model.
  - Analogy: the probabilistic approach reflects potential events done in cyber-threat infrastructures.

Statistics and Insights

- Why Malware visit legitimate domains?
  - Redirection of accesses of legitimate domains to malicious domains, e.g, windows update website redirection, or google-analytics fake web-pages.
  - Stealing mail credentials by exploiting vulnerabilities, e.g., adobe flash player vulnerability used by malware to distribute mail credentials.
  - Example: Kaba variant (plugX) connects to legitimate domains. It resolved DNS lookups through a hosting company. As a result, it was used to hijack legitimate domains since the hosting company did not check if hosted domains are registered elsewhere or not.