30th ANNUAL FIRST CONFERENCE
KUALA LUMPUR
June 24-29, 2018
30 YEARS OF INCIDENT HANDLING

Threat Hunting Techniques at Scale

Dhia Mahjoub, PhD
Head of Security Research, Cisco Umbrella (OpenDNS)

Tuesday, June 26th, 2018

FIRST™
Improving Security Together

Agenda

Intelligence cycle at scale

Big data challenges

Spike detection and classification

Co-occurrences

Tracking Malspam: combining techniques

SSL Data mining

Conclusion

Contributors

Dhia Mahjoub, Head of Security Research, @DhiaLite

PhD graph theory, network security, threat intel

Thomas Mathew, Senior Security Researcher

MS Computer Science, signal analysis, machine learning

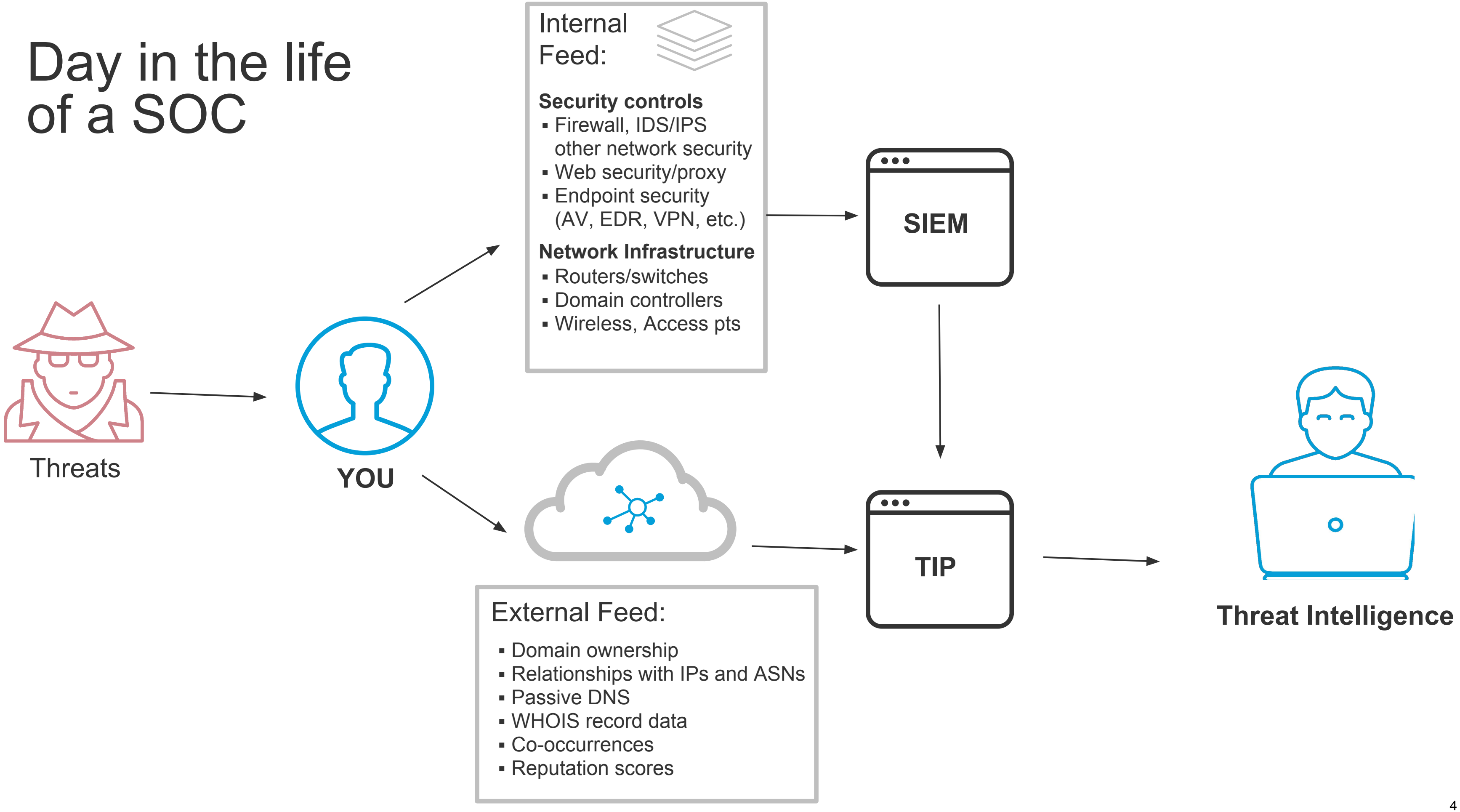
Scott Sitar, Technical Leader

MS Applied Math, big data engineering, algorithms

David Rodriguez, Senior Security Researcher

MA Mathematics, statistics, machine learning

Day in the life of a SOC



Threats

YOU

Internal Feed:

Security controls
- Firewall, IDS/IPS other network security
- Web security/proxy
- Endpoint security (AV, EDR, VPN, etc.)

Network Infrastructure
- Routers/switches
- Domain controllers
- Wireless, Access pts

External Feed:

- Domain ownership
- Relationships with IPs and ASNs
- Passive DNS
- WHOIS record data
- Co-occurrences
- Reputation scores

SIEM

TIP

Threat Intelligence

Umbrella Investigate Intel Production Cycle

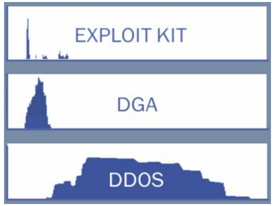
Requirements
Detect and block domains/IPs to protect customers and provide insight and context around domains/IPs

Collection
Retrieve raw DNS, IP, BGP, SSL, whois, hash, crawled web data, etc. at scale

Google
Map data ©2017 Terms of Use

Processing
Caching, indexing, enriching, summarizing data at scale

Analysis
Threat detection and reputation scores using ML & graph models, human domain expertise

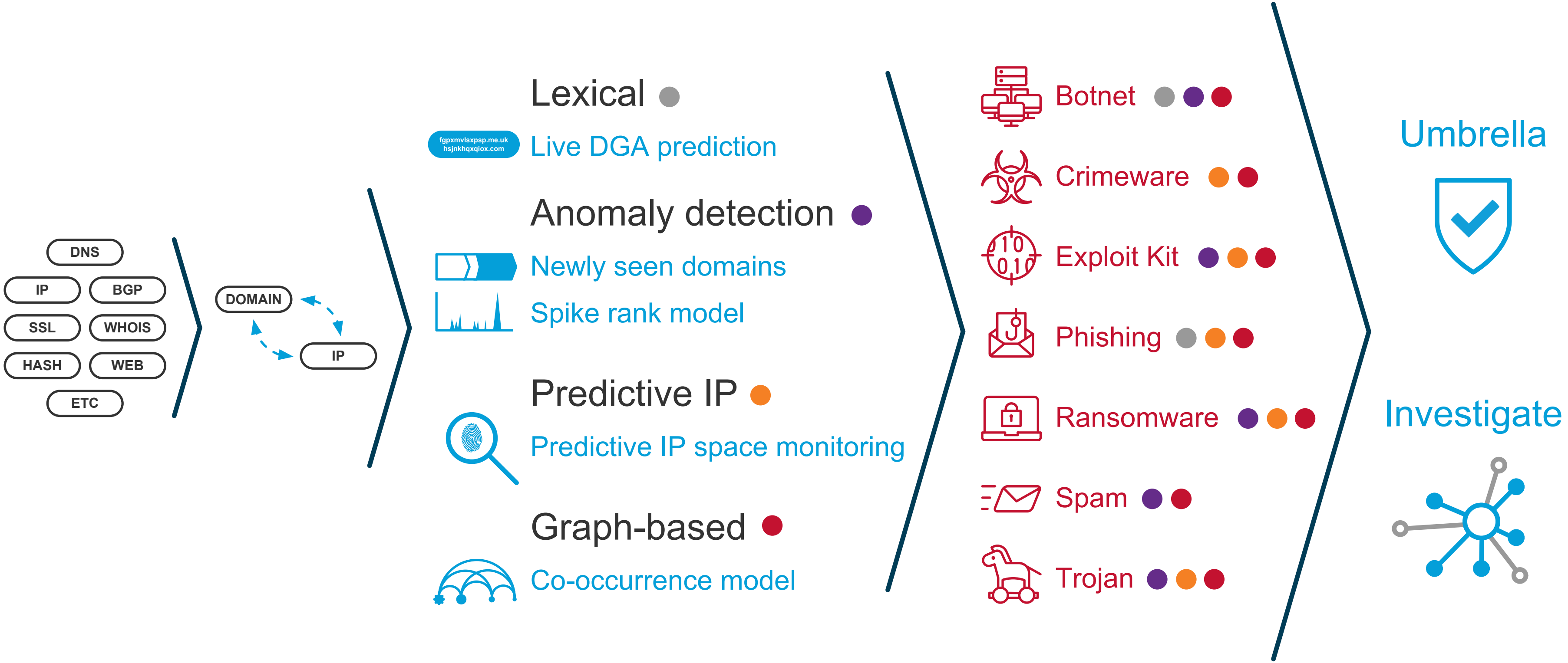


Dissemination
Domains, IPs into block list

Investigate UI and API

Feedback
Protected customers, actionable use cases with Investigate

What makes us different



DNS
IP
BGP
SSL
WHOIS
HASH
WEB
ETC

DOMAIN
IP

Lexical ●

Live DGA prediction

Anomaly detection ●

Newly seen domains

Spike rank model

Predictive IP ●

Predictive IP space monitoring

Graph-based ●

Co-occurrence model

Botnet ● ● ●

Crimeware ● ●

Exploit Kit ● ● ●

Phishing ● ● ●

Ransomware ● ● ●

Spam ● ●

Trojan ● ● ●

Umbrella

Investigate

Malspam/Hancitor

Cybercrime sites

Criminal hosting space

Path of a malspam attack

1. **Phishing email sent from delta@performanceair.com**

2. **Victims click on malicious URLs**

   myhearthstonehomes.org
   ourrealtyguy.info
   ourrealtyguy.org
   ourrealtyguy.us
   package2china.com

3. **Malicious word doc drops Hancitor**

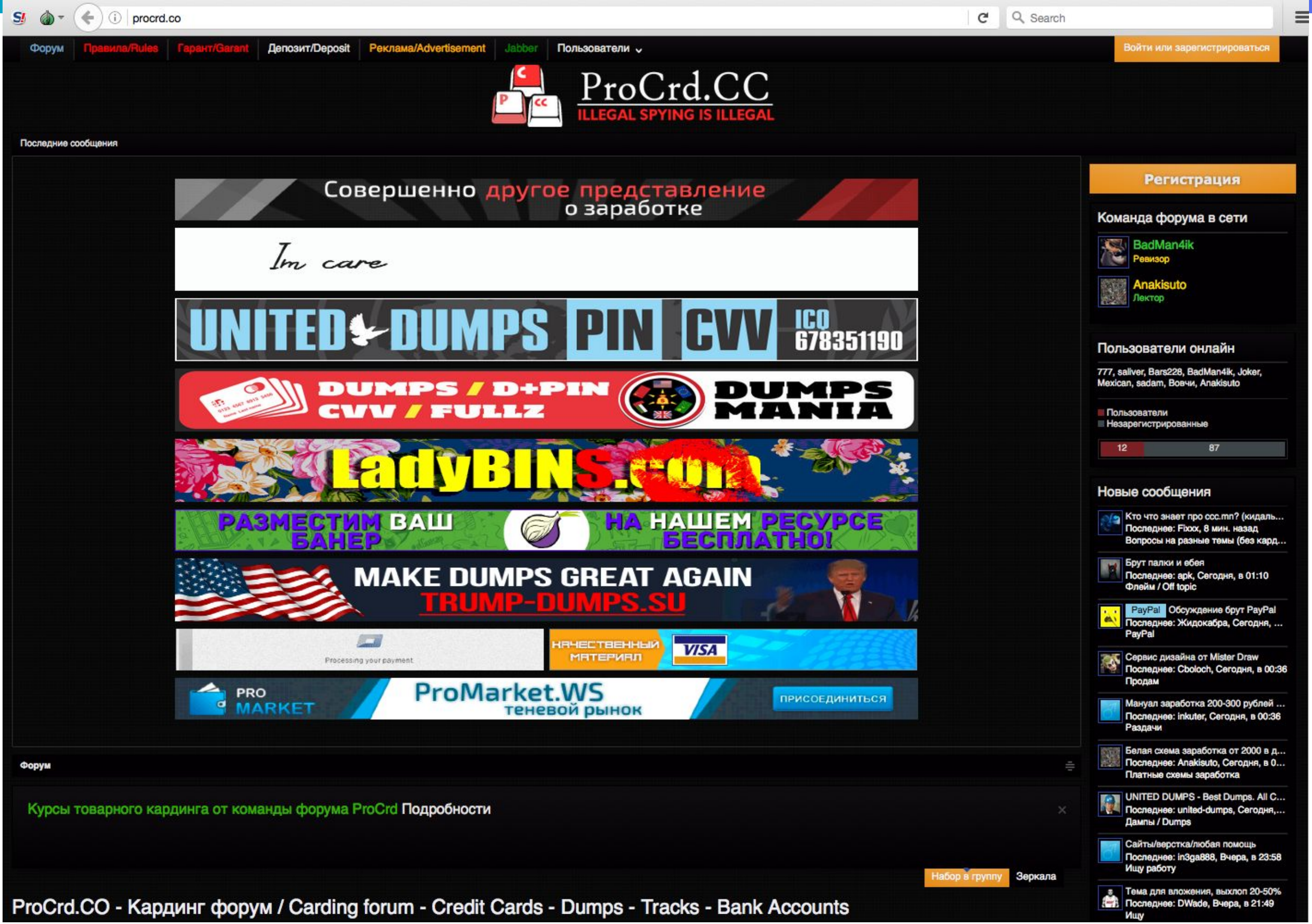
4. **Hancitor makes C2 call to domains for trojans**

   uneventrendi.com
   ketofonerof.ru
   thettertrefbab.ru

5. **Trojans (Pony, Evil Pony, Zloader) make C2 call for extra malware or functionality**

   mebelucci.com.ua
   uneventrendi.com
   lycasofrep.com
   rinbetarrab.com

6. **Infection on device & positioned for data extraction**


procrd.co
Search
Форум
Правила/Rules
Гарант/Garant
Депозит/Deposit
Реклама/Advertisement
Jabber
Пользователи
Войти или зарегистрироваться
ProCrd.CC
ILLEGAL SPYING IS ILLEGAL
Совершенно другое представление о заработке
I'm care
UNITED DUMPS PIN CVV ICQ 678351190
DUMPS / D+PIN CVV / FULLZ DUMPS MANIA
LadyBINS.com
РАЗМЕСТИМ ВАШ БАНЕР НА НАШЕМ РЕСУРСЕ БЕСПЛАТНО!
MAKE DUMPS GREAT AGAIN
TRUMP-DUMPS.SU
Processing your payment
КАЧЕСТВЕННЫЙ МАТЕРИАЛ VISA
PRO MARKET
ProMarket.WS теневой рынок
ПРИСОЕДИНИТЬСЯ
Регистрация
Набор в группу
Зеркала

Последние сообщения

Форум

Курсы товарного кардинга от команды форума ProCrd Подробности

ProCrd.CO - Кардинг форум / Carding forum - Credit Cards - Dumps - Tracks - Bank Accounts

Команда форума в сети

BadMan4ik
Ревизор

Anakisuto
Лектор

Пользователи онлайн

777, saliver, Bars228, BadMan4ik, Joker, Mexican, sadam, Вовчи, Anakisuto

🟥 Пользователи
⬛ Незарегистрированные

12 87

Новые сообщения

Кто что знает про ccc.mn? (кидаль...
Последнее: Fixxx, 8 мин. назад
Вопросы на разные темы (без кард...

Брут палки и ебея
Последнее: apk, Сегодня, в 01:10
Флейм / Off topic

PayPal Обсуждение брут PayPal
Последнее: Жидокабра, Сегодня, ...
PayPal

Сервис дизайна от Mister Draw
Последнее: Cboloch, Сегодня, в 00:36
Продам

Мануал заработка 200-300 рублей ...
Последнее: inkuter, Сегодня, в 00:36
Раздачи

Белая схема заработка от 2000 в д...
Последнее: Anakisuto, Сегодня, в 0...
Платные схемы заработка

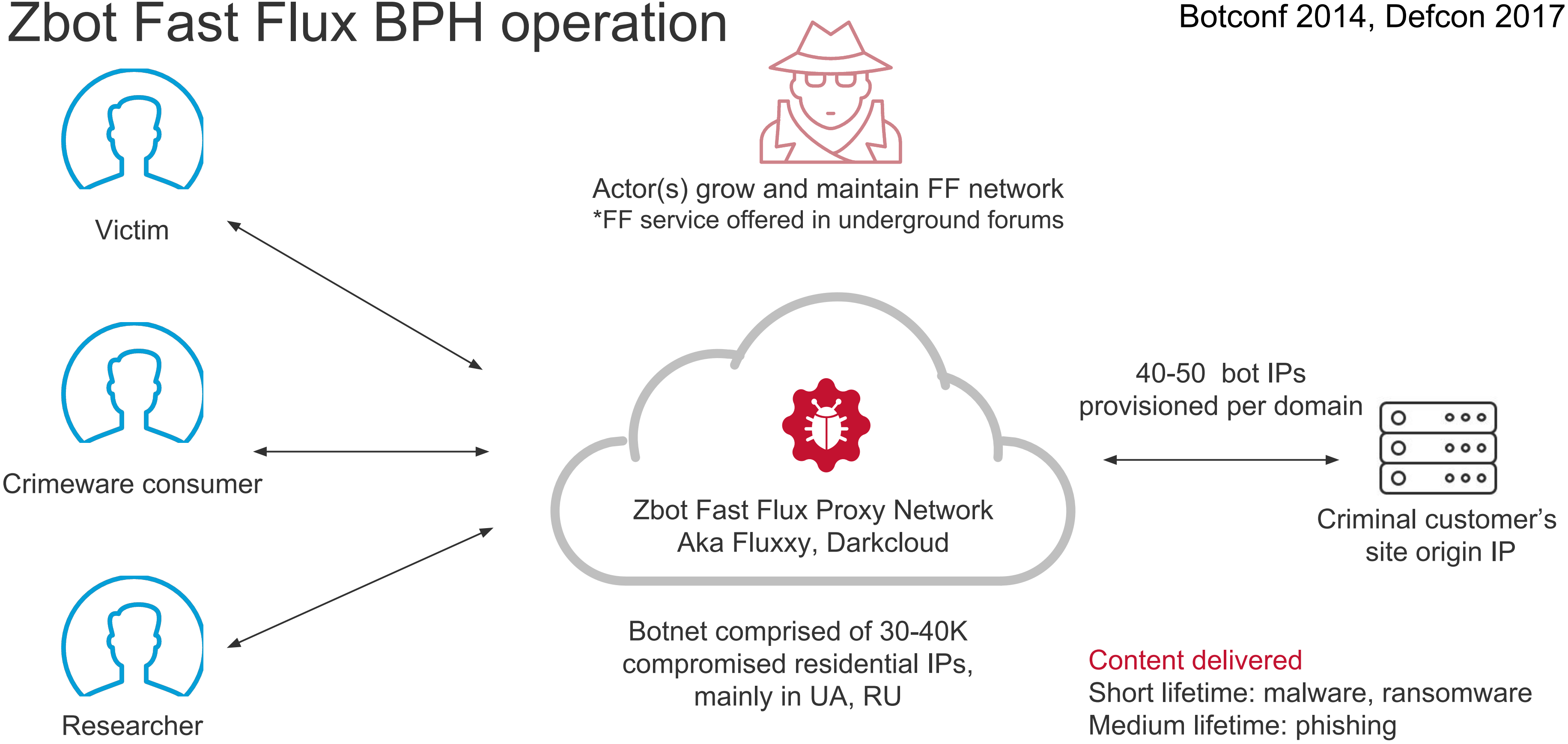
UNITED DUMPS - Best Dumps. All C...
Последнее: united-dumps, Сегодня,...
Дампы / Dumps

Сайты/верстка/любая помощь
Последнее: in3ga888, Вчера, в 23:58
Ищу работу

Тема для вложения, выхлоп 20-50%
Последнее: DWade, Вчера, в 21:49
Ищу

Zbot Fast Flux BPH operation

Introduced at Black Hat 2014, Botconf 2014, Defcon 2017



Actor(s) grow and maintain FF network
*FF service offered in underground forums

Victim

Crimeware consumer

Researcher

Zbot Fast Flux Proxy Network
Aka Fluxxy, Darkcloud

40-50 bot IPs
provisioned per domain

Criminal customer’s
site origin IP

Botnet comprised of 30-40K
compromised residential IPs,
mainly in UA, RU

Content delivered
Short lifetime: malware, ransomware
Medium lifetime: phishing
Long lifetime: carding, cybercrime forums

Data Collection

Working with Big Data

Challenges and Uses

- Large datasets give researchers possibility to uncover widespread network threats
- When working with large data sets - traditional threat-hunting methods have to be modified
- We will explore:
  - Pivoting
  - Classification

Working with Big Data

Pivoting

- Pivoting is useful for analysts when given seeds of information

- Want to convert some seed information into further information about a malicious campaign
  - Difficult to connect multiple data sets together
  - Scale of data can make look-ups difficult

Working with Big Data

Classification

- Classification is the most self-explanatory challenge
- Revolves around sifting through a dataset and classifying threats
- Challenges arise when dealing with scale and class-imbalance problem
    - For example, at Cisco Umbrella we produce around 4 TB of hourly data that needs to be processed in near real-time
    - Class-imbalance refers to the percentage of benign to malicious domains that can be found

Case-Studies

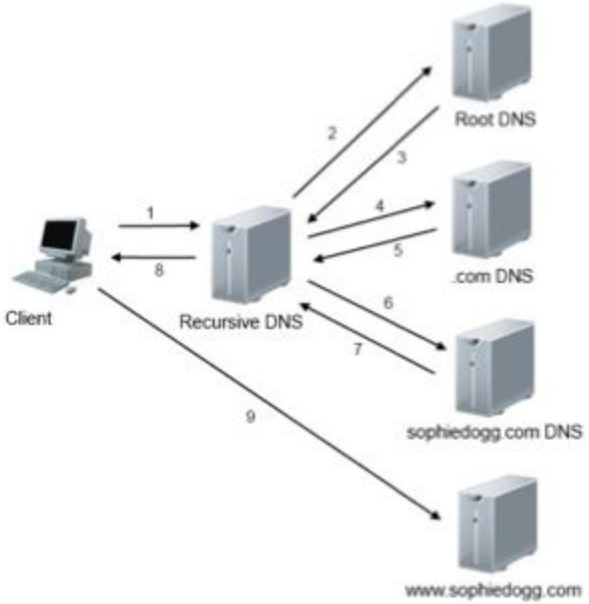
- We will be examining some successful use-cases we worked on on dayjob involving classifier design, and designing a platform for pivoting

- Two datasets:
  - Recursive Layer DNS Traffic
  - IPv4 SSL scans

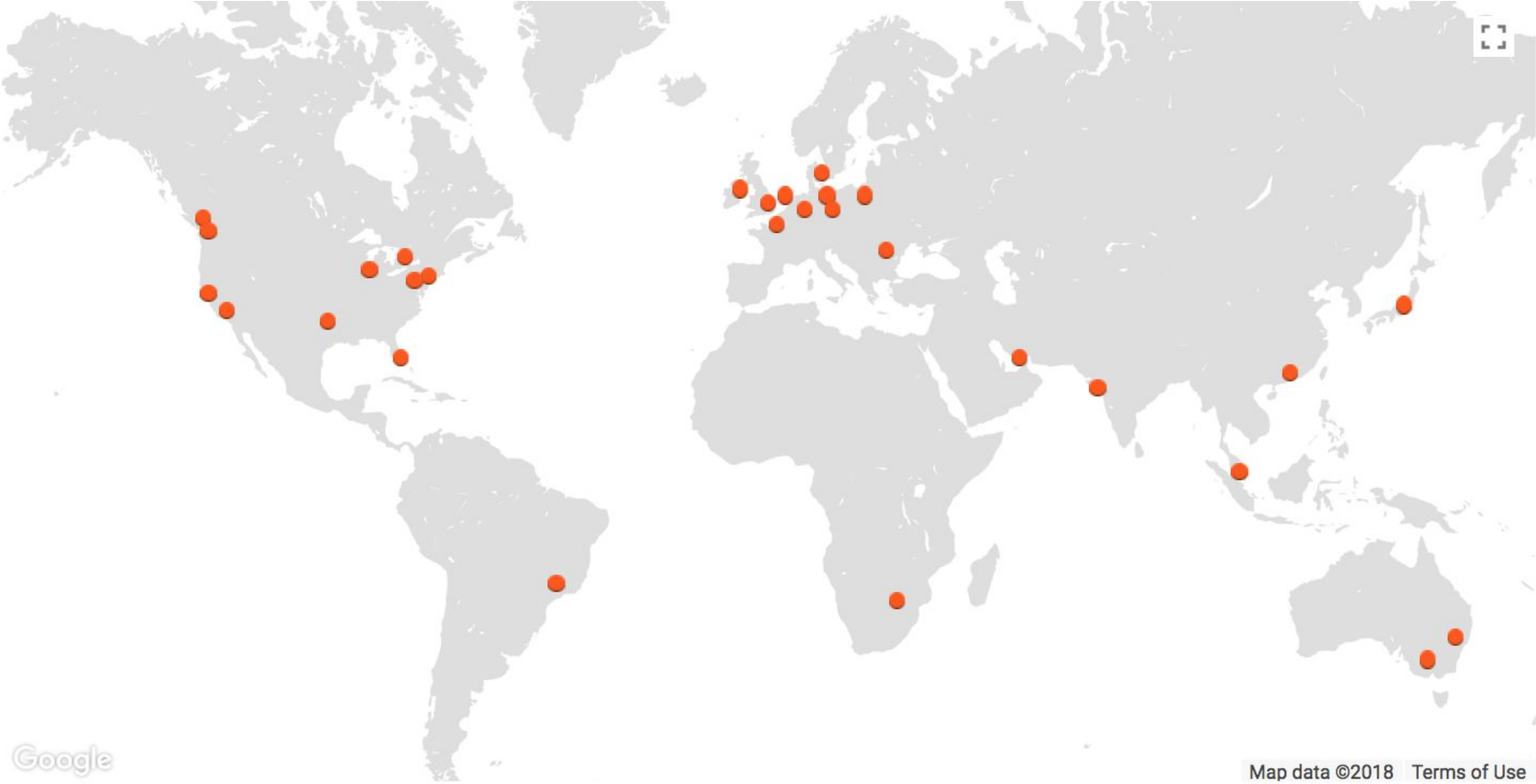
Cisco Umbrella Datasets

Recursive DNS Data

- 28 data centers worldwide
- ~150 billion queries a day
- Translates to around 24 TB a day
- Valuable client query information



Cisco Umbrella data center locations



Open source Datasets

SSL Data

- SSL data collected from internet wide scan of IPv4 space

- Store and retrieve over 2 TB of total data
  - Scalability
  - Speed
  - Flexibility

- Primary source of data (scans.io)
  - Secondary – active scanning

Processing

Challenges Working at Cisco Umbrella Scale

1. Logs are big:
   - We've peaked at over 150 billion user queries per day and growing
2. Interesting algorithms are slow:
   - Always worse than linear, and sometimes far worse
3. Event horizons move:
   - the data you need to make your new idea work is always one day past the current retention window

Lessons Learned

- Currently somewhere between the 4th and 5th iteration of our systems
- Algorithms are always improving and systems are always getting faster, but...

The most cost effective way to improve your search performance is almost always to reduce the size of your search space

The Internet Is...

Noisy

- People doing scans and scraping expeditions
- Misconfigured search domains
- Infrastructure chattering away
- Low value/inconsequential entities fighting each other

The Internet Is...

Repetitive

- Content distribution networks - everyone needs to download that latest javascript framework
- Low TTLs to aid fault tolerance mean clients need to constantly ask the same questions
- The incessant push to move everything to "the cloud"

The Internet Is...

Boring (for the most part)

- Typically, only 5-10% of our raw logs are useful for threat hunting purposes
- Need to strike a balance between cost effective scaling and losing a small amount of signal

Analysis

Spike Detection

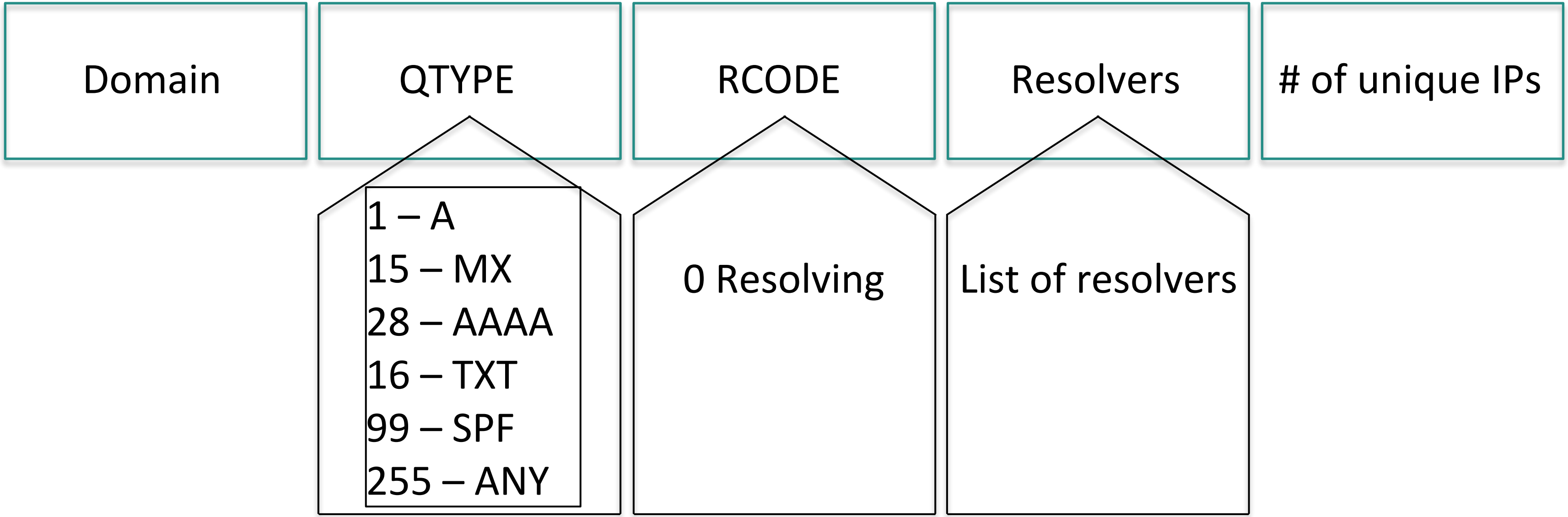
Datasets

Goals - Recursive DNS Data

- Can we identify exploit kit and ransomware domains from DNS client traffic?
- Examine traffic logs for possible signals

Datasets

Recursive DNS data



Domain

QTYPE

1 – A
15 – MX
28 – AAAA
16 – TXT
99 – SPF
255 – ANY

RCODE

0 Resolving

Resolvers

List of resolvers

# of unique IPs

DNS Features Taxonomy

Assigned

-Lexical
-DGA setup
-Hosting
-Registration

Inherent

-DNS query trends
-Diversity of clients across geography and IP space
-DNS query volume
-Query types
-Number of querying IPs
-Distribution of queries across resolvers

Harder to obfuscate and change by actors at global scale

Classification

Recursive DNS Data

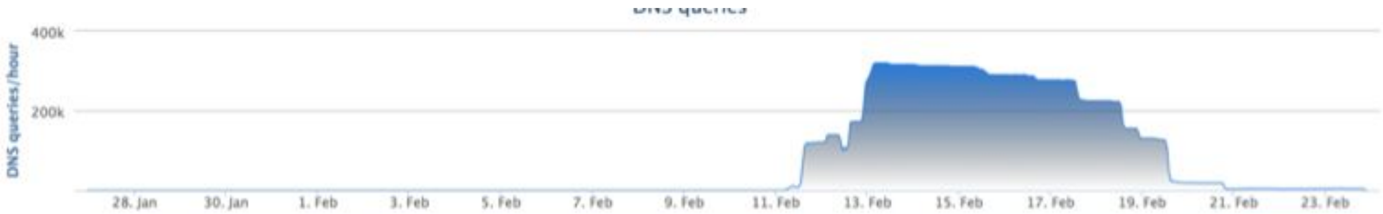
- Classify domains based on two sets of features:
  - Spike DNS data
  - Historical query volume patterns

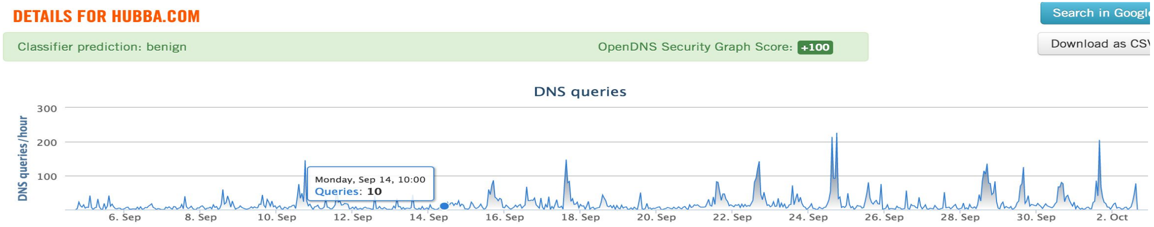
- Spike DNS data
  - Qtype distributions
  - Resolver distributions

- Historical query data:
  - Volatility
  - Sparsity

Classification

Recursive DNS Data



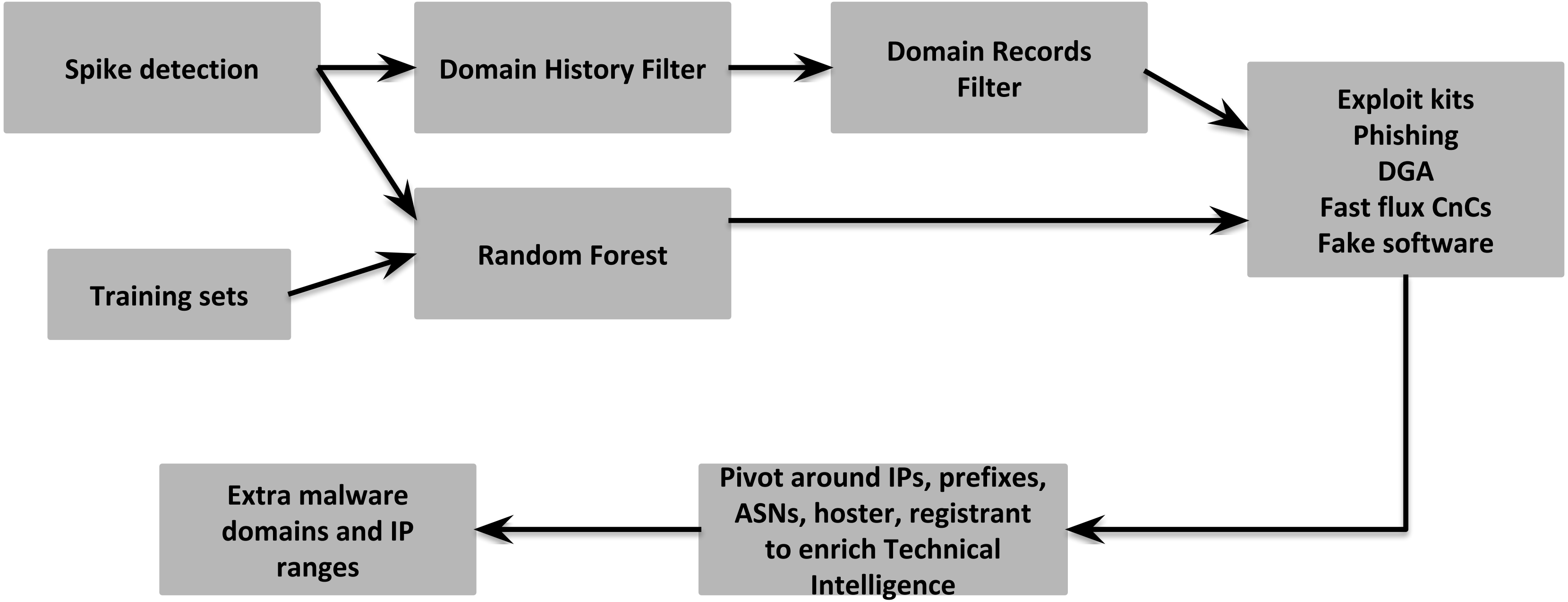

DETAILS FOR HUBBA.COM

Classifier prediction: benign

OpenDNS Security Graph Score: +100

Spike Detection pipeline

Recursive DNS Data



Classification

Random Forests

- Use random forest for classification step
  - Random forests parallelize easily
  - Non-parametric model
  - Handle non-linear boundaries
- Feed in spike domain feature vectors hourly to classify spiked domains
- Out-of-bag error at 3%

Classification

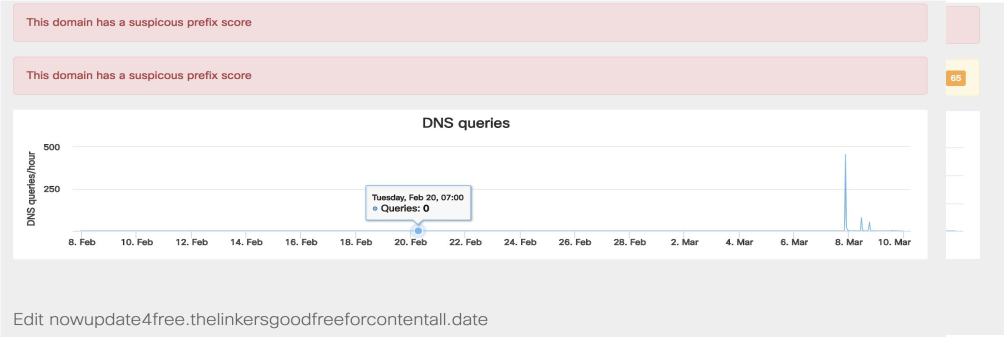
Results

- Two types of domains caught:
    - Dedicated
    - Compromised

- Have different time series for the two classes
- Compromised domains more difficult to detect due to the presence of additional noise

- ‘nowupdate4free.thelinkersgoodfreeforcontentall.date’

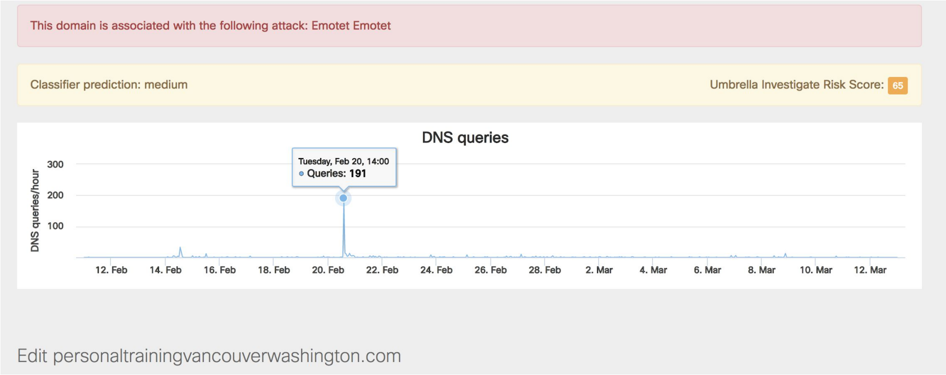
- ‘vancouverwashingtonpersonaltraining.com’



This domain has a suspicous prefix score

This domain has a suspicous prefix score

Edit nowupdate4free.thelinkersgoodfreeforcontentall.date



This domain is associated with the following attack: Emotet Emotet

Classifier prediction: medium

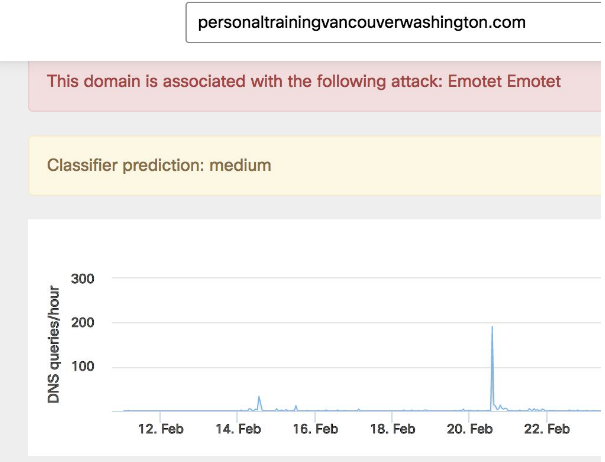
Umbrella Investigate Risk Score: 65

Edit personaltrainingvancouverwashington.com

Classification

Results

- Layering of signals post initial classification
- Good method for identifying compromised domains



personaltrainingvancouverwashington.com

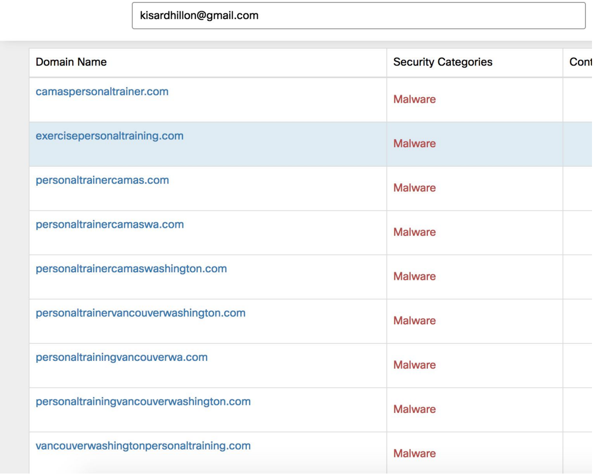
This domain is associated with the following attack: Emotet Emotet

Classifier prediction: medium

Classification

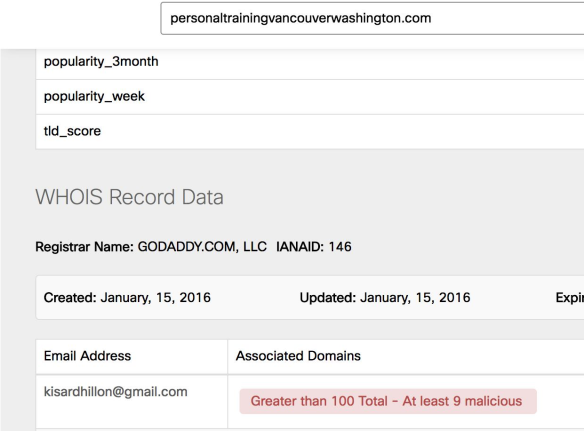
Results

- Use available WHOIS data to pivot
  - Search through other registrant domains
  - Look for similar signal patterns
- Pivoting done through a combination of manual and automated work
- A whole set of compromised domains are found
  - Able to block these compromised domains ahead of going live



kisardhillon@gmail.com

Domain Name	Security Categories	Cont
camaspersonaltrainer.com	Malware	
exercisepersonaltraining.com	Malware	
personaltrainercamas.com	Malware	
personaltrainercamaswa.com	Malware	
personaltrainercamaswashington.com	Malware	
personaltrainervancouverwashington.com	Malware	
personaltrainingvancouverwa.com	Malware	
personaltrainingvancouverwashington.com	Malware	
vancouverwashingtonpersonaltraining.com	Malware	



personaltrainingvancouverwashington.com

popularity_3month

popularity_week

tld_score

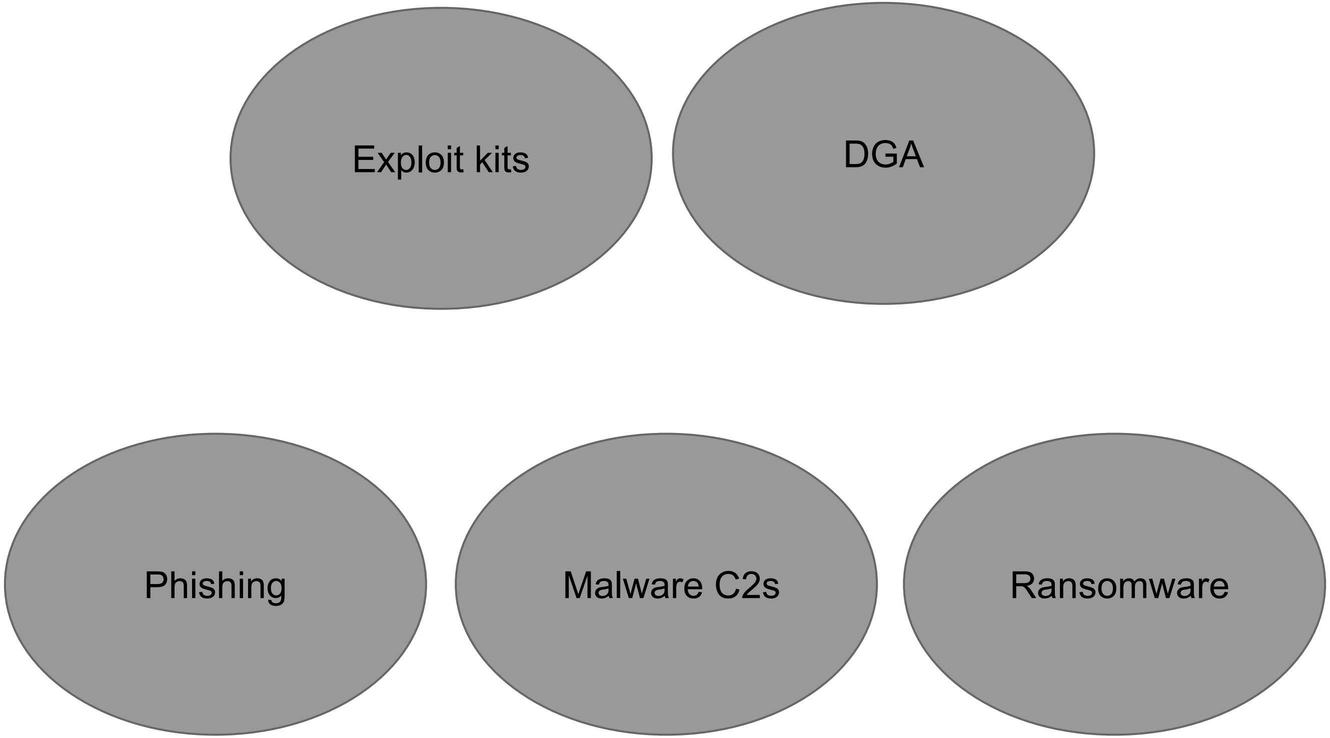
WHOIS Record Data

Registrar Name: GODADDY.COM, LLC **IANAID:** 146

Created: January, 15, 2016 **Updated:** January, 15, 2016 **Expi**

Email Address	Associated Domains
kisardhillon@gmail.com	Greater than 100 Total - At least 9 malicious

Detected Threats



Exploit kits

DGA

Phishing

Malware C2s

Ransomware

Track Malspam; Combining techniques

Path of a malspam attack

1 **Phishing email sent from delta@performanceair.com**

2 **Victims click on malicious URLs**

myhearthstonehomes.org
ourrealtyguy.info
ourrealtyguy.org
ourrealtyguy.us
package2china.com

3 **Malicious word doc drops Hancitor**

4 **Hancitor makes C2 call to domains for trojans**

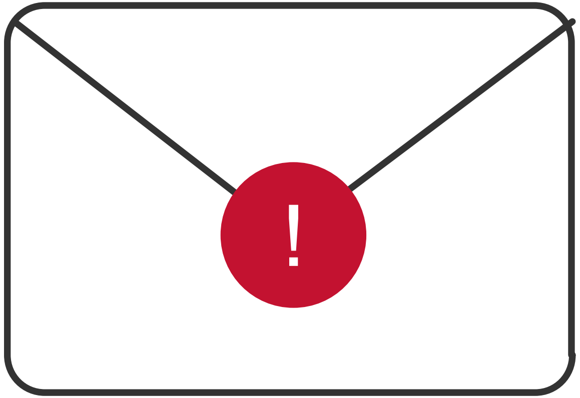
uneventrendi.com
ketofonerof.ru
thettertrefbab.ru

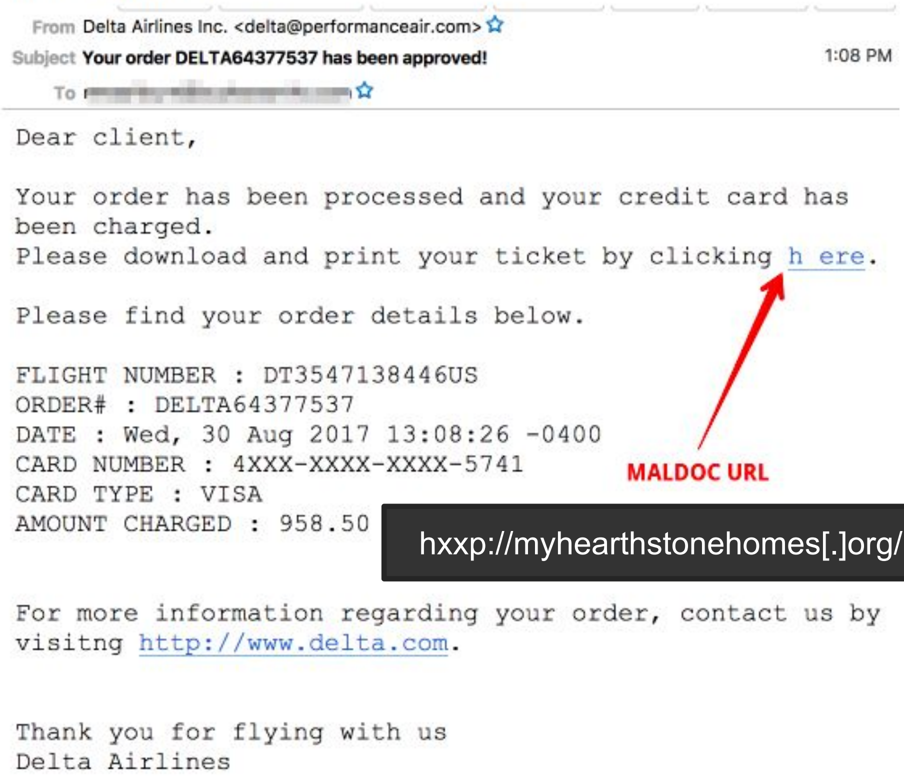
5 **Trojans (Pony, Evil Pony, Zloader) make C2 call for extra malware or functionality**

mebelucci.com.ua
uneventrendi.com
lycasofrep.com
rinbetarrab.com

6 **Infection on device & positioned for data extraction**

Malicious malspam campaign





From Delta Airlines Inc. <delta@performanceair.com>

Subject **Your order DELTA64377537 has been approved!** 1:08 PM

To [redacted]

```
Dear client,

Your order has been processed and your credit card has
been charged.
Please download and print your ticket by clicking h ere.

Please find your order details below.

FLIGHT NUMBER : DT3547138446US
ORDER# : DELTA64377537
DATE : Wed, 30 Aug 2017 13:08:26 -0400
CARD NUMBER : 4XXX-XXXX-XXXX-5741
CARD TYPE : VISA
AMOUNT CHARGED : 958.50


For more information regarding your order, contact us by
visitng http://www.delta.com.


Thank you for flying with us
Delta Airlines
```

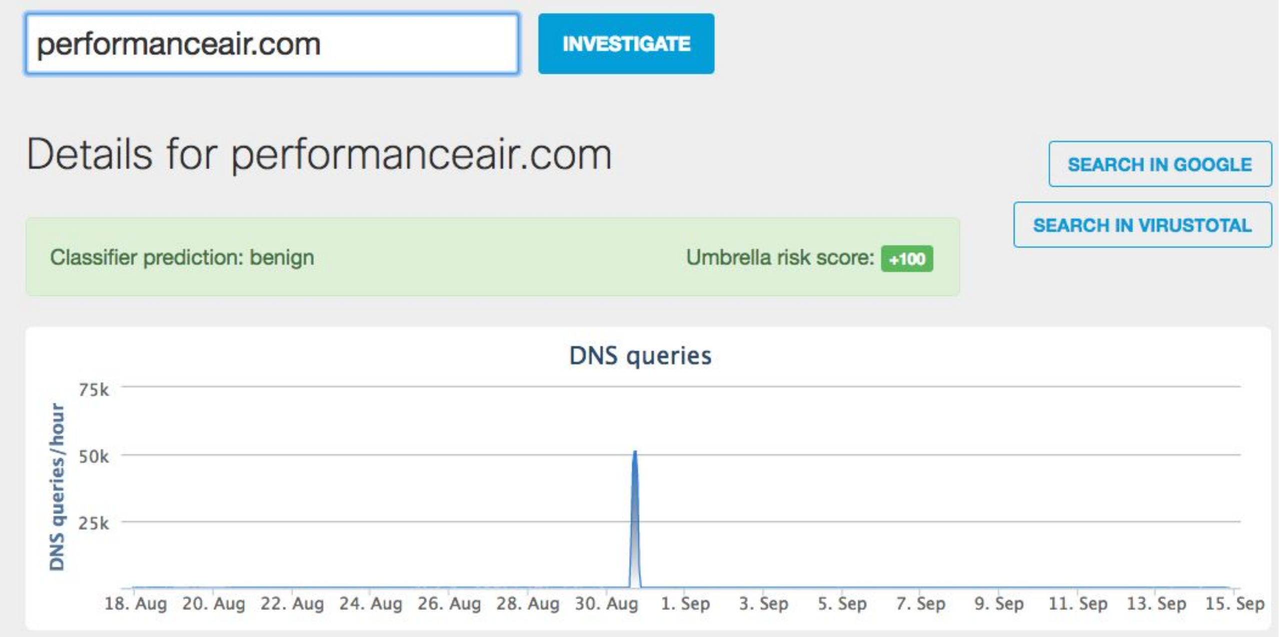
MALDOC URL

hxxp://myhearthstonehomes[.]org/i.php?d=

Reference hazmalware.wordpress.com

performanceair.com

Spoofed email used in mailspam attack

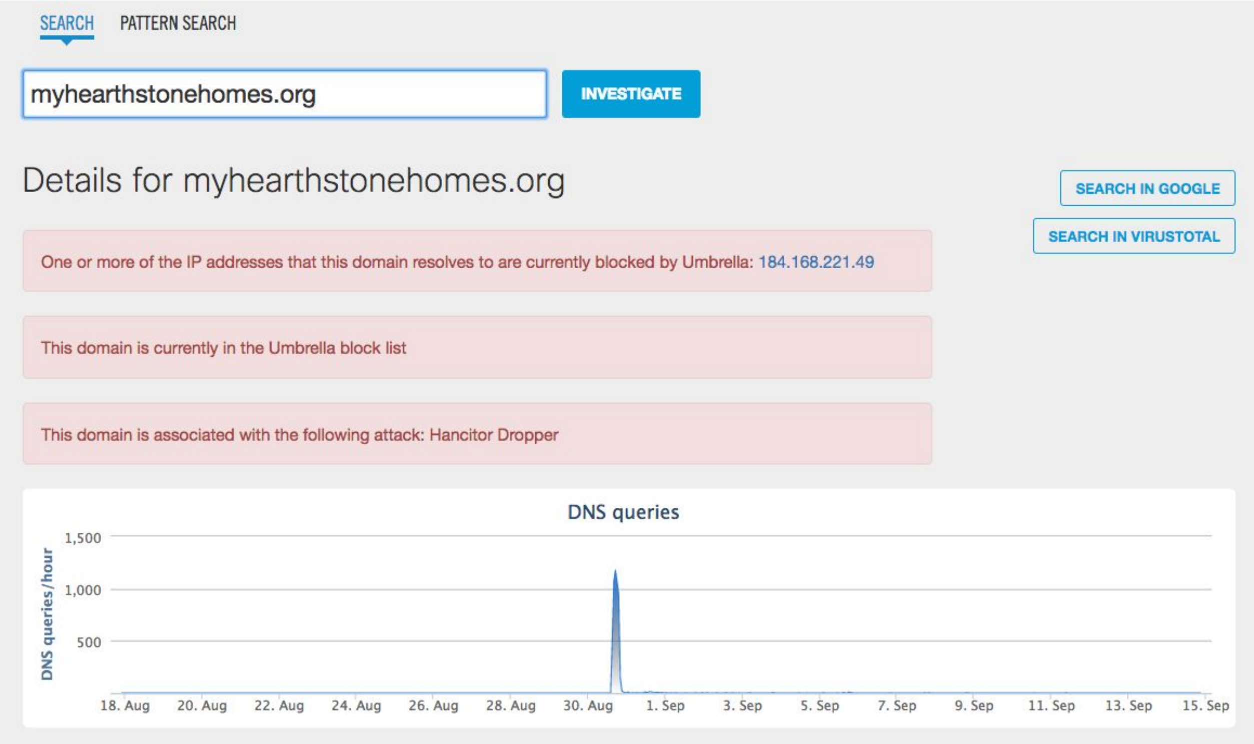


Details for performanceair.com

Classifier prediction: benign

Umbrella risk score: +100

August 30: Peak of malicious redirect



Details for myhearthstonehomes.org

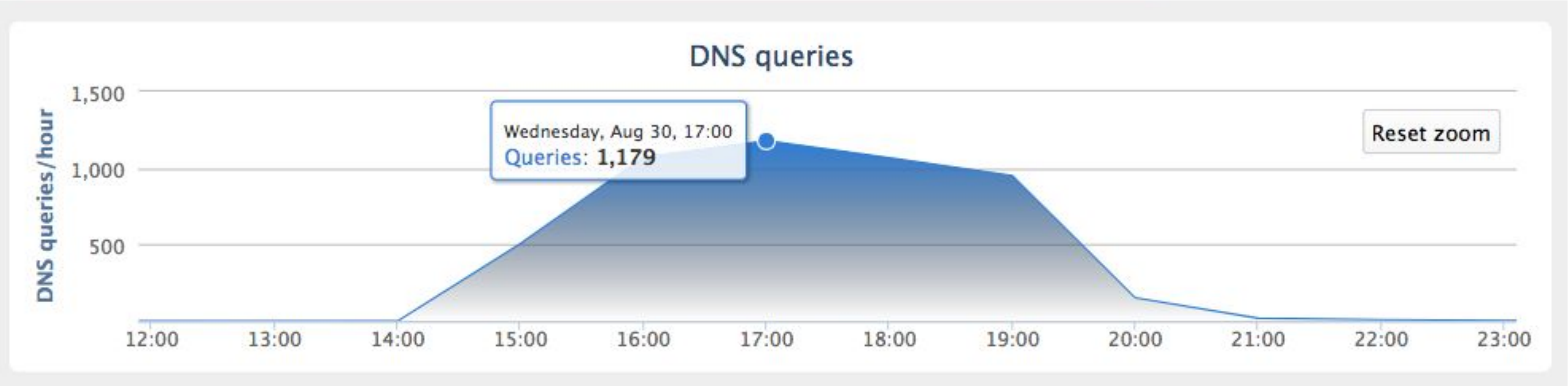
One or more of the IP addresses that this domain resolves to are currently blocked by Umbrella: 184.168.221.49

This domain is currently in the Umbrella block list

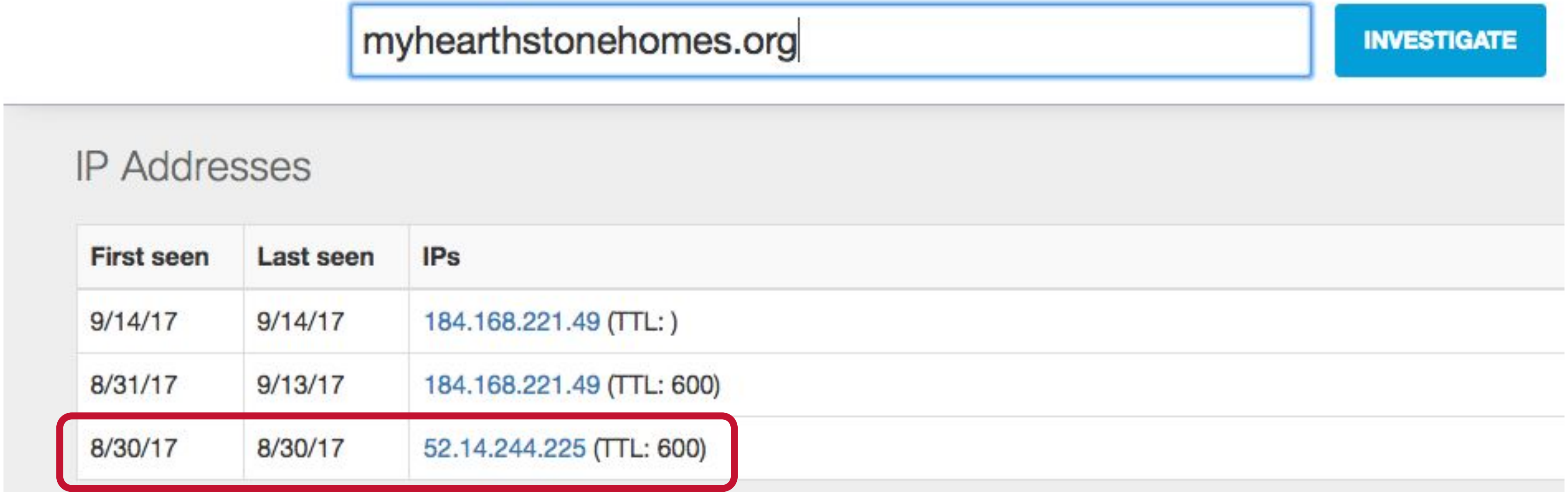
This domain is associated with the following attack: Hancitor Dropper

Duration: 7 hour period

Attack took place between 14:00-21:00 UTC



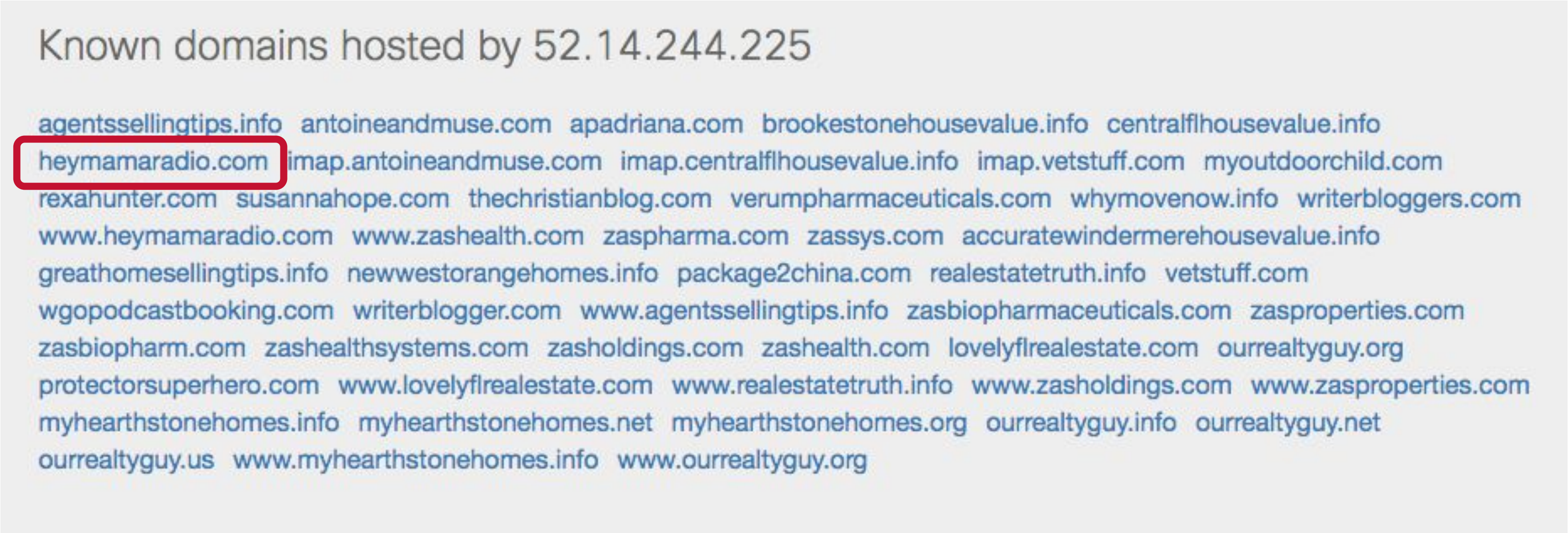
Insight into the IP network



IP Addresses

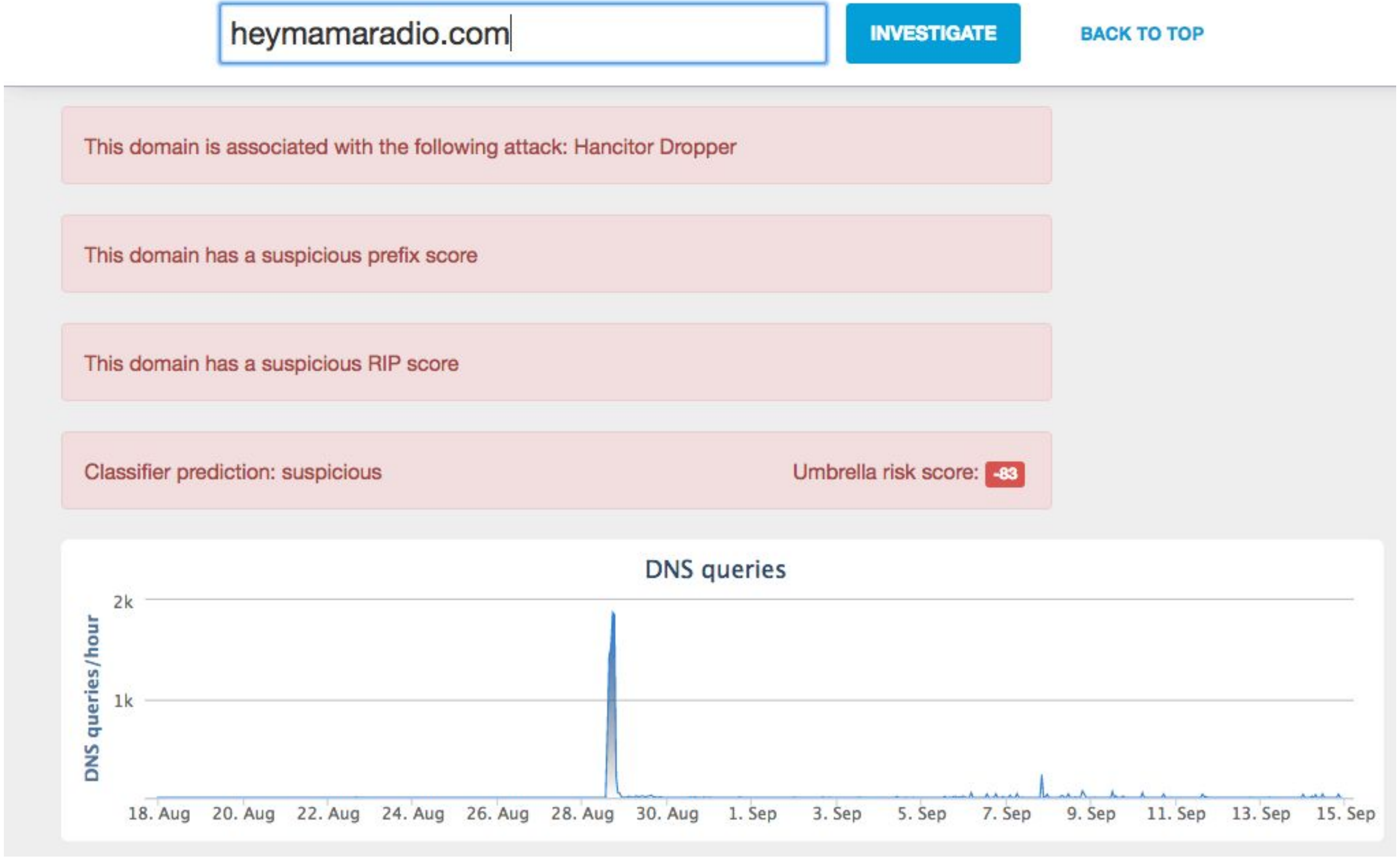
First seen	Last seen	IPs
9/14/17	9/14/17	184.168.221.49 (TTL:)
8/31/17	9/13/17	184.168.221.49 (TTL: 600)
8/30/17	8/30/17	52.14.244.225 (TTL: 600)

Known malicious domains on the same IP



Known domains hosted by 52.14.244.225

agentssellingtips.info antoineandmuse.com apadriana.com brookestonehousevalue.info centralflhousevalue.info
heymamaradio.com imap.antoineandmuse.com imap.centralflhousevalue.info imap.vetstuff.com myoutdoorchild.com
rexahunter.com susannahope.com thechristianblog.com verumpharmaceuticals.com whymovenow.info writerbloggers.com
www.heymamaradio.com www.zashealth.com zaspharma.com zassys.com accuratewindermerehousevalue.info
greathomesellingtips.info newwestorangehomes.info package2china.com realestatetruth.info vetstuff.com
wgopodcastbooking.com writerblogger.com www.agentssellingtips.info zasbiopharmaceuticals.com zasproperties.com
zasbiopharm.com zashealthsystems.com zasholdings.com zashealth.com lovelyflrealestate.com ourrealtyguy.org
protectorsuperhero.com www.lovelyflrealestate.com www.realestatetruth.info www.zasholdings.com www.zasproperties.com
myhearthstonehomes.info myhearthstonehomes.net myhearthstonehomes.org ourrealtyguy.info ourrealtyguy.net
ourrealtyguy.us www.myhearthstonehomes.info www.ourrealtyguy.org



heymamaradio.com

This domain is associated with the following attack: Hancitor Dropper

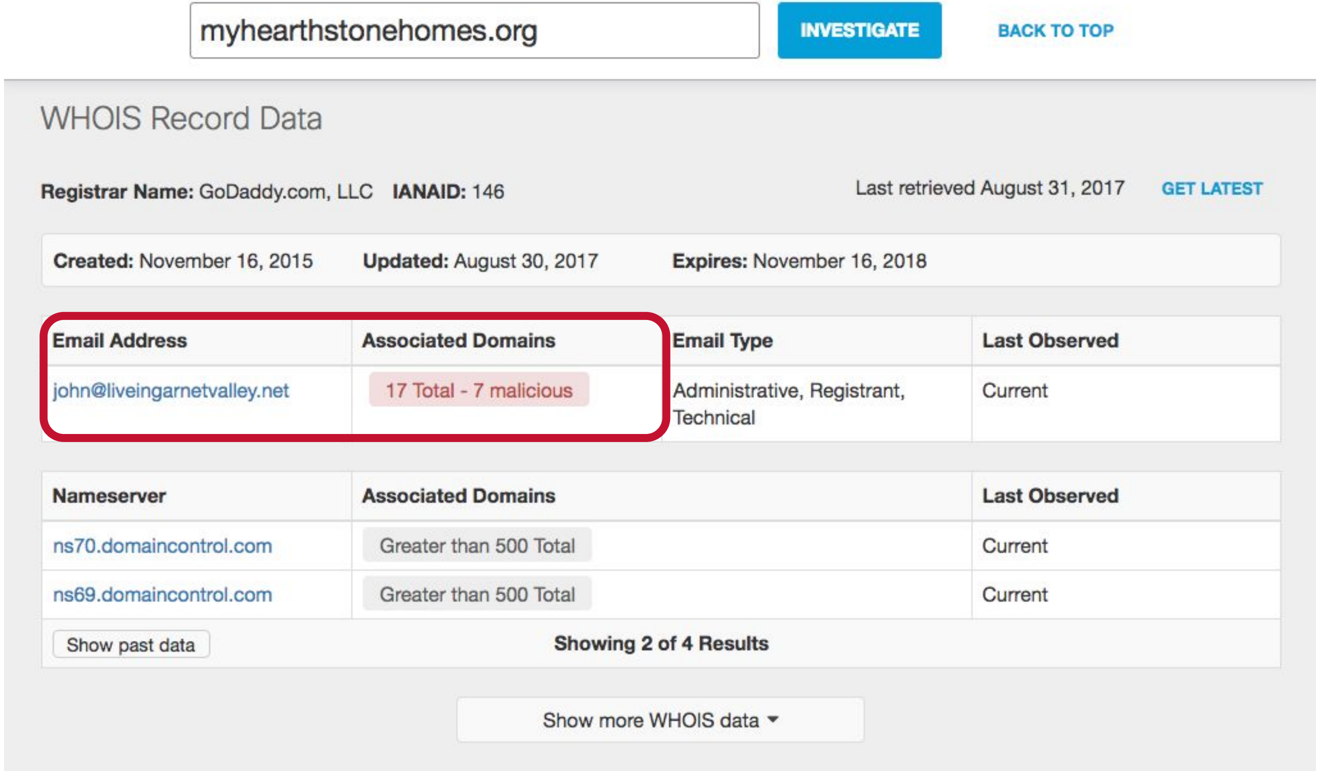
This domain has a suspicious prefix score

This domain has a suspicious RIP score

Classifier prediction: suspicious

Umbrella risk score: -83

WHOIS information of myhearthstronehomes.org



WHOIS Record Data

Registrar Name: GoDaddy.com, LLC **IANAID:** 146

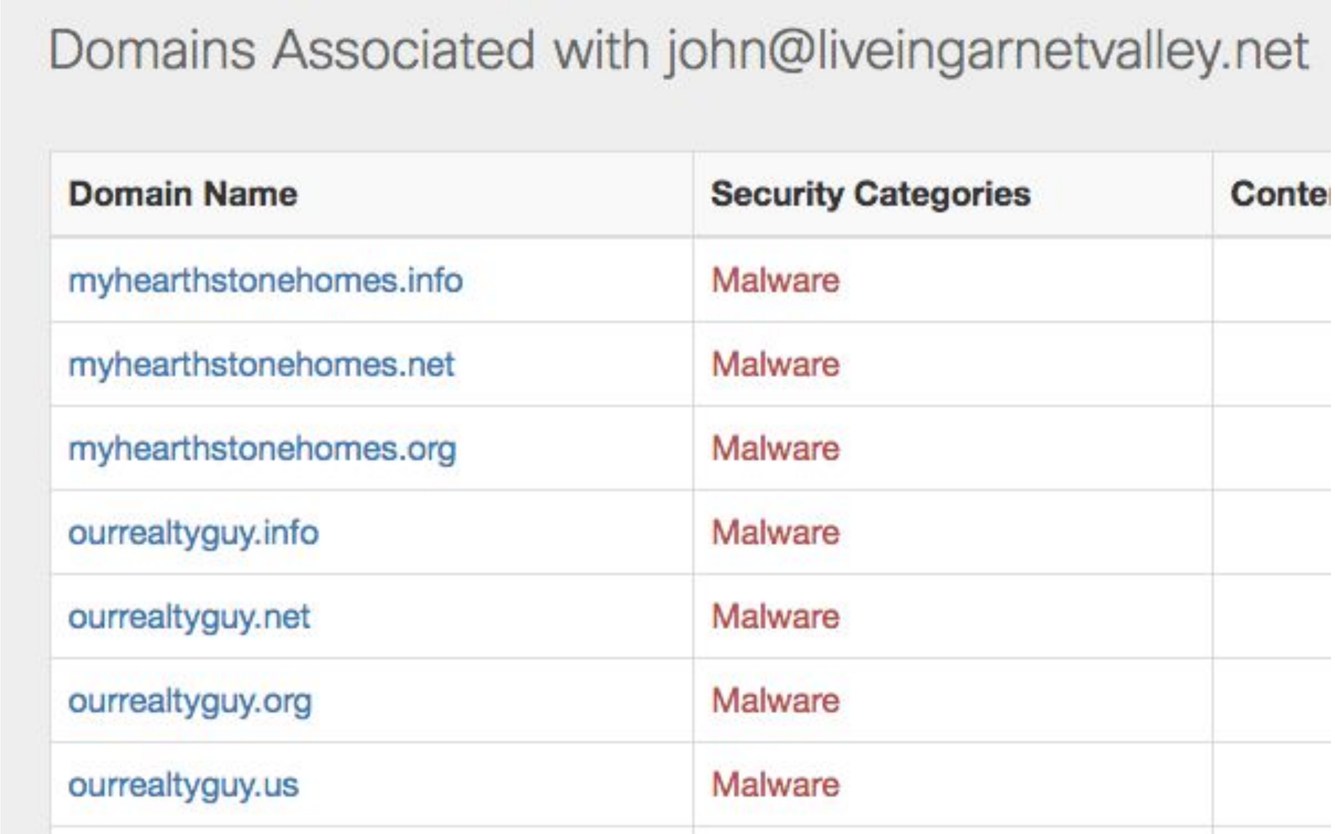
Last retrieved August 31, 2017 GET LATEST

Created: November 16, 2015 **Updated:** August 30, 2017 **Expires:** November 16, 2018

Email Address	Associated Domains	Email Type	Last Observed
john@liveingarnetvalley.net	17 Total - 7 malicious	Administrative, Registrant, Technical	Current

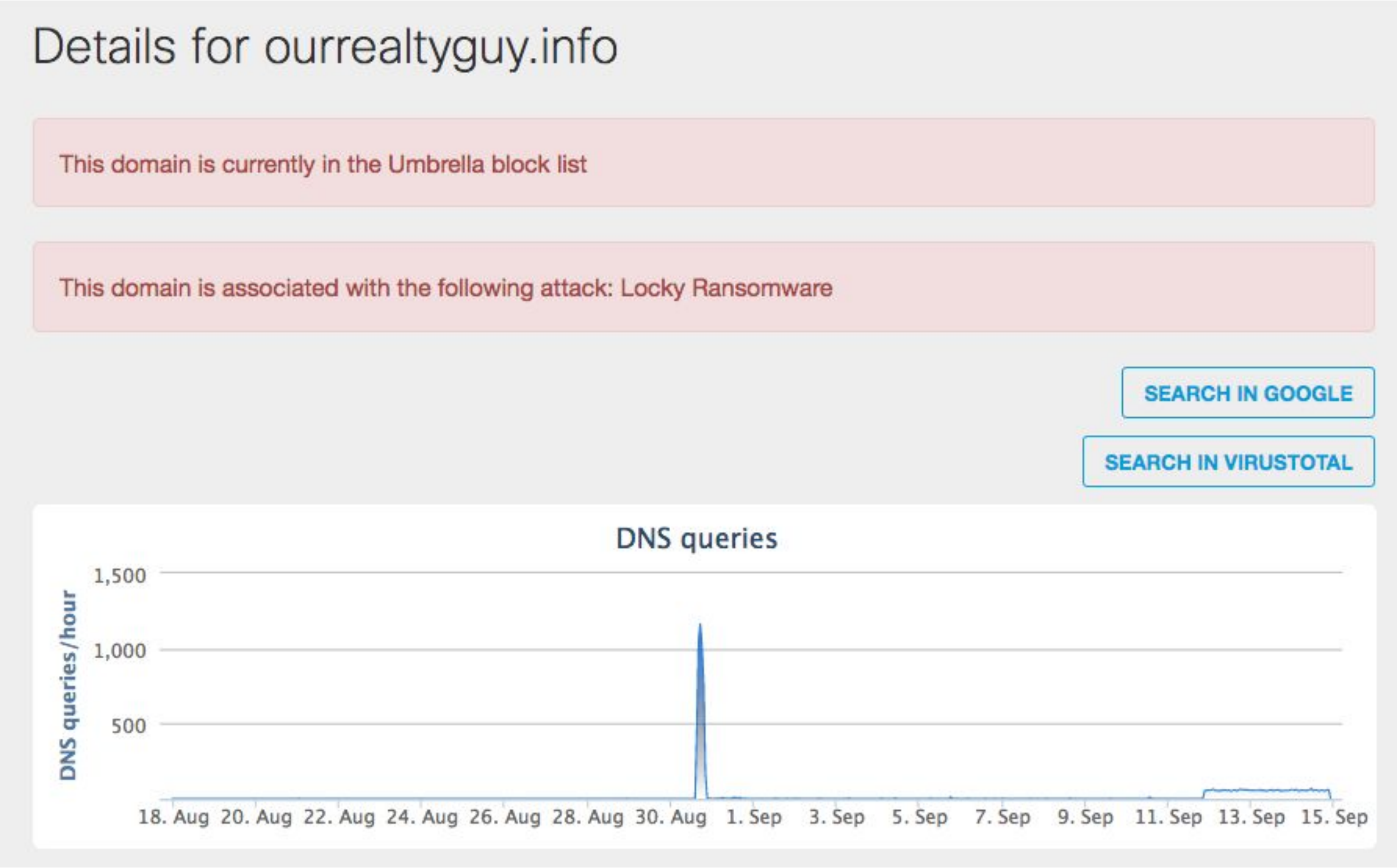
Nameserver	Associated Domains	Last Observed
ns70.domaincontrol.com	Greater than 500 Total	Current
ns69.domaincontrol.com	Greater than 500 Total	Current

Showing 2 of 4 Results



Domains Associated with john@liveingarnetvalley.net

Domain Name	Security Categories	Conte
myhearthstonehomes.info	Malware	
myhearthstonehomes.net	Malware	
myhearthstonehomes.org	Malware	
ourrealtyguy.info	Malware	
ourrealtyguy.net	Malware	
ourrealtyguy.org	Malware	
ourrealtyguy.us	Malware	

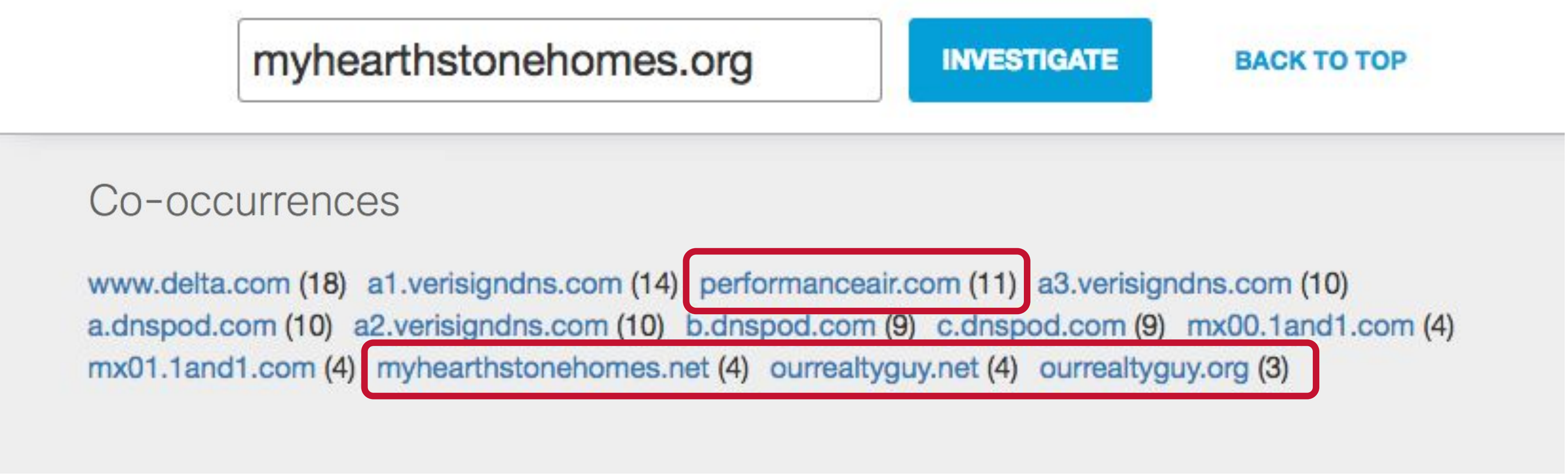


Details for ourrealtyguy.info

This domain is currently in the Umbrella block list

This domain is associated with the following attack: Locky Ransomware

Co-occurring domains tied to the same malspam campaign



Co-occurrences

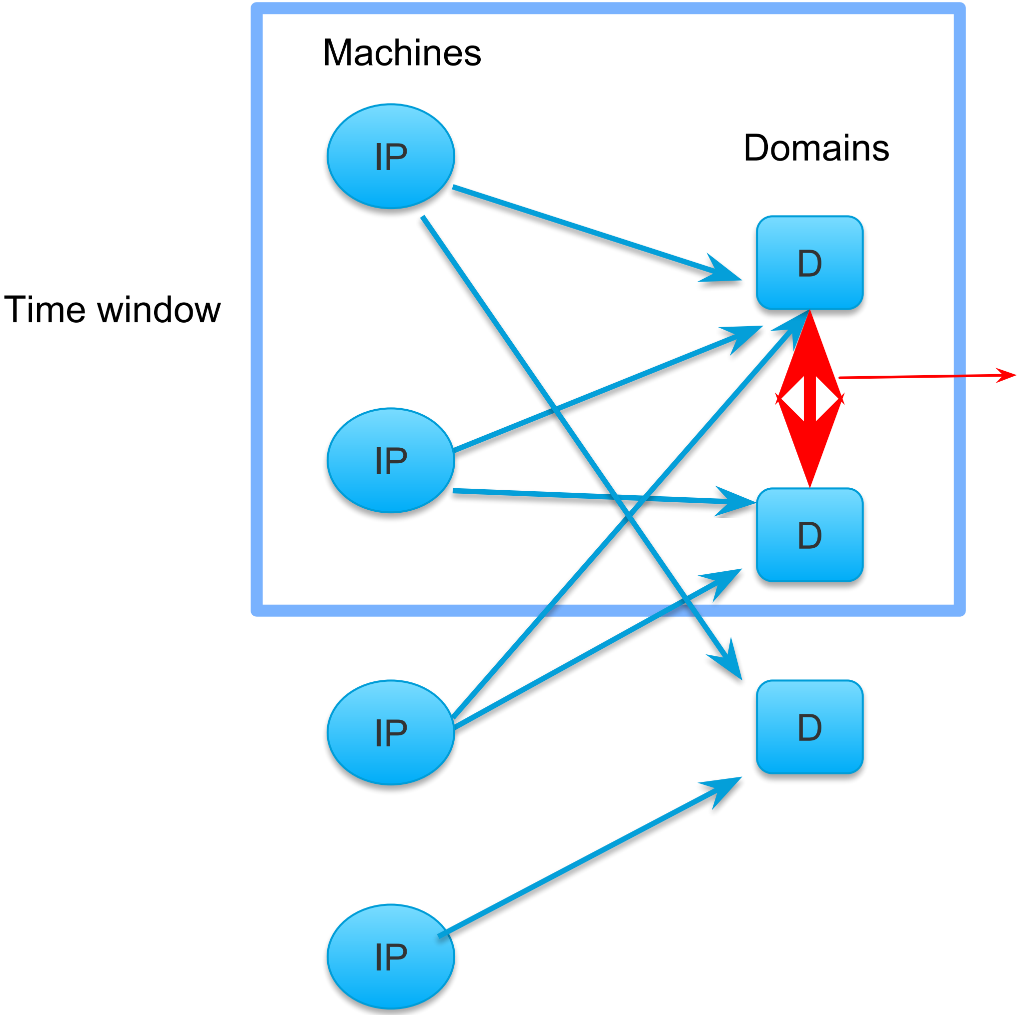
www.delta.com (18) a1.verisigndns.com (14) performanceair.com (11) a3.verisigndns.com (10) a.dnspod.com (10) a2.verisigndns.com (10) b.dnspod.com (9) c.dnspod.com (9) mx00.1and1.com (4) mx01.1and1.com (4) myhearthstonehomes.net (4) ourrealtyguy.net (4) ourrealtyguy.org (3)

Co-occurrences

- Find domains queried by clients in close temporal proximity
- Data is one hour of querylog traffic - 2TB of raw data
- Identify domains looked up by same clients within one minute window of one another
- Output {domain: [List of Domains]}
- Example:

  100luimg.361lu.com. -> {"ucsec1.ucweb.com":3.0,"d2.avgc.us":3.0,"home.1100lu.info":4.0}

Co-occurrences

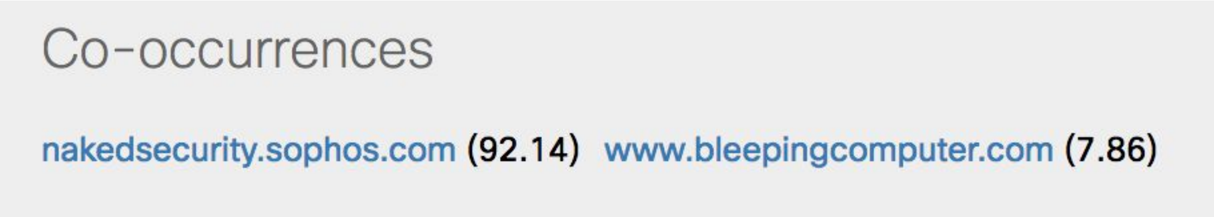


Edge in the co-occurrence graph

- The closer in time, the higher the co-occurrence score
- The more clients exhibiting this behavior, the higher the score

Co-occurrences

- Domains having similar topic, e.g. security sites, hacking, carding sites
  - Visited by users with related interest
- Example: first.org



Co-occurrences

nakedsecurity.sophos.com (92.14) www.bleepingcomputer.com (7.86)

- Botnet CnC domains, e.g. DGAs
- Infection chains: compromised sites -> Exploit kit landing domains

Scaling Up co-occurrence detection algorithms

Before	After
job ran daily	job runs hourly
job used heavily sampled logs	no sampling apart from initial "algorithm relevant" cleaning
heuristics used to further cut down data size to help catch initial compromise/infection	no further data size reduction necessary

Path of malspam attack

1 **Phishing email sent from delta@performanceair.com**

2 **Victims click on malicious URLs**

myhearthstonehomes.org
ourrealtyguy.info
ourrealtyguy.org
ourrealtyguy.us
package2china.com

3 **Malicious word doc drops Hancitor**

4 **Hancitor makes C2 call to domains for trojans**

uneventrendi.com
ketofonerof.ru
thettertrefbab.ru

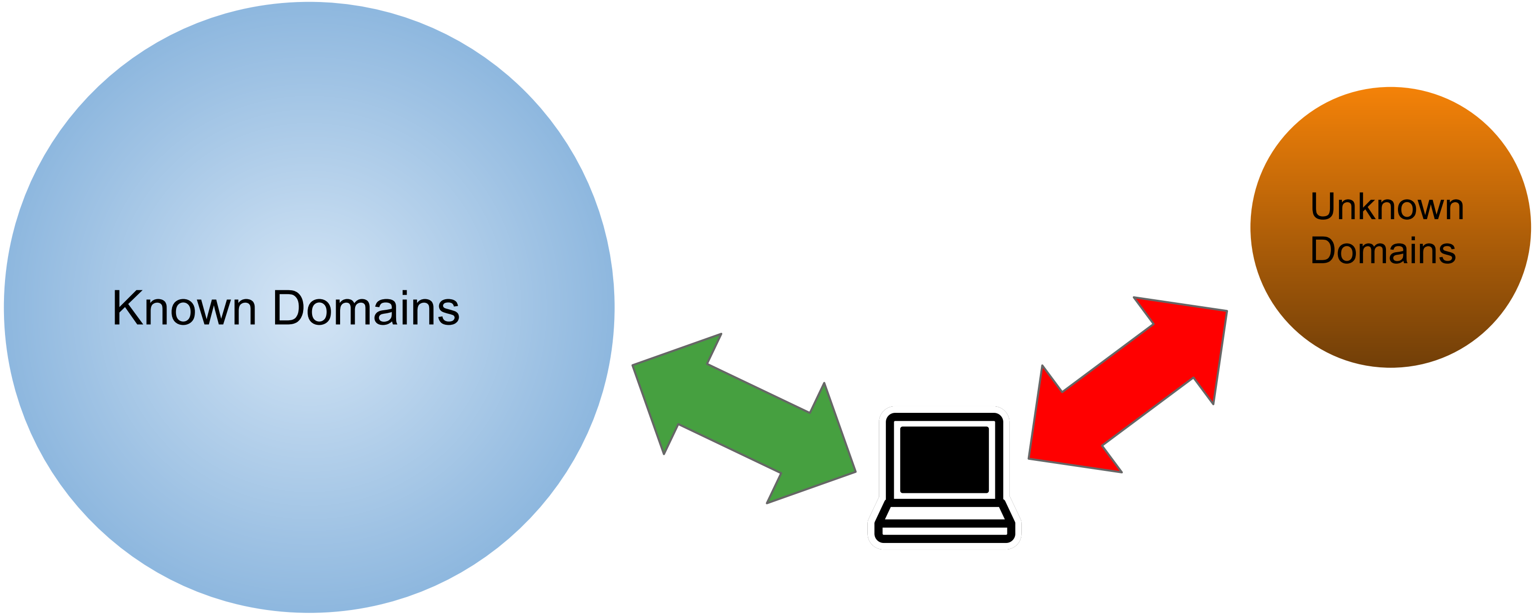
Newly seen domains

5 **Trojans (Pony, Evil Pony, Zloader) make C2 call for extra malware or functionality**

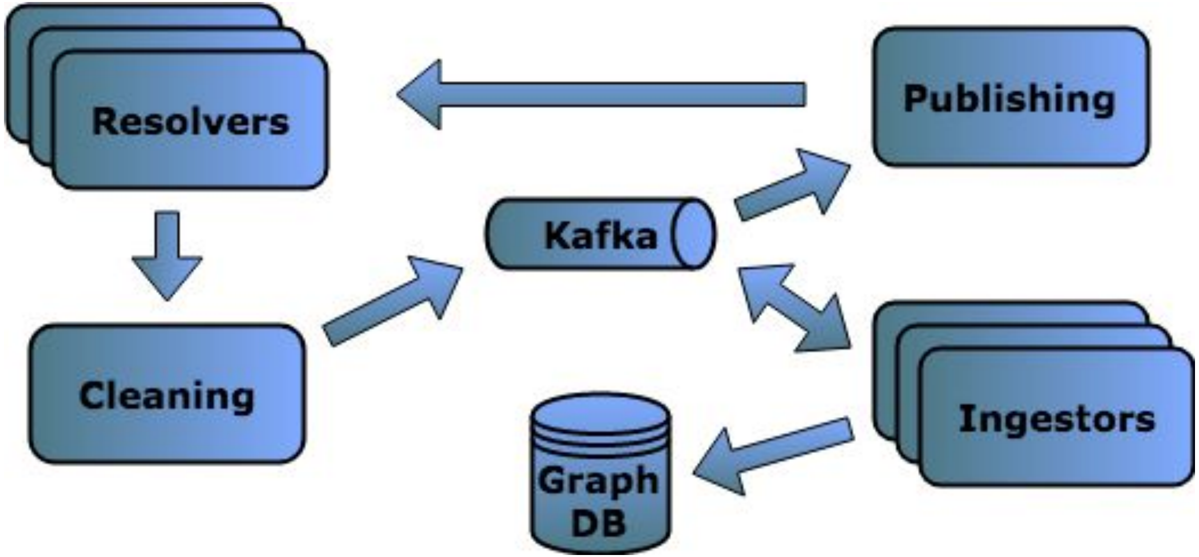
mebelucci.com.ua
uneventrendi.com
lycasofrep.com
rinbetarrab.com

6 **Infection on device & positioned for data extraction**

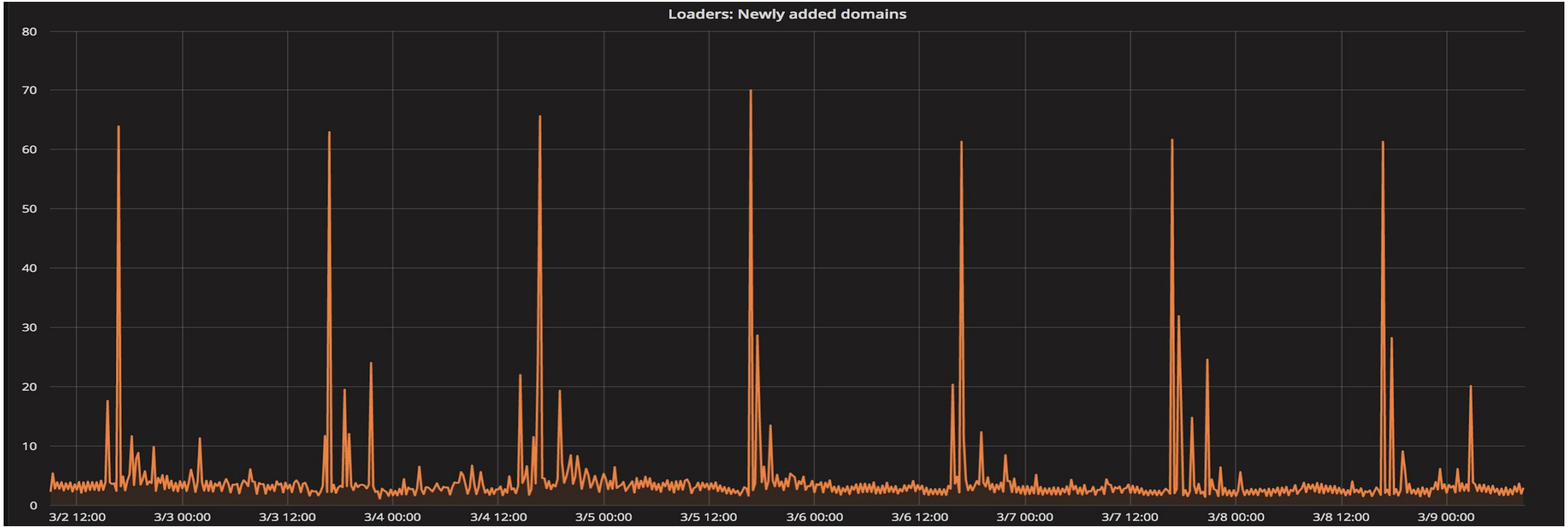
Newly Seen Domains: A real-time Stream



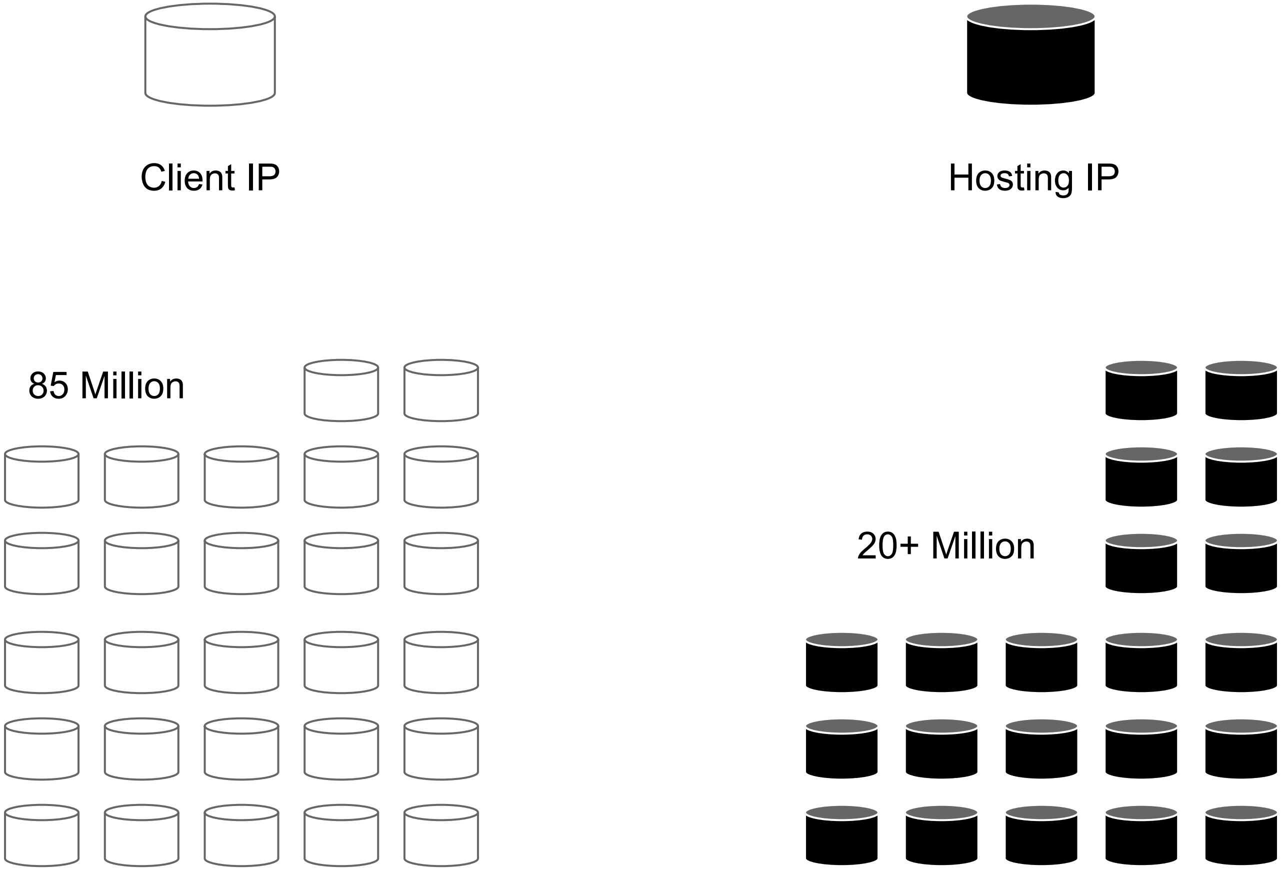
Overview of Newly Seen Domains System



Clear Patterns Emerge



Cybercrime sites



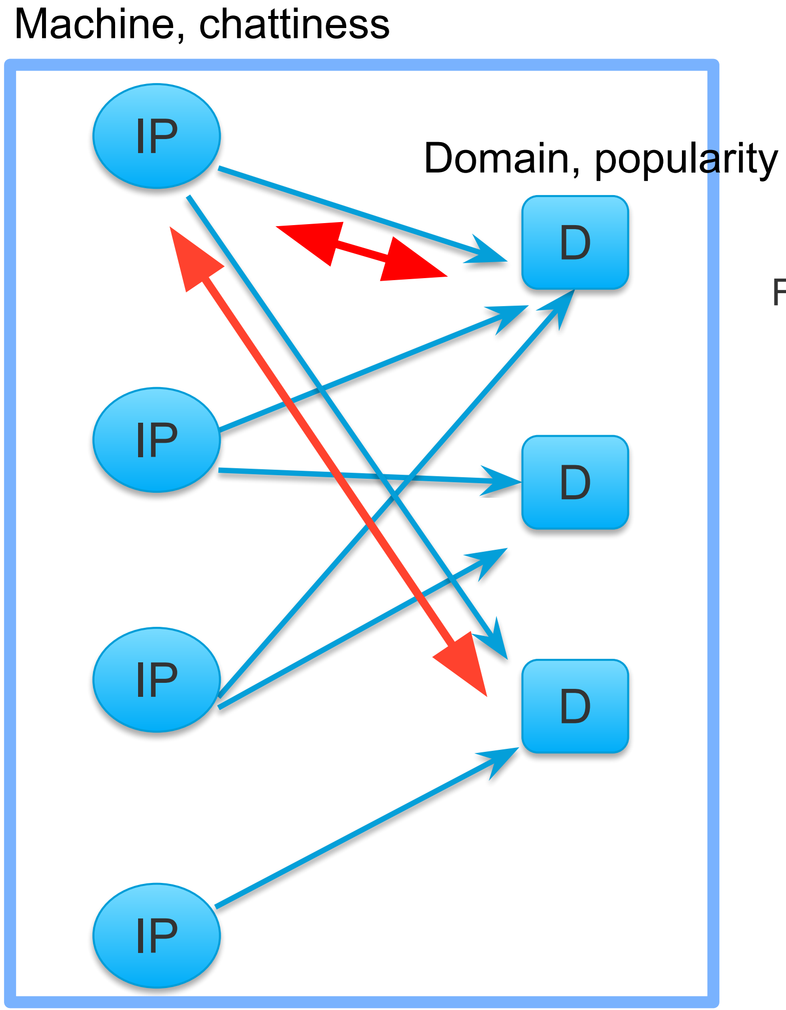
Client IP

Hosting IP

85 Million

20+ Million

Amplify through signals using seeds

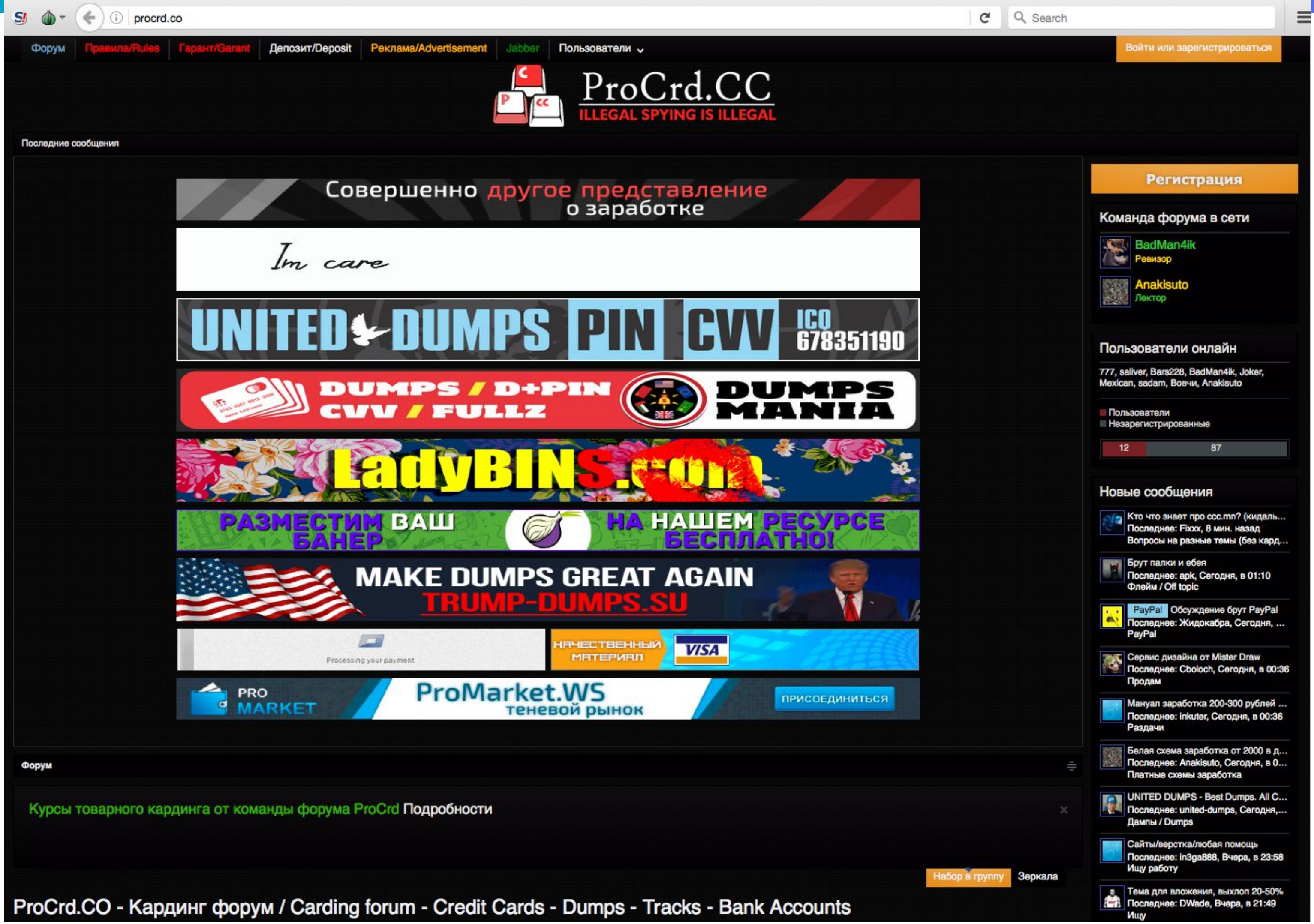


For every 1 hour of traffic, we define:

- Chattiness: # unique domains a machine queries
- Popularity: # unique machines that queried the domain

- amplify dom domain chattiness popularity nbhourspast level

- Pivot through domains and machines by keeping a threshold of chattiness and popularity



Последние сообщения

Форум

Курсы товарного кардинга от команды форума ProCrd Подробности

ProCrd.CO - Кардинг форум / Carding forum - Credit Cards - Dumps - Tracks - Bank Accounts

Команда форума в сети

BadMan4ik
Ревизор

Anakisuto
Лектор

Пользователи онлайн

777, saliver, Bars228, BadMan4ik, Joker, Mexican, sadam, Вовчи, Anakisuto

🟥 Пользователи
▨ Незарегистрированные

12 87

Новые сообщения

Кто что знает про ccc.mn? (кидаль...
Последнее: Fixxx, 8 мин. назад
Вопросы на разные темы (без кард...

Брут палки и ебея
Последнее: apk, Сегодня, в 01:10
Флейм / Off topic

PayPal Обсуждение брут PayPal
Последнее: Жидокабра, Сегодня, ...
PayPal

Сервис дизайна от Mister Draw
Последнее: Cboloch, Сегодня, в 00:36
Продам

Мануал заработка 200-300 рублей ...
Последнее: inkuter, Сегодня, в 00:36
Раздачи

Белая схема заработка от 2000 в д...
Последнее: Anakisuto, Сегодня, в 0...
Платные схемы заработка

UNITED DUMPS - Best Dumps. All C...
Последнее: united-dumps, Сегодня,...
Дампы / Dumps

Сайты/верстка/любая помощь
Последнее: in3ga888, Вчера, в 23:58
Ищу работу

Тема для вложения, выхлоп 20-50%
Последнее: DWade, Вчера, в 21:49
Ищу

Amplify through signals using seeds

Pivot from procrd.co -> other crimeware sites

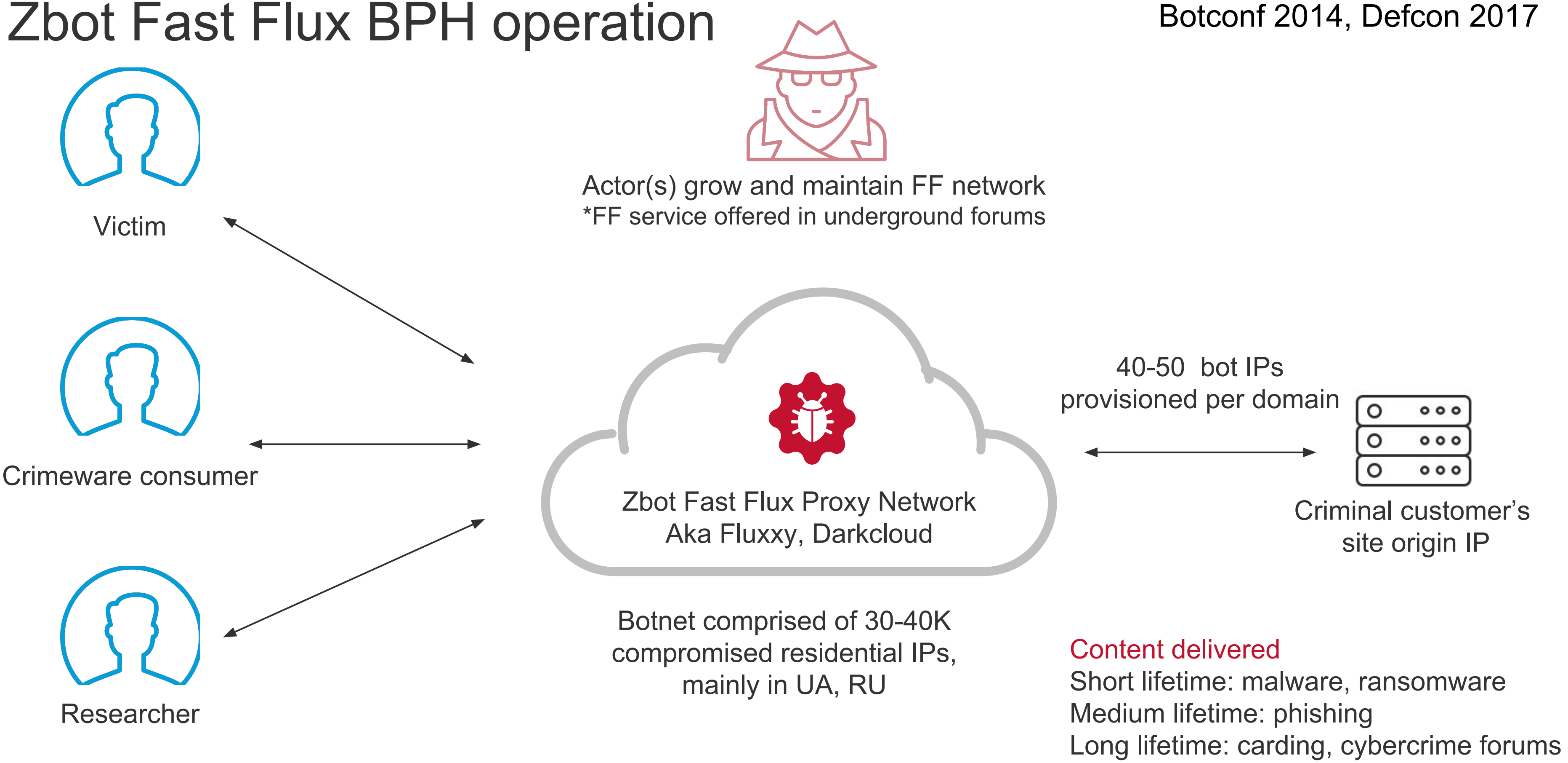
```
amplify.sh dom domain chattiness popularity nbhourspast level
amplify.sh dom procrd.co 10 10 120 2
```

Carding/dump shops: carder007.org, carder.site, cardx.biz, mastercvv.in, trump-dumps.ru
Crimeware forums: fuckav.ru
Jabber/chat servers used by criminals: jabber.ru, blah.im
Anonymous, vpn, proxy, socks: doublevpn.com, hidemevpn.de, vpmmonster.ru, hidevpn.me
Stolen accounts, shell, RDP: dedicrdp.ru

Criminal hosting space

Introduced at Black Hat 2014, Botconf 2014, Defcon 2017

Zbot Fast Flux BPH operation


Victim
Crimeware consumer
Researcher
Zbot Fast Flux Proxy Network
Aka Fluxxy, Darkcloud
Criminal customer’s site origin IP

Actor(s) grow and maintain FF network
*FF service offered in underground forums

40-50 bot IPs provisioned per domain

Botnet comprised of 30-40K compromised residential IPs, mainly in UA, RU

Content delivered
Short lifetime: malware, ransomware
Medium lifetime: phishing
Long lifetime: carding, cybercrime forums

SSL Data mining

SSL

Goals

- What questions will we ask?
  - How can we connect domain, IP data with SSL

- IPv4 SSL scans: 2TB of data - a few million IPs and a few million SHAs

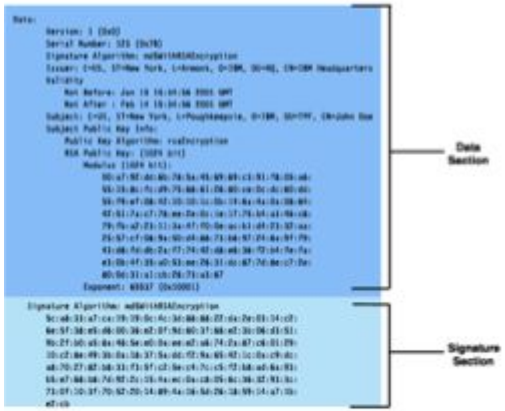
SSL Backend Architecture - Data Format

- Two separate data formats
  - x509 certificate
  - IP ← → SSL SHA mappings

- Require different forms of indexing
  - Document store
  - Key/Value store

- Communication between data stores

Certificate Example



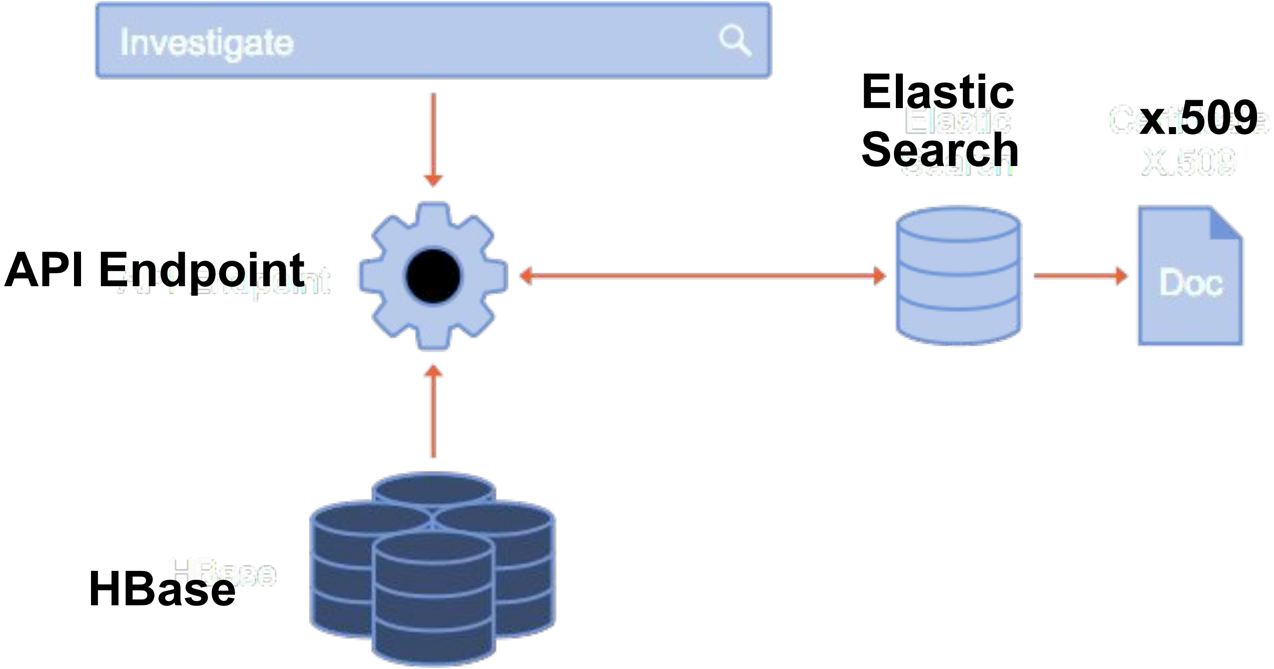
```
Data:
        Version: 1 (0x0)
        Serial Number: 123 (0x7B)
        Signature Algorithm: md5WithRSAEncryption
        Issuer: C=US, ST=New York, L=Armonk, O=IBM, OU=RQ, CN=IBM Headquarters
        Validity
            Not Before: Jun 18 16:05:56 2001 GMT
            Not After : Feb 14 15:05:56 2003 GMT
        Subject: C=US, ST=New York, L=Poughkeepsie, O=IBM, OU=TPF, CN=John Doe
        Subject Public Key Info:
            Public Key Algorithm: rsaEncryption
            RSA Public Key: (1024 bit)
                Modulus (1024 bit):
                    [illegible]
                Exponent: 65537 (0x10001)
    Signature Algorithm: md5WithRSAEncryption
        [illegible]
```

SSL Backend Architecture

Data Challenges

- Different data stores for different data types
  - Documents don’t store well in traditional RDBMS
  - X509 certs are sparse documents
- Combination of big data technologies
  - HBASE
  - ElasticSearch

SSL Backend Architecture



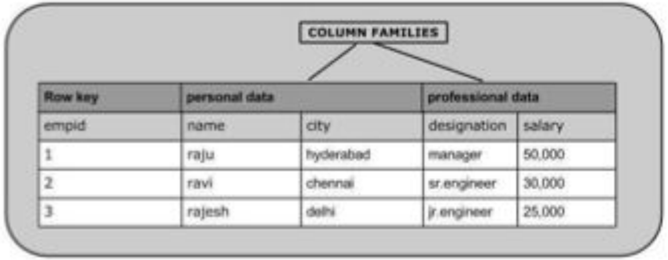
SSL Backend Architecture

HBase and Elastic Search Table Design

- 4 HBase major components:
  - Rowkey
  - Column Family
  - Column Qualifier
  - Cell

- Design of the rowkey most important
  - Use RK to match SHAs to IP spaces

- Elasticsearch stores parsed x509 docs
  - Indexed on fields; e.g. search on CNs


COLUMN FAMILIES

Row key	personal data		professional data	
empid	name	city	designation	salary
1	raju	hyderabad	manager	50,000
2	ravi	chennai	sr.engineer	30,000
3	rajesh	delhi	jr.engineer	25,000

Reveal origin IP of domains hiding behind reverse proxies

- darkmoney.cc, a cybercrime forum hides behind Cloudflare

```
darkmoney.cc.    299    IN    A    104.31.1.166
darkmoney.cc.    299    IN    A    104.31.0.166
```

- Search CN=darkmoney.cc in SSL data base

```
33b64e11a6e8529d9b719bf9e91bf8b9fd0ad6fa,darkmoney.cc
```

- Search the sha in SSL data base

```
33b64e11a6e8529d9b719bf9e91bf8b9fd0ad6fa 2016-06-27 181.174.164.101
```

- Confirm content is hosted on the hidden IP

```
curl --header 'Host: darkmoney.cc' http://181.174.164.101
```

Conclusion

Dealing with large scale threat intel problems, you need to:

- Know your requirements: what are you looking for?
- Know what to collect
- Know how to store and process the data at scale
- Know what analysis to apply: human or machine based at scale or a combination
- What is your final product: discrete IOCs, or trends and TTPs

Some of our related work

- Hack in the Box 2018 https://youtu.be/co2cvi_5FIc
- SANS CTI Summit 2018 https://www.sans.org/summit-archives/file/summit-archive-1517343456.pdf
- Flocon 2018 https://schd.ws/hosted_files/flocon2018/d7/2.%20FloCon%202018_.pdf
- https://schd.ws/hosted_files/flocon2018/16/2.%20Flocon_2018_Thomas_Dhia_Jan_10.pdf
- Virus Bulletin 2017 https://www.youtube.com/watch?v=sbzvZ8ChTiU
- Defcon 2017 https://www.youtube.com/watch?v=AbJCOVLQbjs
- Black Hat 2017 https://www.youtube.com/watch?v=PGTTRN6Vs-Y&feature=youtu.be
- Usenix Enigma 2017 https://www.youtube.com/watch?v=ep2gHQgjYTs&t=818s
- Black Hat 2016 https://www.youtube.com/watch?v=m9yqnwuqdSk
- RSA 2016 https://www.rsaconference.com/events/us16/agenda/sessions/2336/using-large-scale-data-to-provide-attacker
- BruCon 2015 https://www.youtube.com/watch?v=8edBgoHXnwg
- Virus Bulletin 2014 https://www.virusbtn.com/conference/vb2014/abstracts/Mahjoub.xml
- Black Hat 2014 https://www.youtube.com/watch?v=UG4ZUaWDXSs

Thank you

dhia@opendns.com

@DhiaLite