RISKIQ®

Internet Snapshots

Utilizing web crawling and internet scanning to connect infrastructure

Stephen Ginty

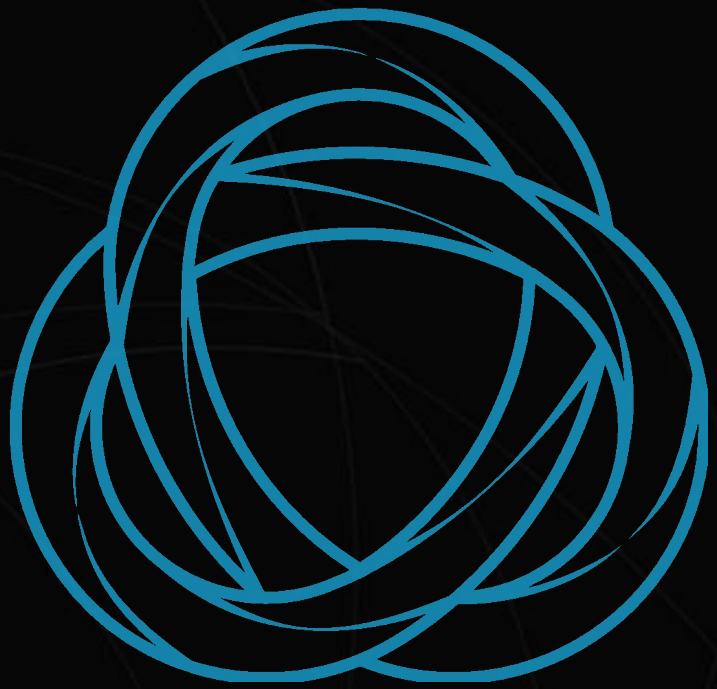
Who Am I?

- Co-founder of PassiveTotal (acquired by RiskIQ September 2015)
- Security researcher
- Focused on
  - Improving analyst workflows
  - Making threat research approachable
  - Operationalizing infrastructure data

About RiskIQ

RiskIQ protects the externally-facing digital assets – known and unknown – of any organization from malicious actors.

Our products provide actionable and timely pictures of your attack surface and attackers' infrastructure to proactively defend against threats.



Agenda

- Attack impressions
- Research methodologies
- Common datasets
- Derived datasets
- Wrap-up

Attack Impressions

No Connection, No Attack

"Security Protip: Never connect to the Internet and you'll be fine"

- Actors must use the Internet to mount an attack online
- Signals are everywhere
  - Connection locations
  - Service providers
  - Obscure techniques
- By not leaving signals, you leave signals

Mapping the Signals

- Chrome Extensions deployed
- Email addresses used to register extensions
- Py2Exe tool for packaging
- Use of specific affiliate providers
- Track infection progress/engagement with specific providers
- Social media landing page for attacks
- Provocative images used for lures
- Copy used inside of posts to friends
- Custom Javascript encoders
- Javascript functions to interface with social media
- Executables compiled with python
- Use of specific hosting providers
- Use of specific registrars
- Email addresses used for registration of domains
- 3rd-party hosting of infrastructure and infection pages
- Specific apps used for affiliate mining
- Metadata from the images used in attacks
- Hard-coded account IDs for payment processing
- Outbound IP addresses from local ISP
- Social media accounts used for communications
- Frequency of communications
- Fuzzing of social media controls
- .Net wrappers used for payloads
- Specific .net encoders used
- Comments used from landing pages on attacks
- Heavy use of redirection sequences
- Specific shortlink providers used for tracking
- Metadata from shortlink providers outlining statistics
- Operational tempo and campaign release times
- Willingness to change and adjust when caught
- Use of Chromium browser deployed on victim machines
- Open directories on command and control servers
- Reuse of server infrastructure
- Purchase of obscure TLDs not often used
- Timing of new infrastructure deployments
- Specific versions of python used
- Redirection order based on service type
- Infrastructure used to access command and control
- Database names used for storage
- Profiling information on clients and their software
- Specific use of exploits or lack thereof
- Pivot between old methods and then new methods
- Language use inside of code
- Self-signed certificates for command and control

Fewer Places to Hide

- Technology allow us to easily collect from the Internet and save the contents for later processing
- Leverage highly-connected data to capture mistakes, find poor OPSEC, and connect infrastructure

Research Methodologies

Infrastructure Analysis

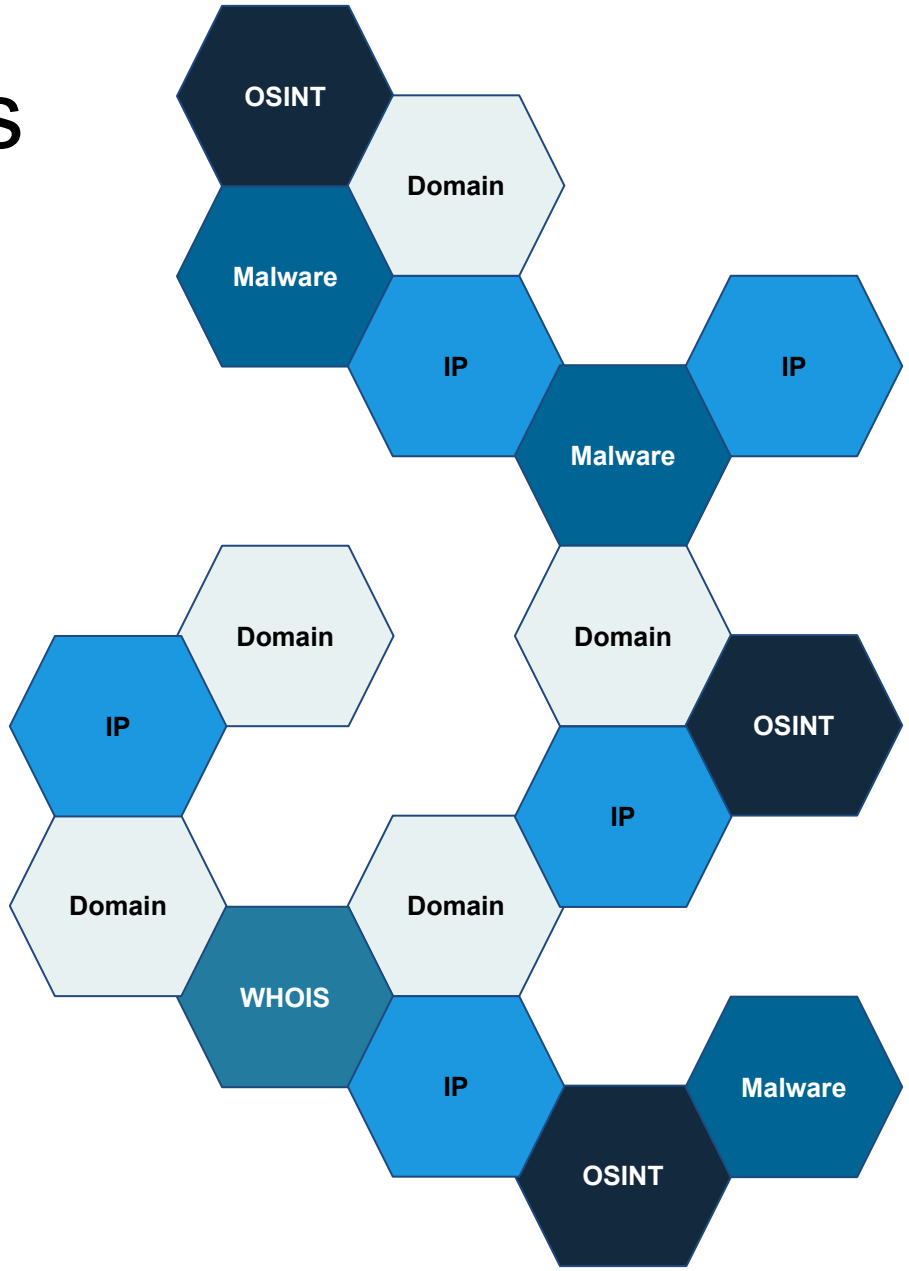
"Research process that brings context to incidents through related entities"

- Why is this successful?
  - –Bad actors can't avoid interacting with core components of the internet
  - –Chaining together multiple datasets to create a larger narrative
  - –Surface new investigative leads

Infrastructure Chains

Leverage the relationships between highly-connected datasets to build out an investigation

- Surface new connections
- Group similar activity
- Substantiate assumptions



Common Datasets

Common Datasets

Passive DNS

Historical set DNS of records for domains and IP addresses. Reveals patterns of attackers or derives timelines.

WHOIS

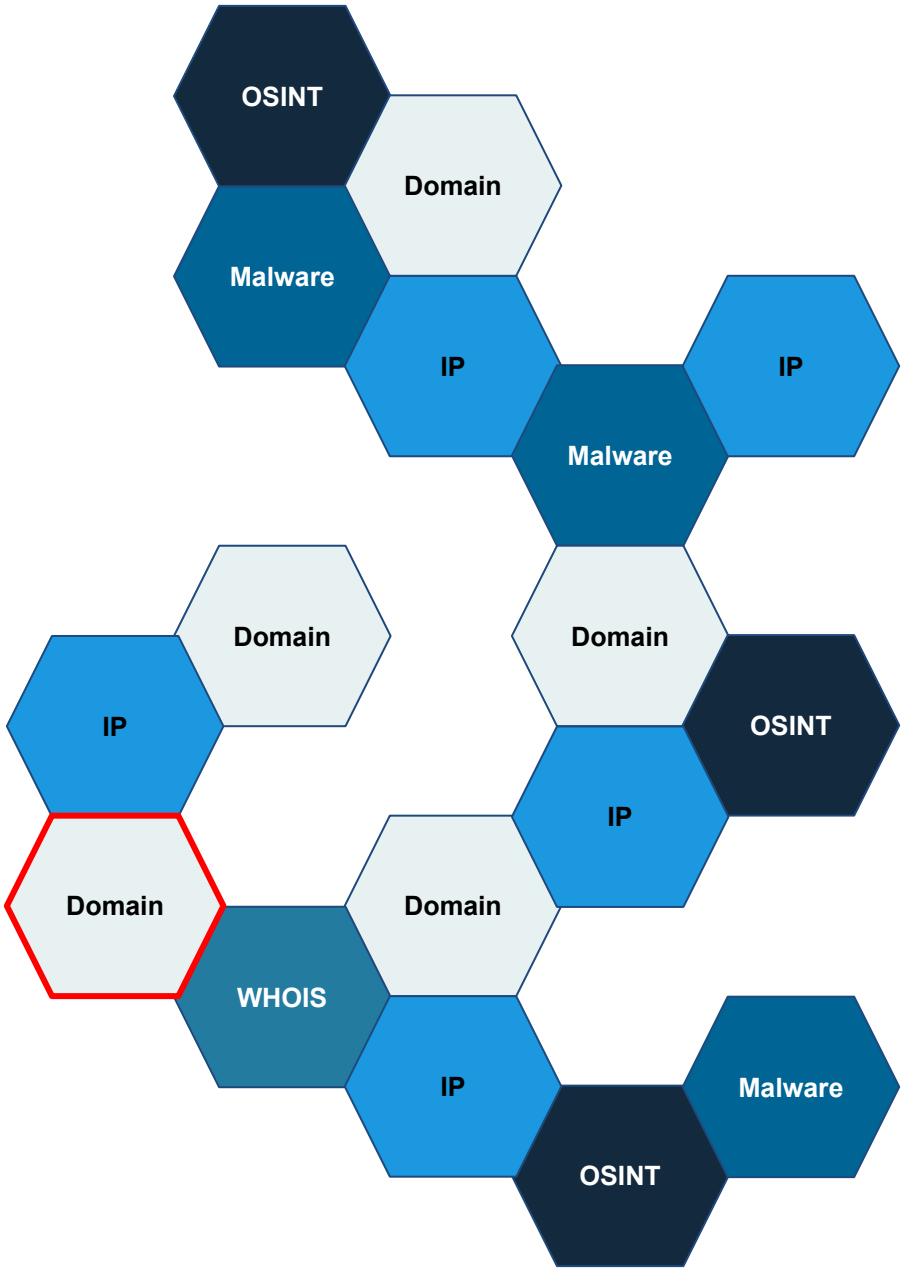
Repository of registrant information that can provide leads on connecting different data points together. Sometimes reveals actor patterns.

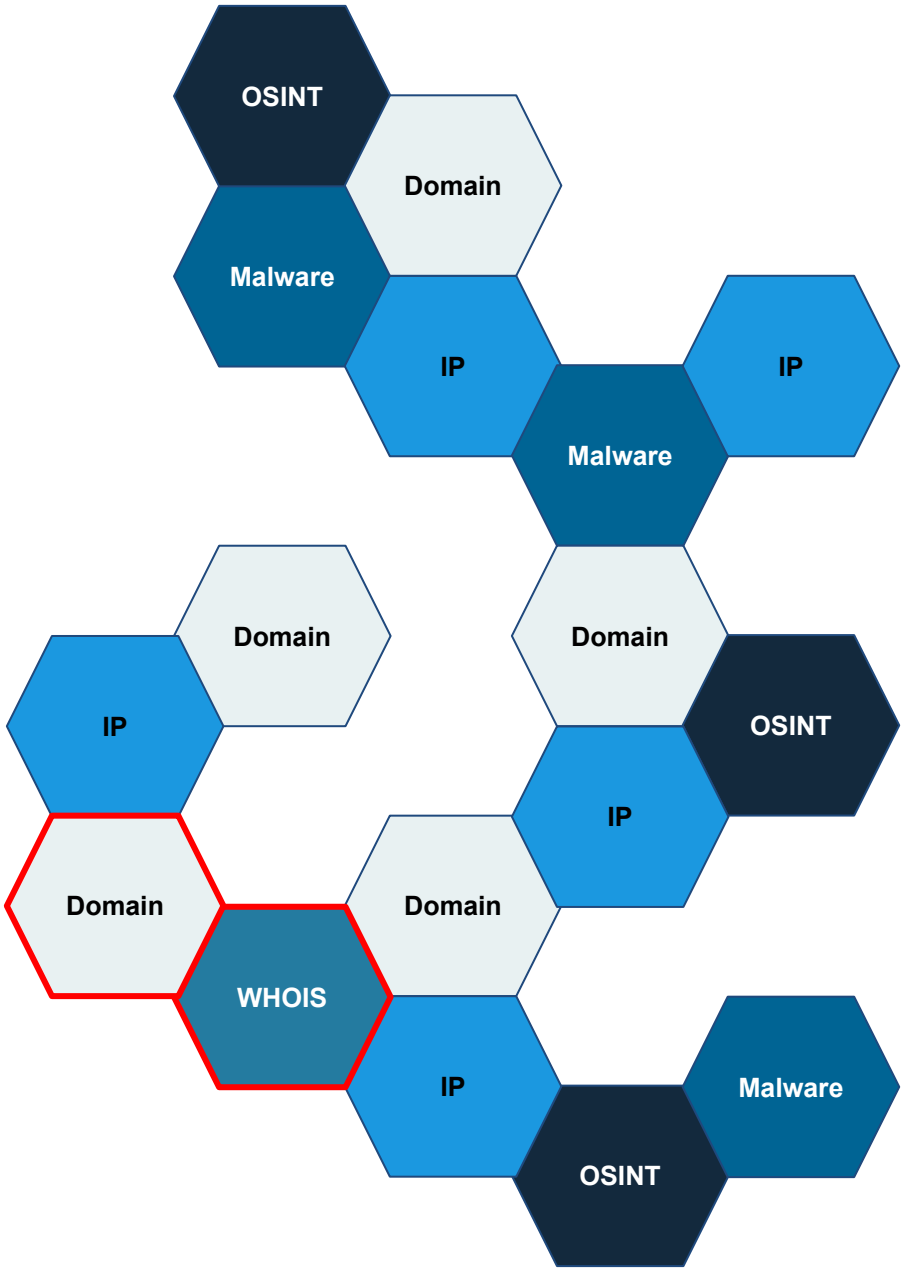
OSINT

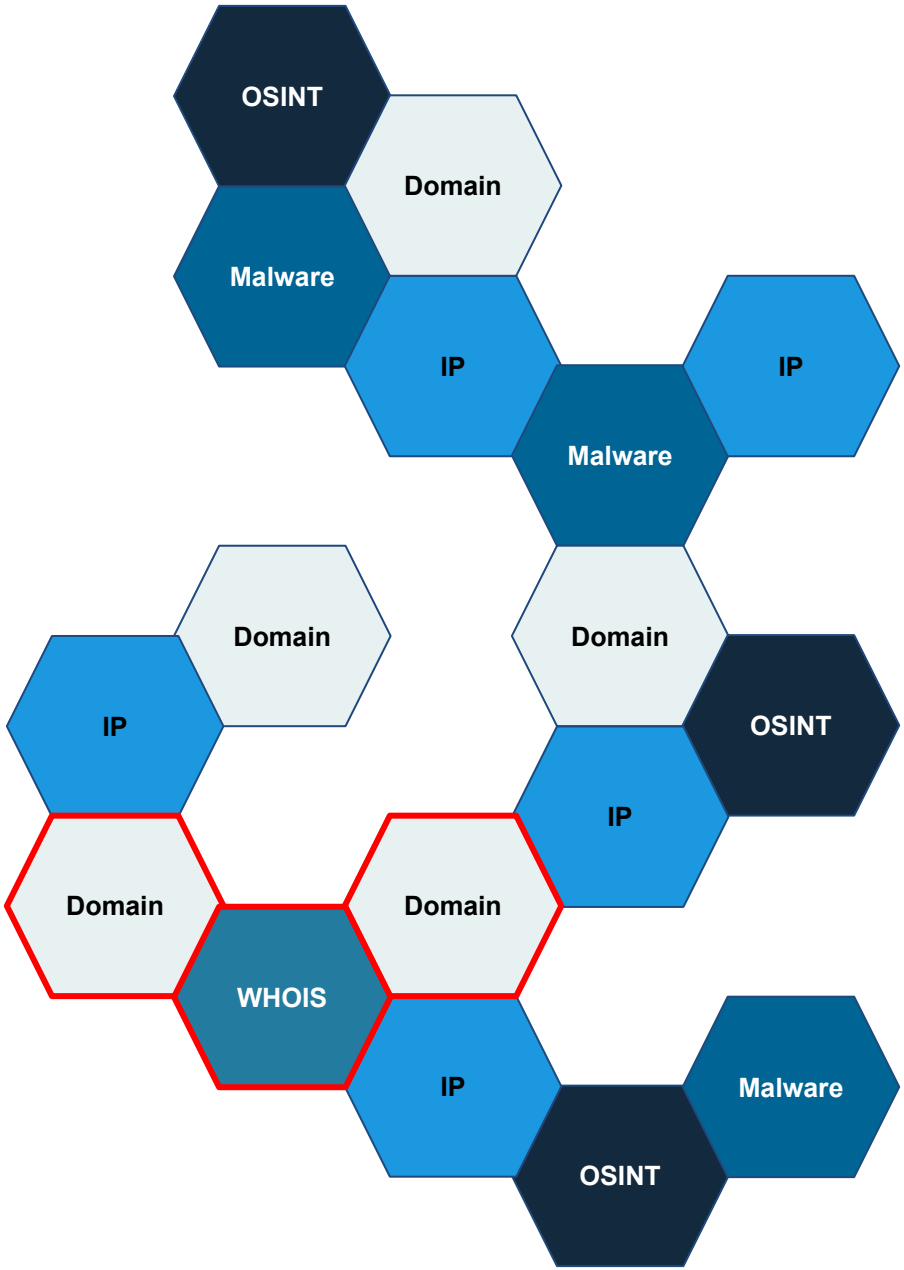
Open source intelligence, both long and short form developed by individuals and companies. Provides context to the actors, campaign or malicious infrastructure.

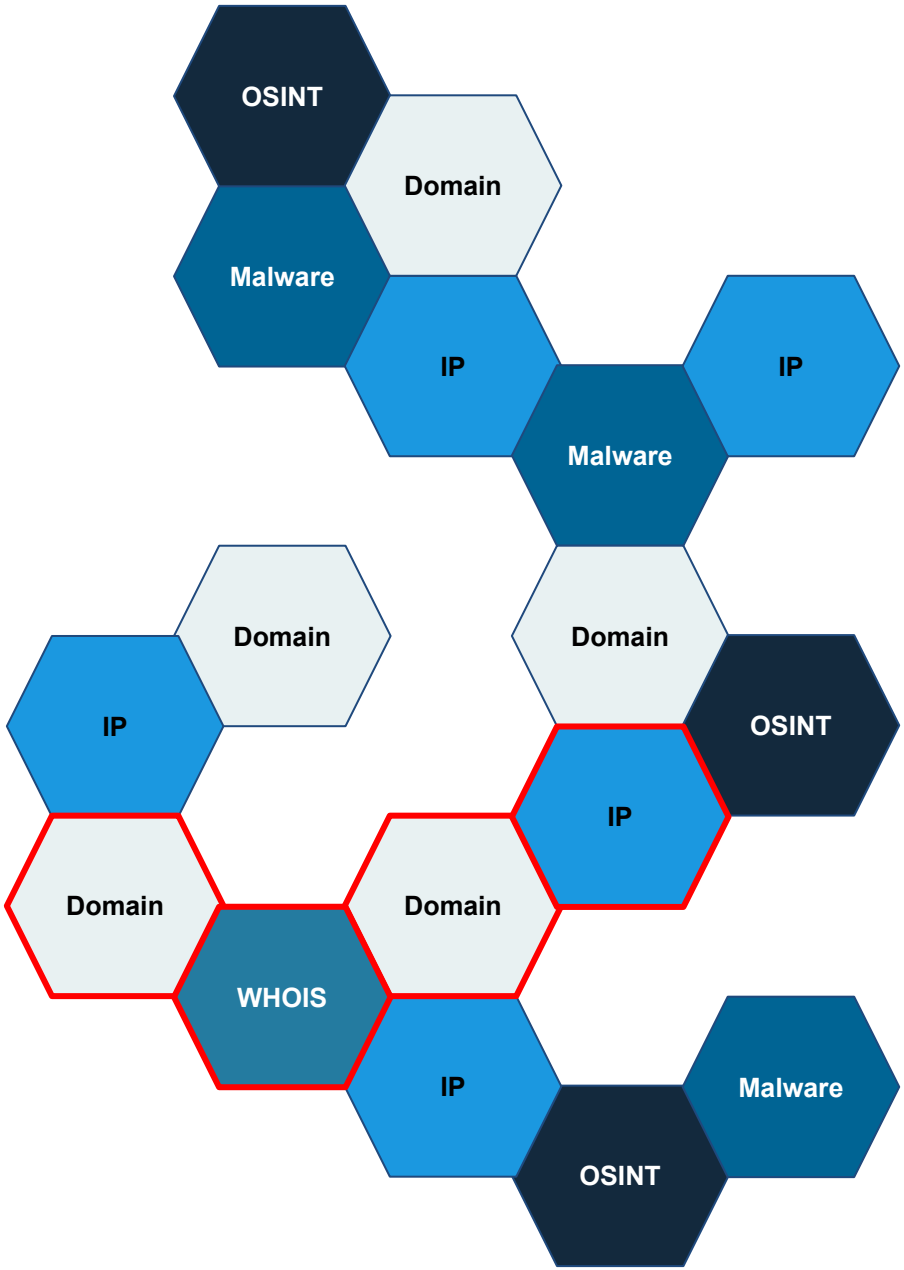
Malware

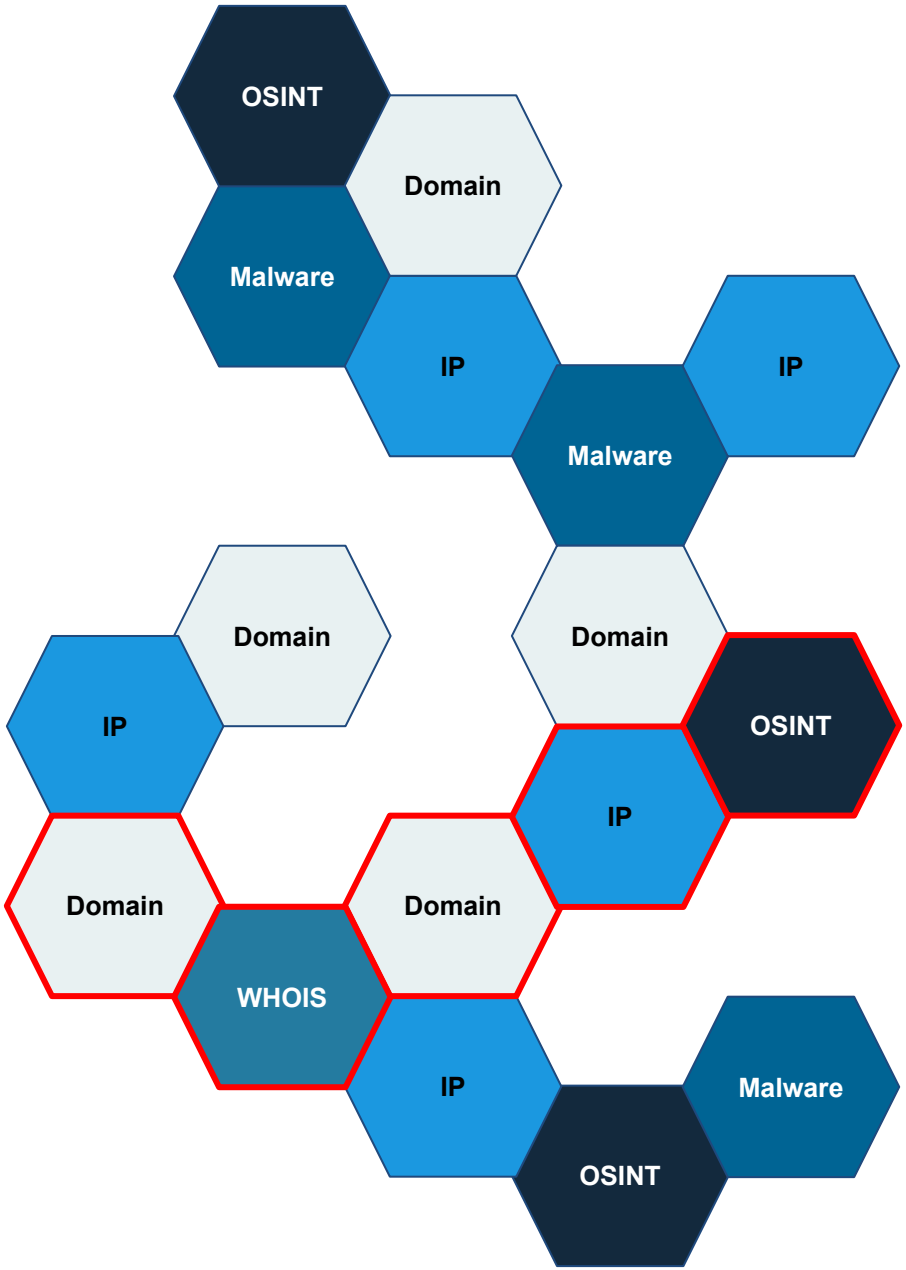
Malicious software being used in attacks or found on the Internet. Outlines capabilities, intent and motives of an attacker. Aids in connecting back to infrastructure.


OSINT
Domain
Malware
IP
IP
Malware
Domain
Domain
IP
OSINT
IP
Domain
Domain
WHOIS
IP
Malware
OSINT


OSINT
Domain
Malware
IP
IP
Malware
Domain
Domain
IP
OSINT
IP
Domain
Domain
WHOIS
IP
Malware
OSINT


OSINT
Domain
Malware
IP
IP
Malware
Domain
Domain
IP
OSINT
IP
Domain
Domain
WHOIS
IP
Malware
OSINT


OSINT
Domain
Malware
IP
IP
Malware
Domain
Domain
IP
OSINT
IP
Domain
Domain
WHOIS
IP
Malware
OSINT


OSINT
Domain
Malware
IP
IP
Malware
Domain
Domain
IP
OSINT
IP
Domain
Domain
WHOIS
IP
Malware
OSINT

Actor TTPs Evolving

- Since 2011, hundreds of blogs, reports and media articles have been published
- Targeting is becoming more precise based on operational focus
- False flags are cropping up amongst better actor groups
- Blurred lines between criminal and nation-state actors

Passive DNS Shortcomings

- DNS as a service (Dynamic DNS)
  - Obfuscates the owner of the infrastructure
- Content delivery networks (CDNs)
  - Obfuscate the actual hosting provider of the actor
  - Additional party to coordinate with
- 3rd-party command and control
  - Free services are dead leads and reveal nothing
- Dead drop command and control
  - Requires expertise to uncover

WHOIS Shortcomings

- Privacy protection services
  - Being offered by default or low-cost
- High opportunity for false flags
  - Take indicators from existing reporting
- Lack of coordination with registrars
  - Obtaining records in bulk is difficult

OSINT Shortcomings

- Reveals methodologies of both defenders and actors
  - Loss of visibility into operations
  - Infrastructure is abandoned
  - Techniques change or become better
- Reporting bias or inaccuracies
  - Unclear how connections are made
  - Data sources are not revealed
  - Mistakes often go uncorrected

Malware Shortcomings

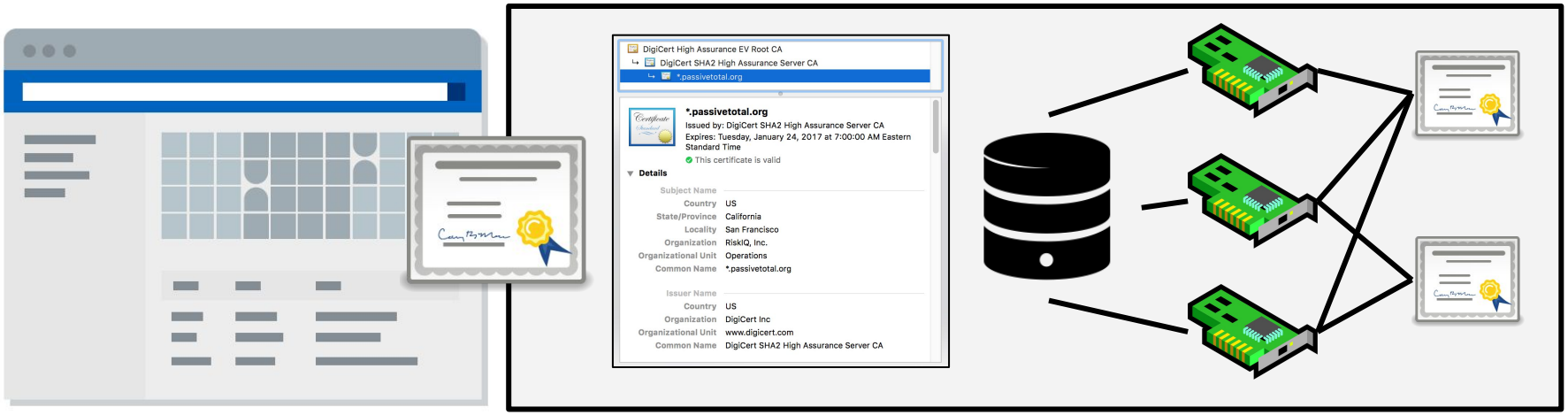
- Often lacks context to the larger attack
  - Not always a clear signal of sophistication
- Reuse across different actors
  - Shared supply chains
- Numerous different providers
  - Open source code
  - Commercial providers
  - Custom implants

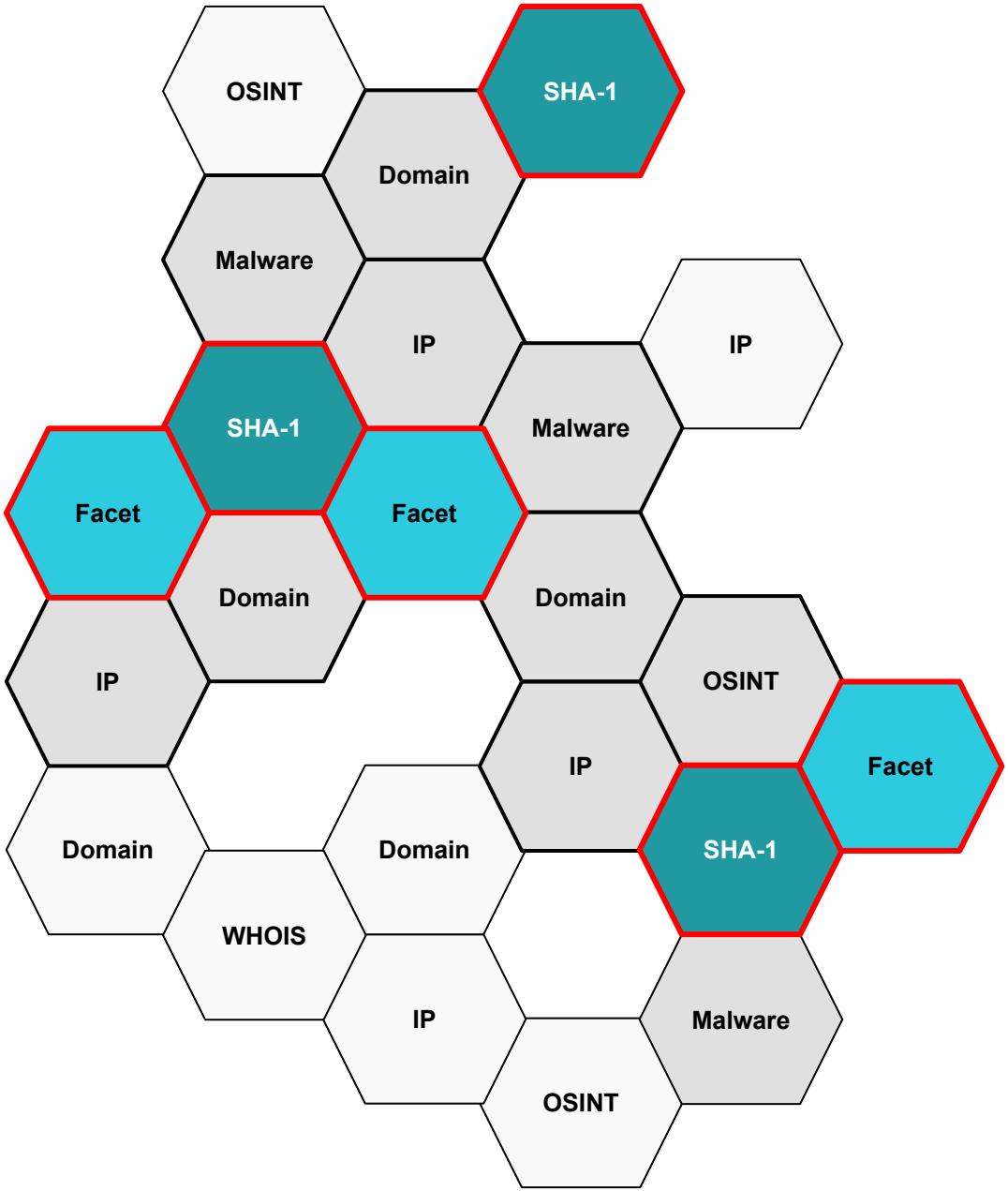
Derived Datasets

SSL Certificates

Digital certificates observed from Internet scanning of IPv4 addresses

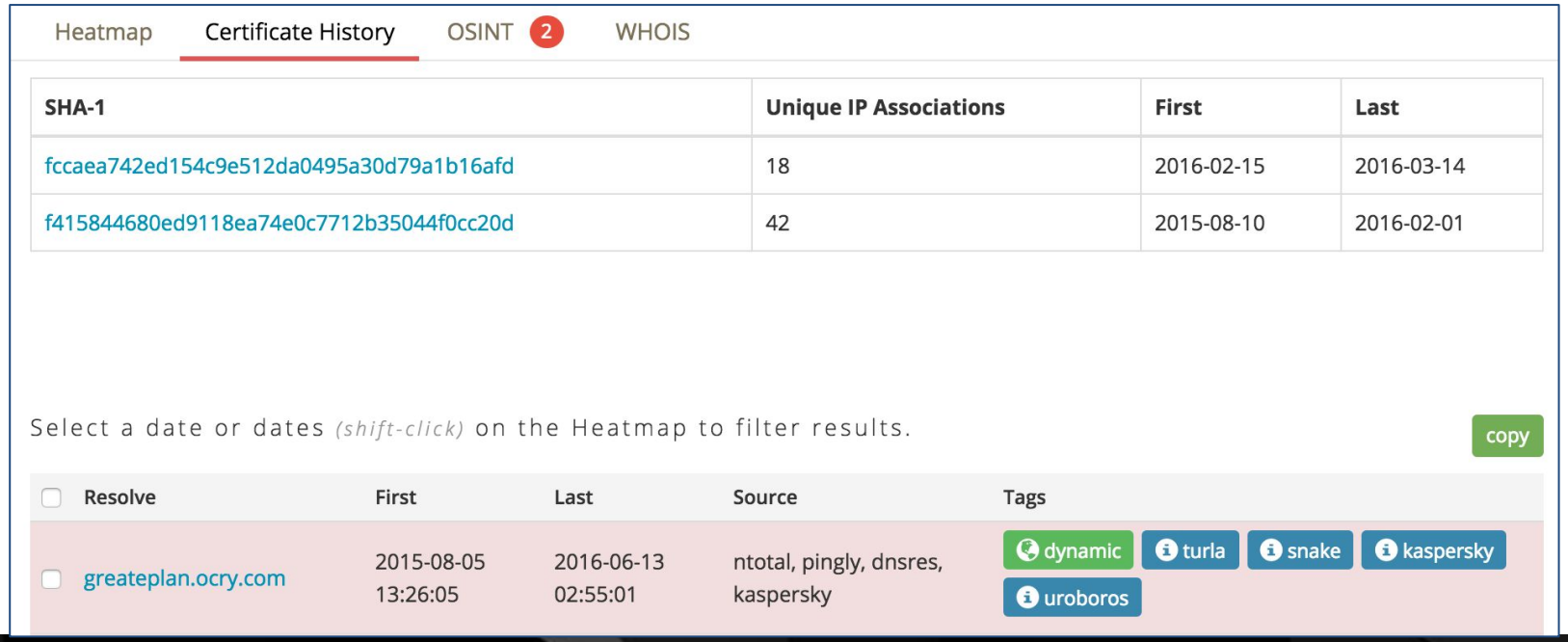
- Overlapping certificate details (facets)
- SHA-1 hash of the certificate
- IP address of the host

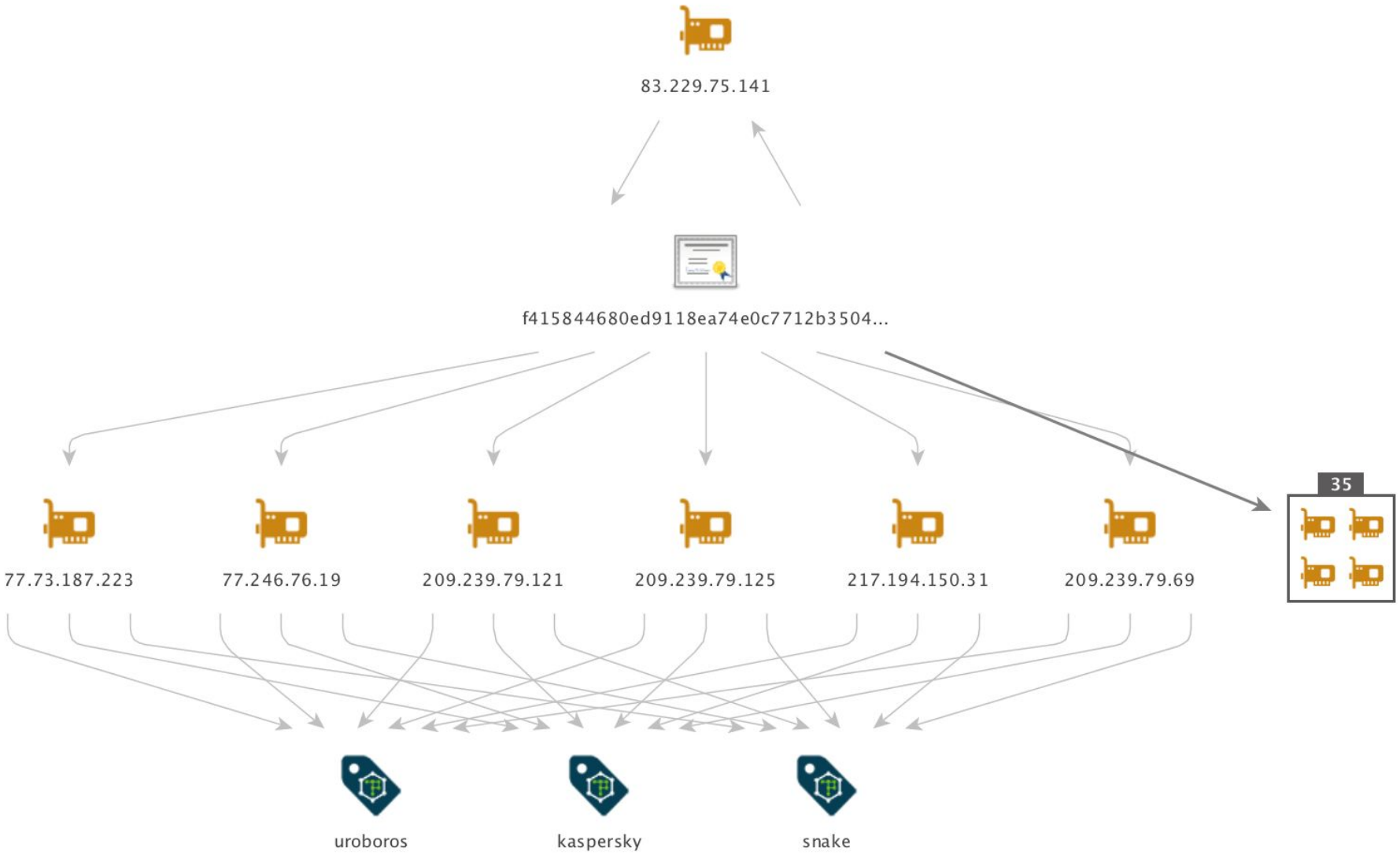


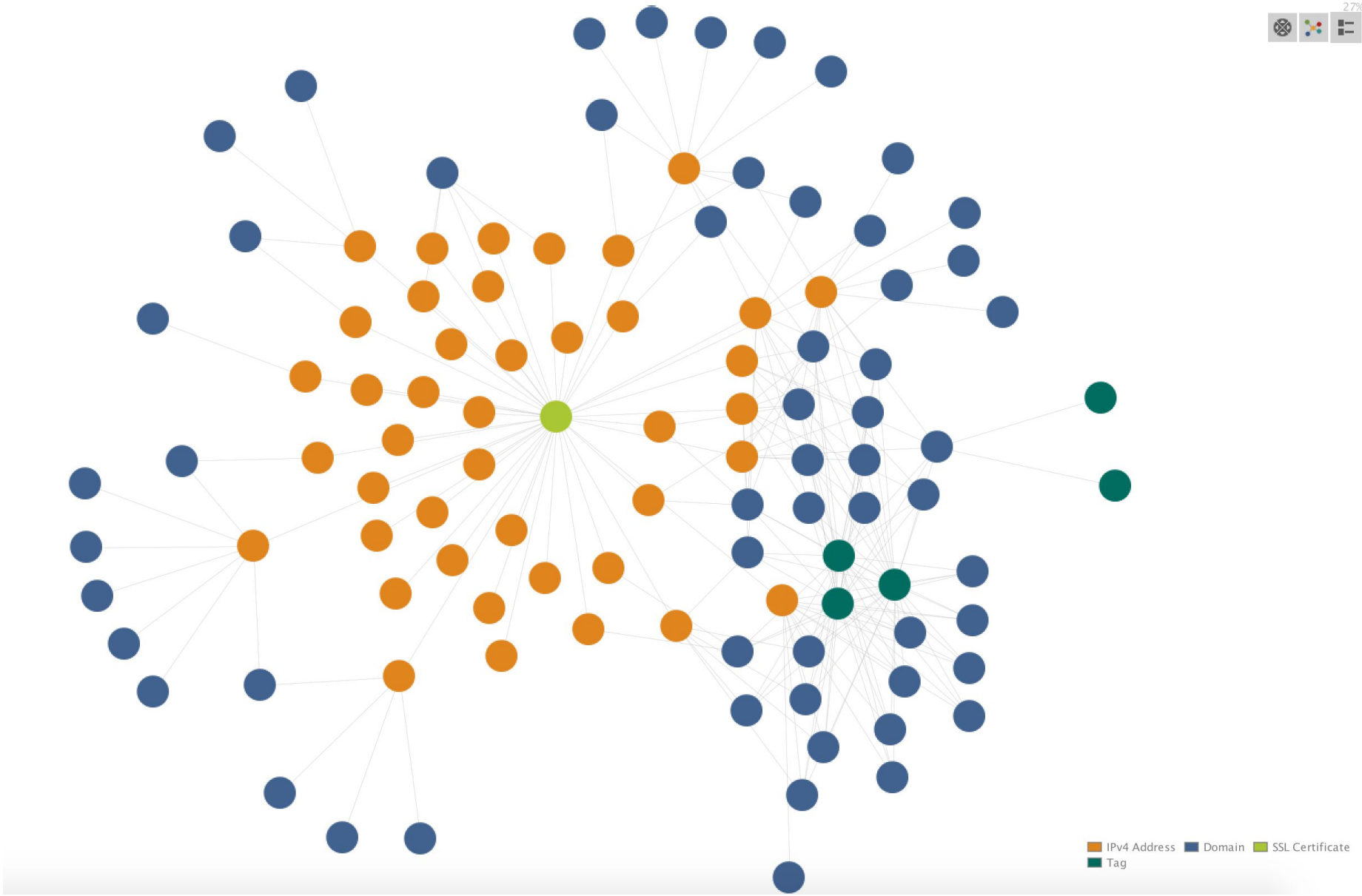

OSINT
SHA-1
Domain
Malware
IP
IP
SHA-1
Malware
Facet
Facet
Domain
Domain
IP
OSINT
IP
Facet
Domain
Domain
SHA-1
WHOIS
IP
Malware
OSINT

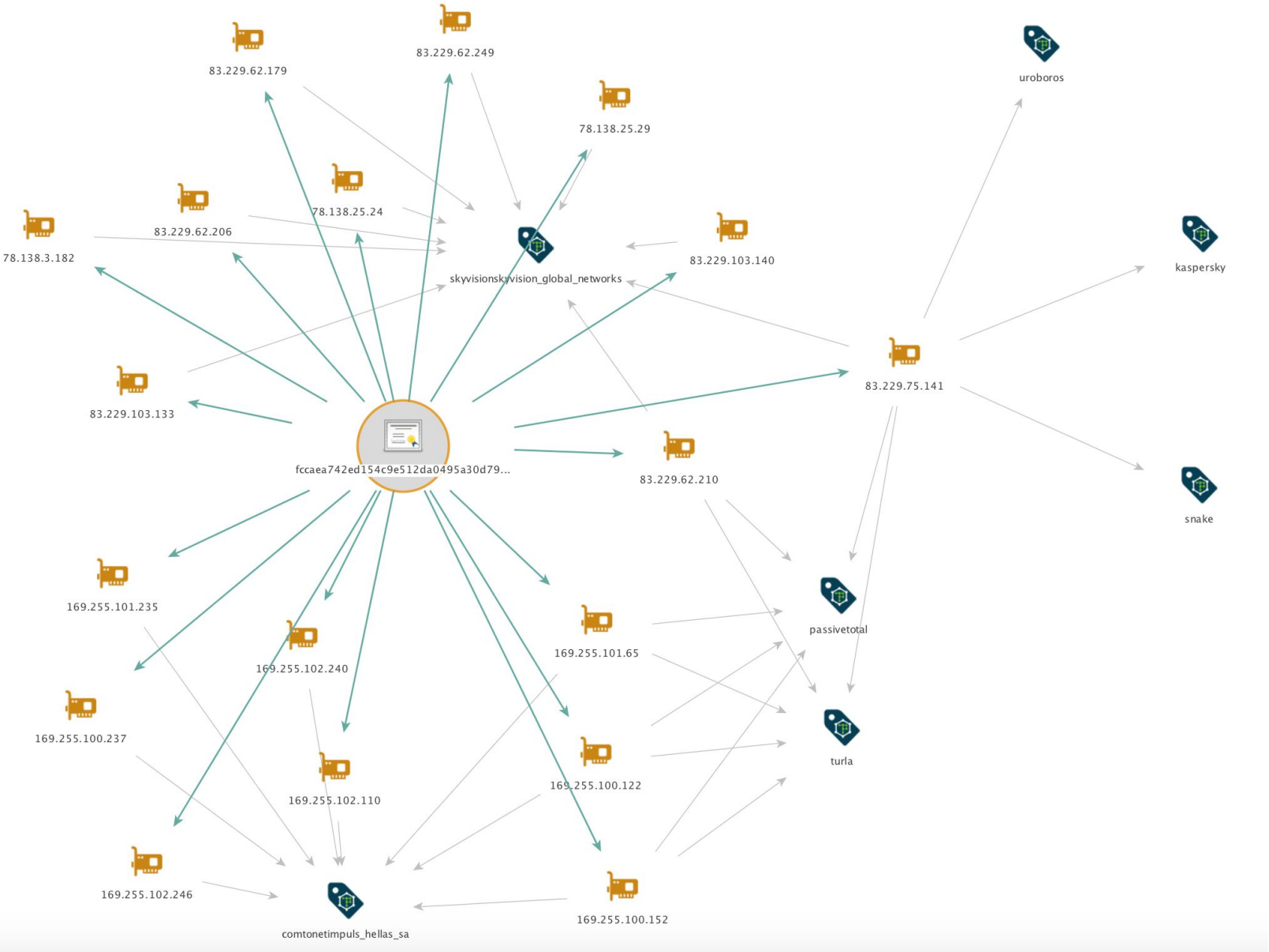
Example: Turla Infrastructure

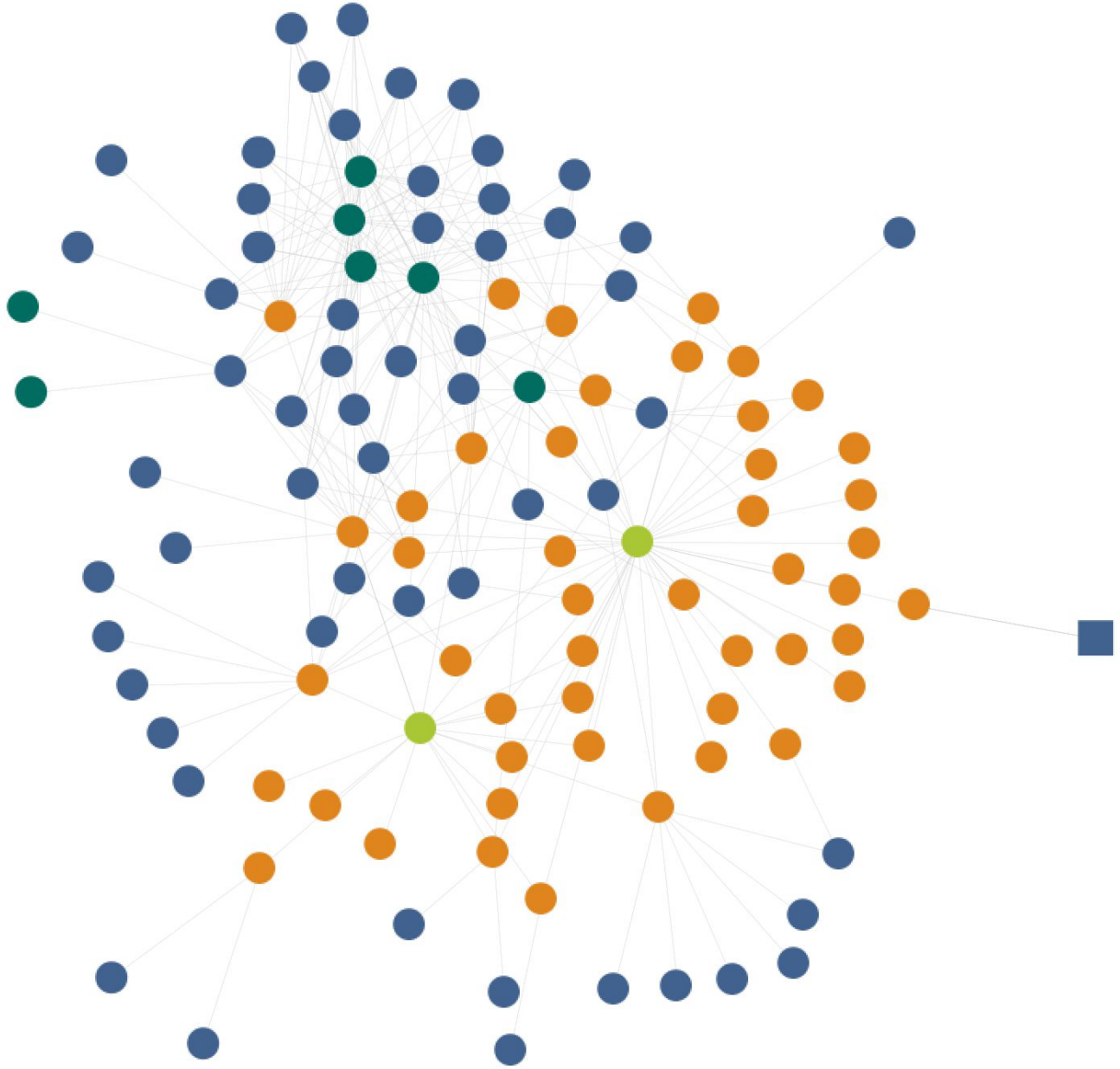
- Connections via SHA-1 Hash
  - 42 IP addresses - 16 new
  - 36 Dynamic DNS domains - 15 new
  - 4 additional satellite providers
  - Active infrastructure identified




83.229.75.141
f415844680ed9118ea74e0c7712b3504...
77.73.187.223
77.246.76.19
209.239.79.121
209.239.79.125
217.194.150.31
209.239.79.69
35
uroboros
kaspersky
snake


27%
IPv4 Address
Domain
SSL Certificate
Tag


83.229.62.179
83.229.62.249
uroboros
78.138.25.29
78.138.25.24
83.229.62.206
78.138.3.182
skyvisionskyvision_global_networks
83.229.103.140
kaspersky
83.229.75.141
83.229.103.133
fccaea742ed154c9e512da0495a30d79...
83.229.62.210
snake
169.255.101.235
passivetotal
169.255.101.65
169.255.102.240
169.255.100.237
turla
169.255.100.122
169.255.102.110
169.255.102.246
169.255.100.152
comtonetimpuls_hellas_sa



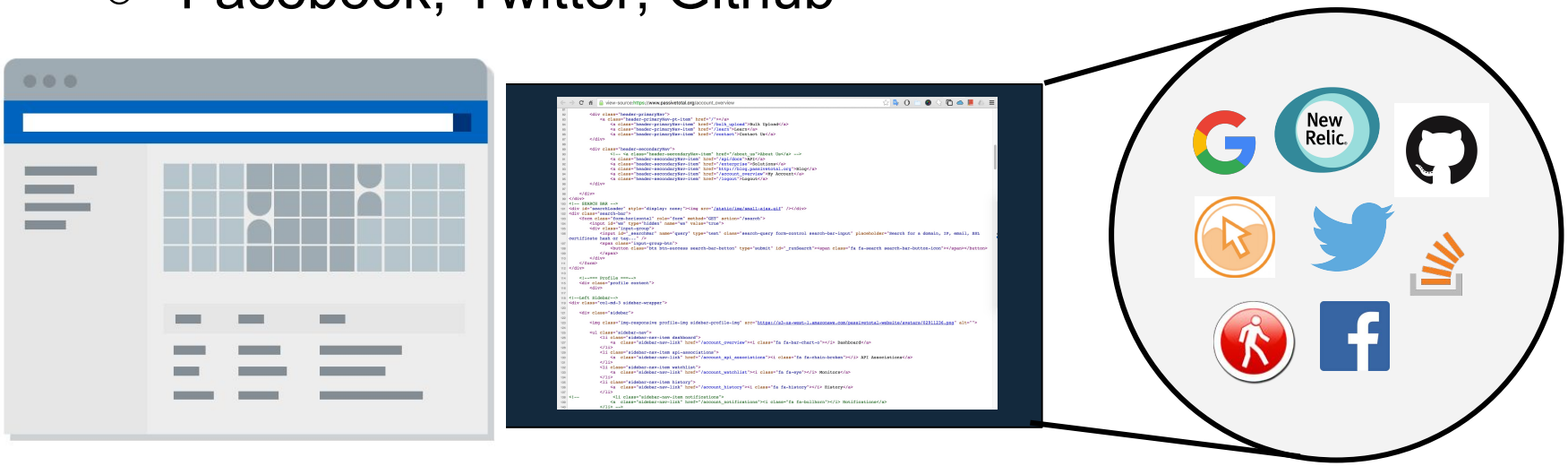
Caveats and Considerations

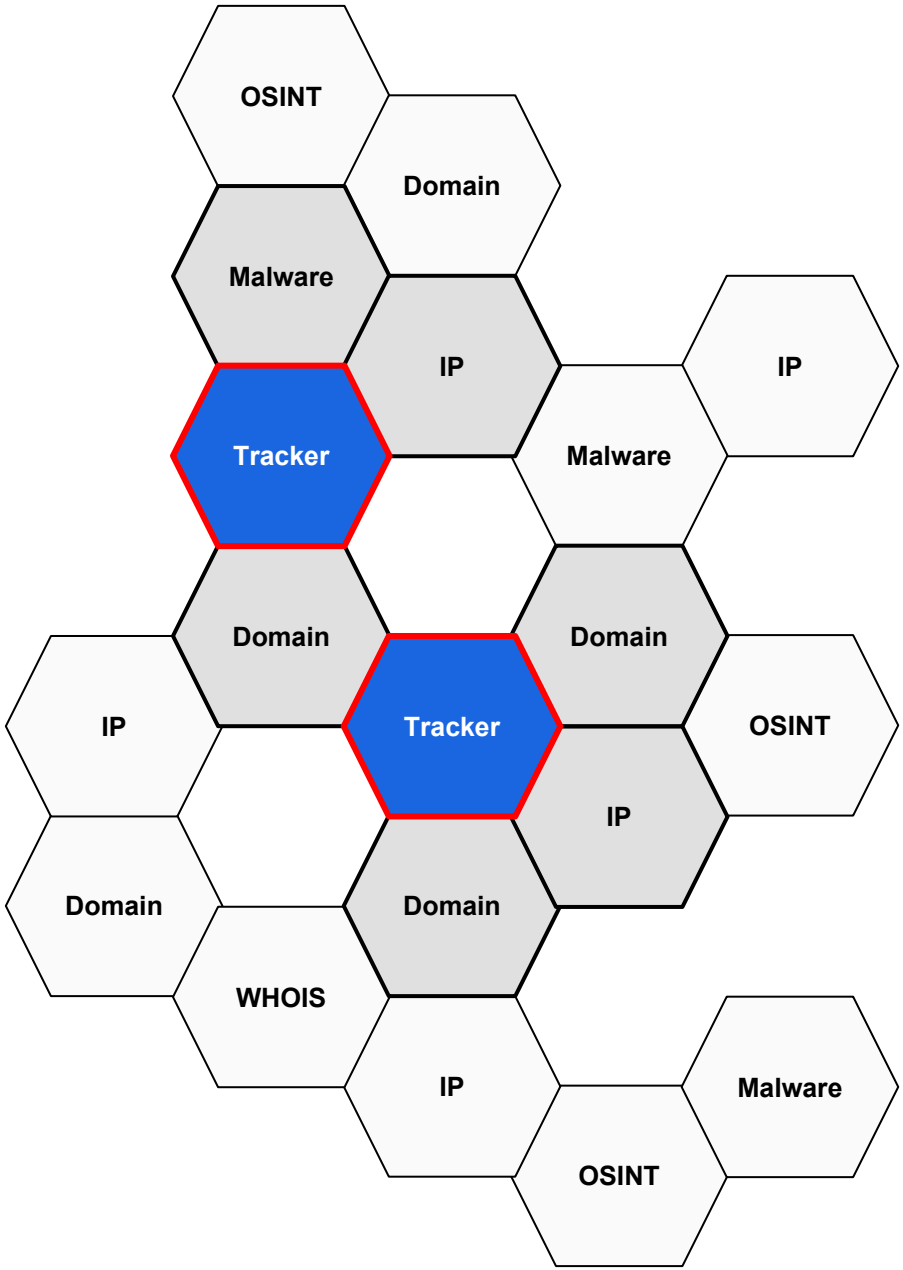
- Certificate validity (date, self-signed)
- Overlap on facet or IP addresses
  - Frequent facets (country, issuer, etc.)
  - Shared certificate usage for companies
- Hosting providers
  - Being used by a CDN?
  - Is the certificate deployed with web services?
- Where was the certificate used?

Tracking Codes

Analytics as a service that leave unique fingerprints in web page content

- Affiliate providers
  - Google, Yandex, Clicky
- Social media
  - Facebook, Twitter, Github



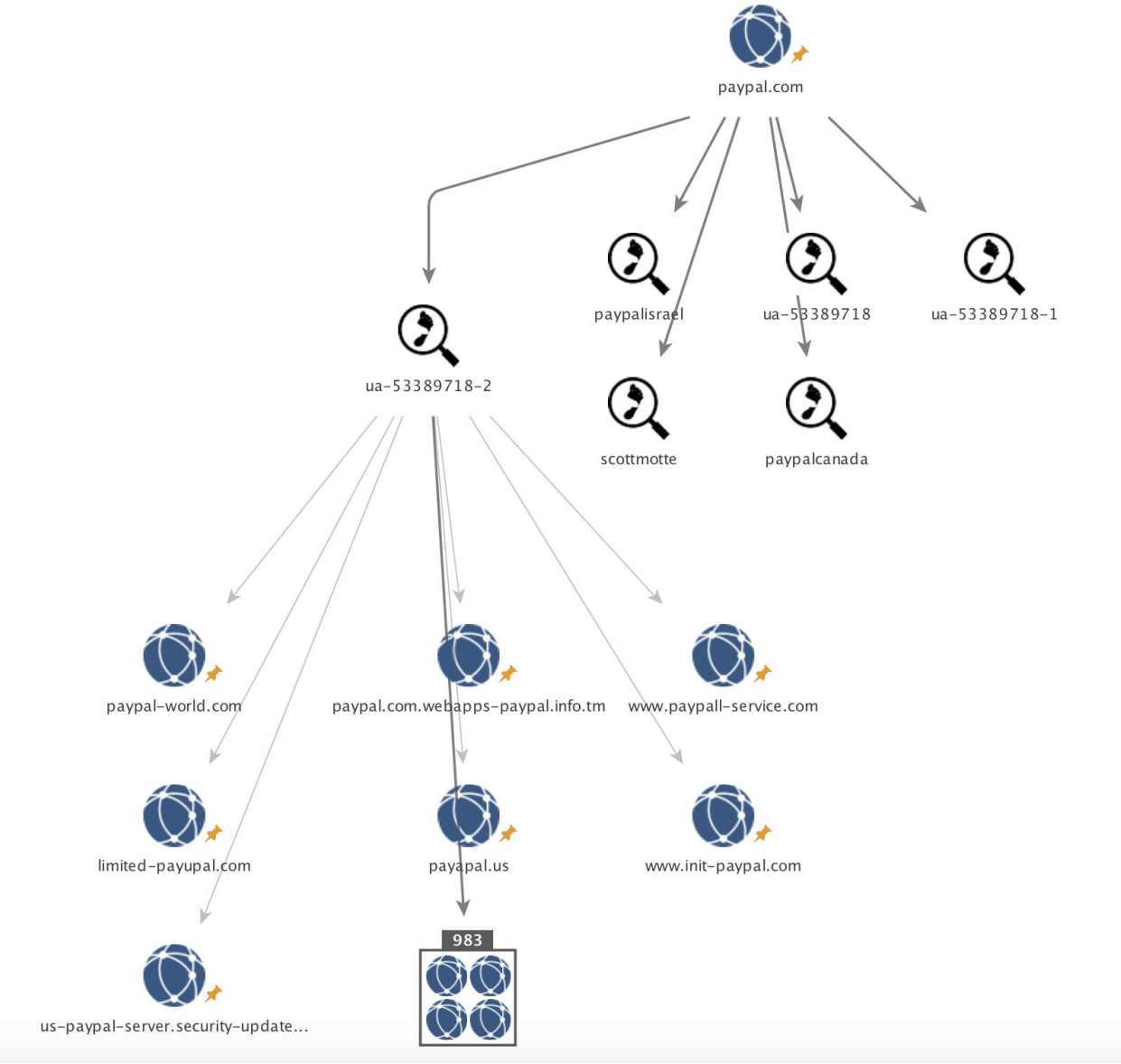

OSINT
Domain
Malware
IP
IP
Tracker
Malware
Domain
Domain
IP
Tracker
OSINT
IP
Domain
Domain
WHOIS
IP
Malware
OSINT

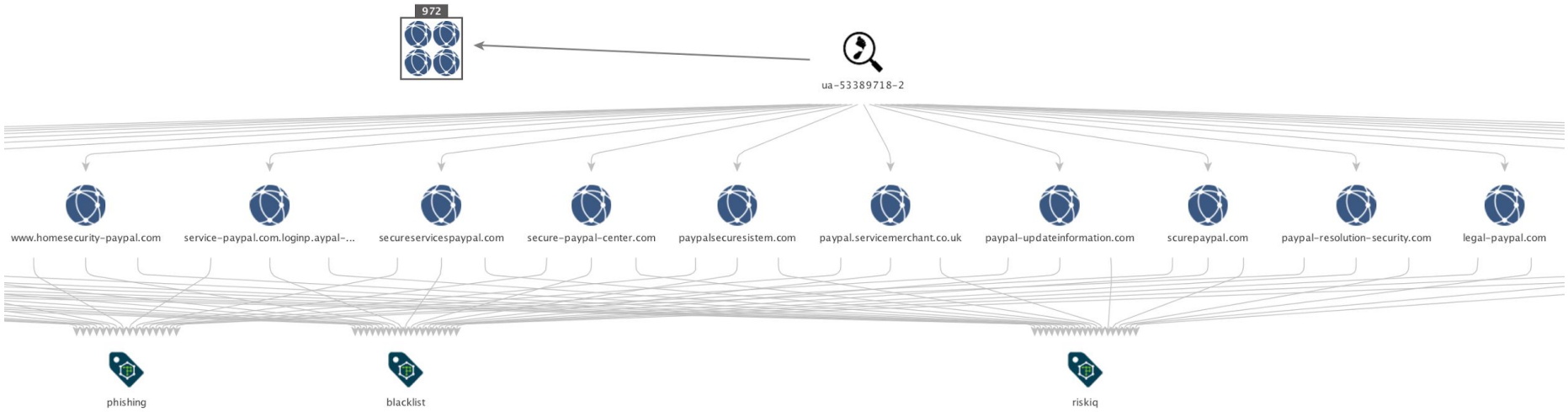
Example: Paypal Phishing

- 37 different trackers
  - Social, Marketing, Analytics
- Google Analytics ID (UA-53389718-2)
  - Linked to hundreds of domains
  - Several blacklisted/phishing pages

Hostname	First Seen	Last Seen	Type	Value
demo.paypal.com	2015-09-28 01:01:40	2016-04-23 17:51:27	GoogleAnalyticsTrackingId	ua-49901229-1
devblog.paypal.com	2016-01-26 15:02:47	2016-04-17 17:05:22	TwitterId	cbetta'
devblog.paypal.com	2016-04-03 13:44:47	2016-04-17 16:48:24	TwitterId	igaganm
devblog.paypal.com	2016-01-29 00:25:50	2016-04-17 17:00:20	TwitterId	joannaschwartz'
devblog.paypal.com	2016-01-06 22:32:04	2016-04-15 00:52:50	TwitterId	piyush252009
devblog.paypal.com	2016-01-01 10:09:23	2016-04-12 08:34:43	TwitterId	scottmotte
developer.paypal.com	2016-01-04 10:18:06	2016-01-04 10:18:06	GitHubId	2830403.js
developer.paypal.com	2016-01-04 08:41:38	2016-01-04 10:18:06	GitHubId	assets
developer.paypal.com	2013-12-15 19:48:14	2016-04-25 03:51:14	GitHubId	paypal
developer.paypal.com	2014-01-20 01:18:51	2015-07-13 02:10:56	GoogleAnalyticsTrackingId	ua-37419067-1
merchant.paypal.com	2012-08-24 19:17:13	2013-10-17 23:21:48	TwitterId	paypalca


paypal.com
paypalsinfronteras
paypalbrasil
paypal
paypalisrael
5uckerpunchpt
sjtrees
2830403.js
assets
paypal
ua-69786029
ua-53389718-2
ua-53389718-1
ua-37419067-1
ua-49901229-1
paypal
6b79c1d412
paypaldev'
peerbackers
paypal_br
paypalde
askpaypal
joannpeach
orkns
piyush252009
scottmotte
paypal
paypalca
igaganm
joannaschwartz'
cbetta'
paypalde
paypal
paypalcanada


paypal.com
paypalisrael
ua-53389718
ua-53389718-1
ua-53389718-2
scottmotte
paypalcanada
paypal-world.com
paypal.com.webapps-paypal.info.tm
www.paypall-service.com
limited-payupal.com
payapal.us
www.init-paypal.com
983
us-paypal-server.security-update...


972
ua-53389718-2
www.homesecurity-paypal.com
service-paypal.com.loginp.aypal-...
secureservicespaypal.com
secure-paypal-center.com
paypalsecuresistem.com
paypal.servicemerchant.co.uk
paypal-updateinformation.com
scurepaypal.com
paypal-resolution-security.com
legal-paypal.com
phishing
blacklist
riskiq

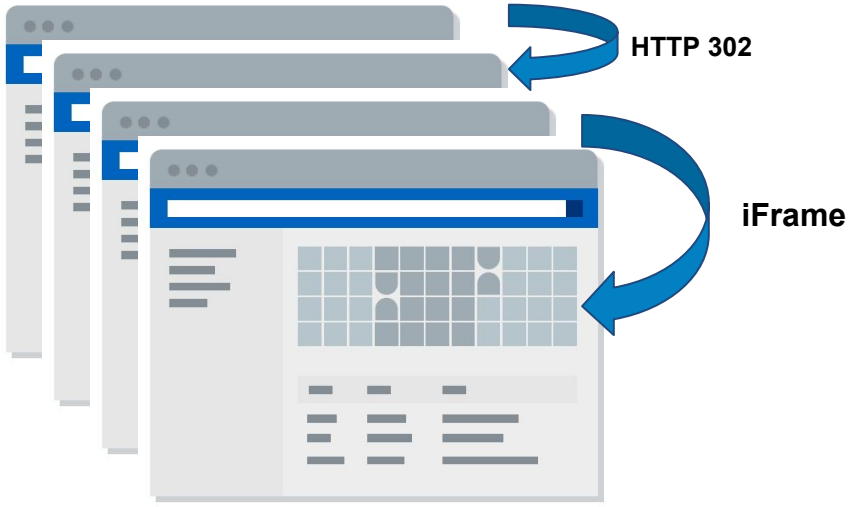
Caveats and Considerations

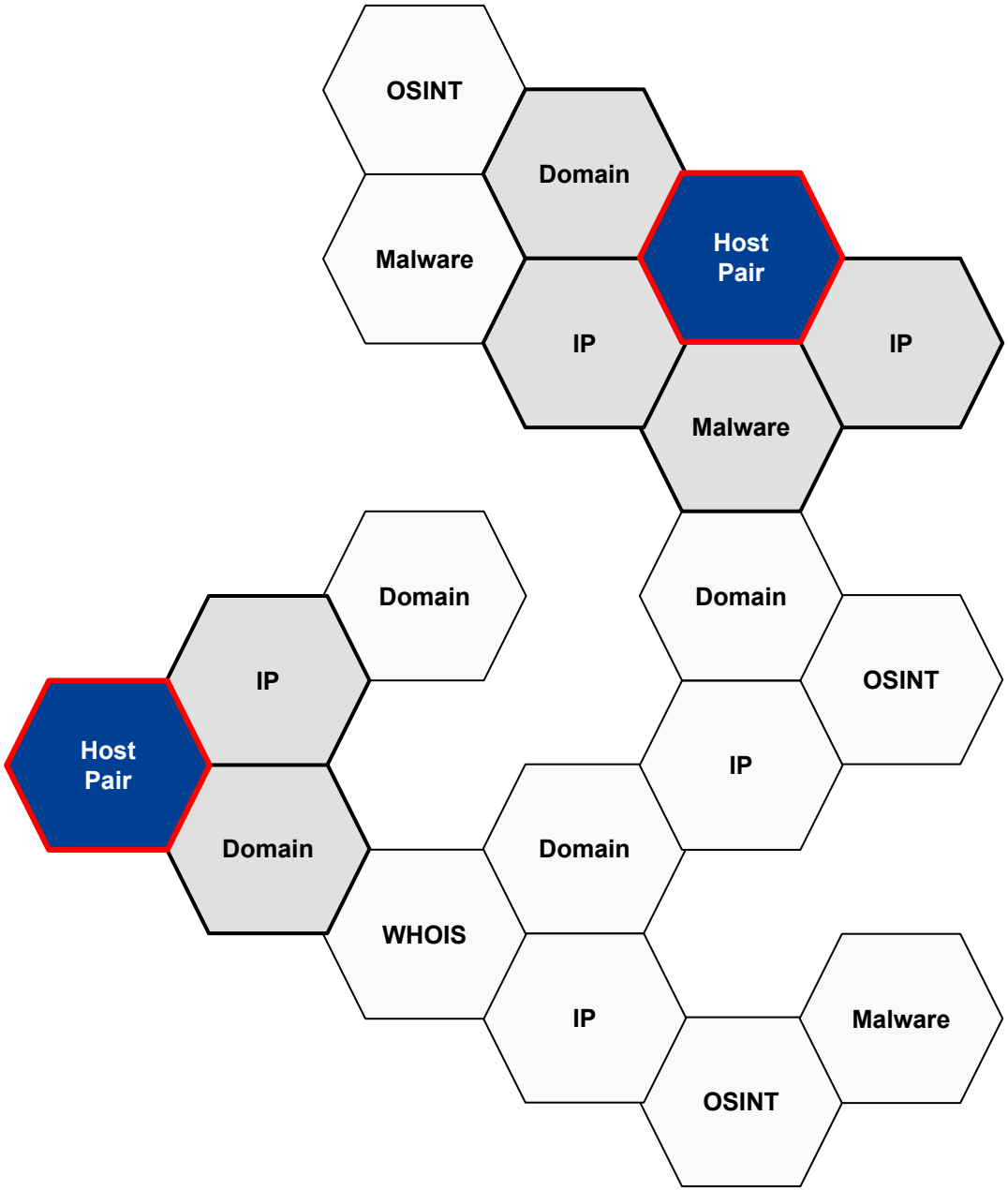
- Valid vs malicious
  - Who owned it first
  - Known good to known bad issue
- Outsourced web presence
  - Third-party developed sites
  - Analytics as a service
- Social providers will have an extreme amount of overlap

Host Pairs

Infrastructure pairs based on observed sequence chains from web crawls

- Sequence pairing direction (parent or child)
- Cause for the pairing
  - iFrame, HTTP 302, Javascript



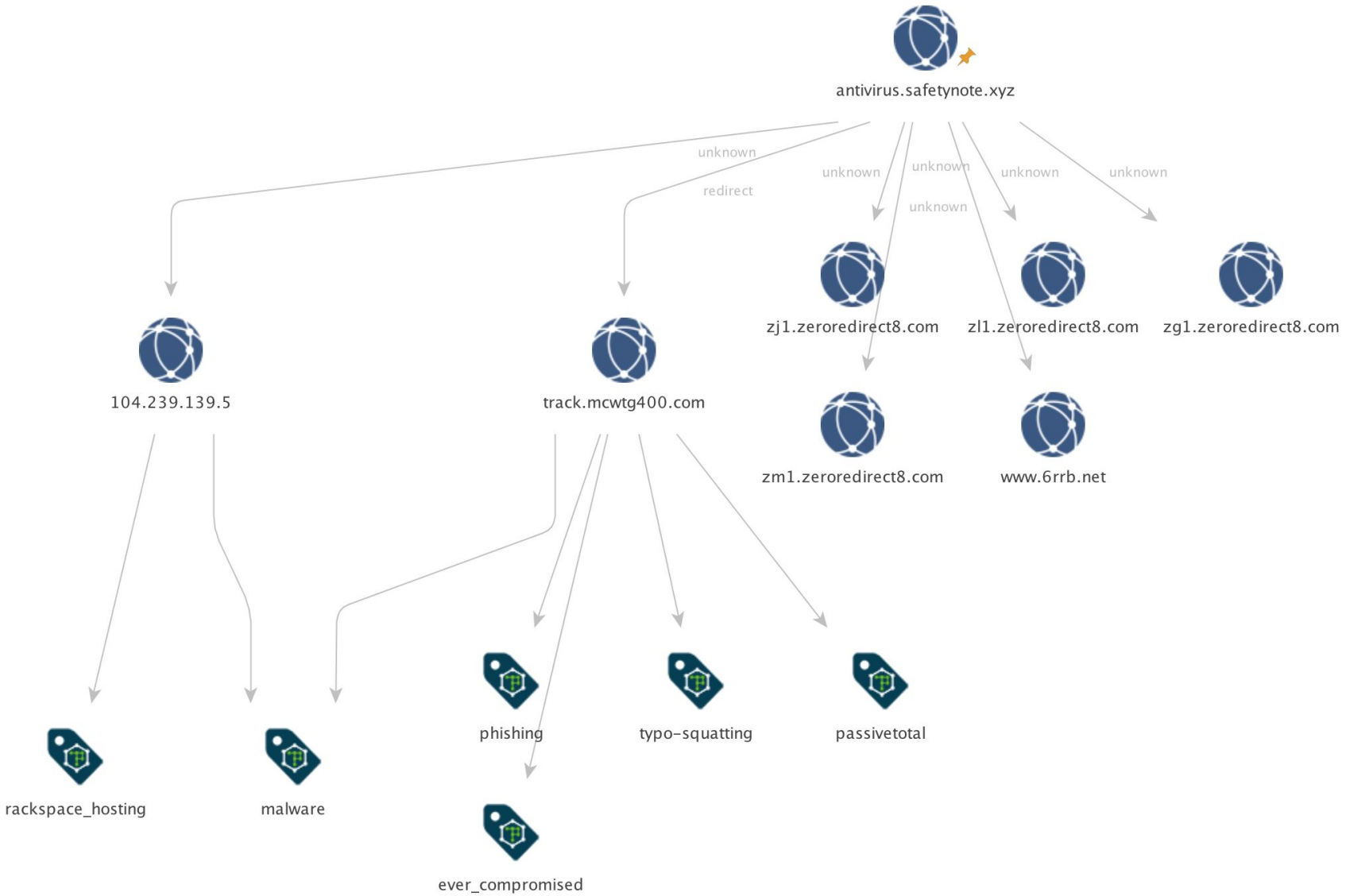

OSINT
Domain
Malware
Host Pair
IP
IP
Malware
Domain
Domain
IP
OSINT
Host Pair
IP
Domain
Domain
WHOIS
IP
Malware
OSINT

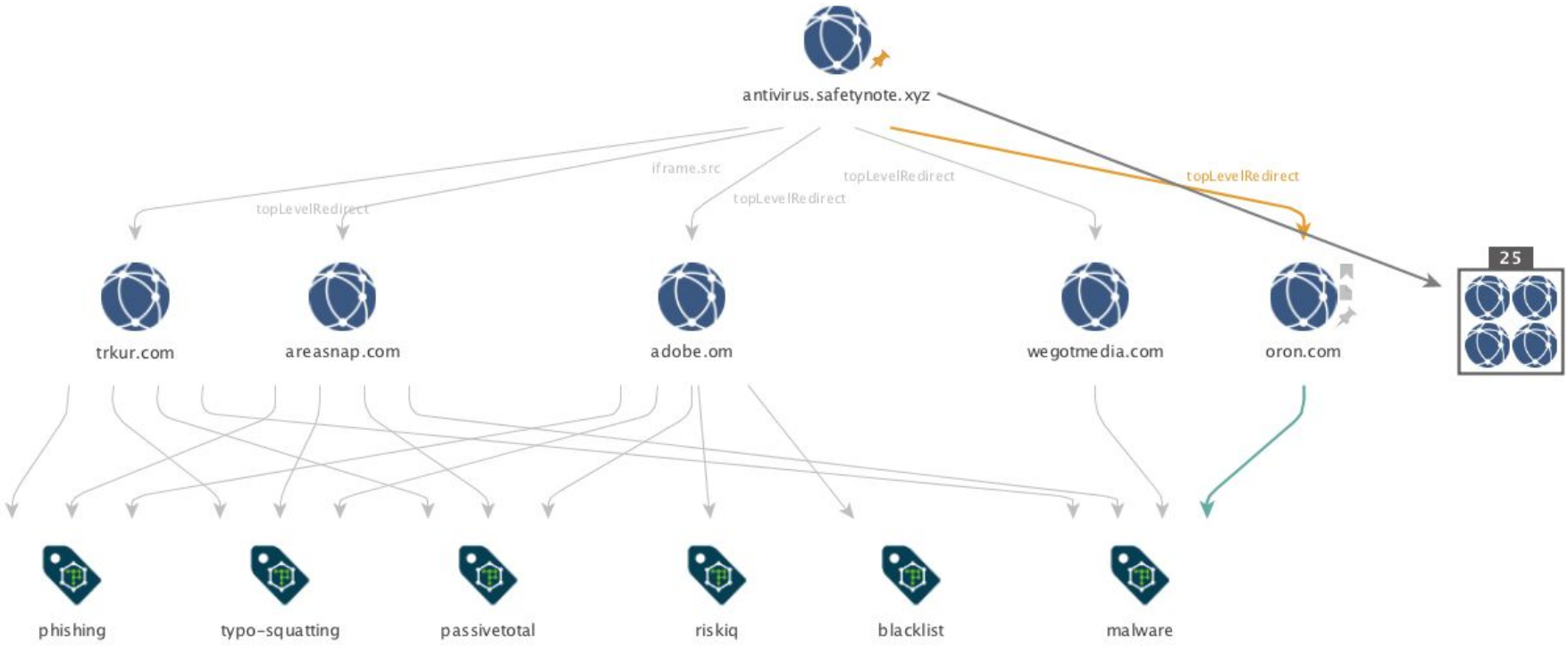
Example: antivirus.safetynote.xyz

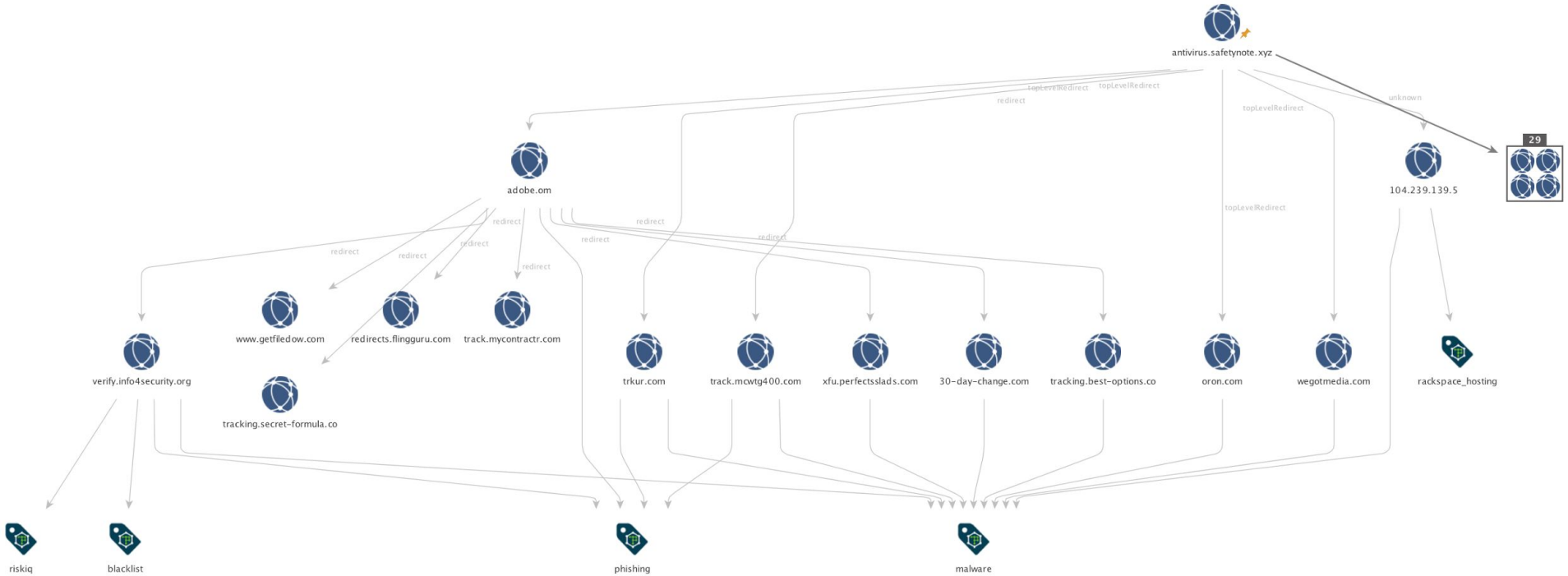
- Connections via host attributes
  - 38 host pairs (6 parents, 32 children)
  - Mostly redirects to typo-squat domains
  - Connections to additional blacklisted domains
- Domain WHOIS is privacy protected
- Resolving IP address leads to Amazon instance

Heatmap | WHOIS | Host Pairs 38

Hostname	First Seen	Last Seen	Direction	Cause
areasnap.com	2016-03-10 23:37:41	2016-04-23 05:01:01	child	iframe.src
track.mcwtg400.com	2016-02-01 08:00:22	2016-02-04 04:13:50	parent	redirect
hadie.persianfbpages.com	2016-02-04 04:13:49	2016-02-04 04:13:49	child	topLevelRedirect
wsj.om	2016-02-02 01:44:21	2016-02-04 02:27:22	child	topLevelRedirect
pogo.om	2016-02-01 19:26:21	2016-02-04 02:09:22	child	topLevelRedirect
toysrus.om	2016-02-02 03:17:21	2016-02-03 21:17:21	child	topLevelRedirect
united.om	2016-02-03 02:28:20	2016-02-03 19:57:21	child	topLevelRedirect
sohu.om	2016-02-03 19:27:43	2016-02-03 19:27:43	child	topLevelRedirect


antivirus.safetynote.xyz
unknown
redirect
unknown
unknown
unknown
unknown
unknown
zj1.zeroredirect8.com
zl1.zeroredirect8.com
zg1.zeroredirect8.com
104.239.139.5
track.mcwtg400.com
zm1.zeroredirect8.com
www.6rrb.net
phishing
typo-squatting
passivetotal
rackspace_hosting
malware
ever_compromised


antivirus.safetynote.xyz
iframe.src
topLevelRedirect
topLevelRedirect
topLevelRedirect
topLevelRedirect
25
trkur.com
areasnap.com
adobe.om
wegotmedia.com
oron.com
phishing
typo-squatting
passivetotal
riskiq
blacklist
malware

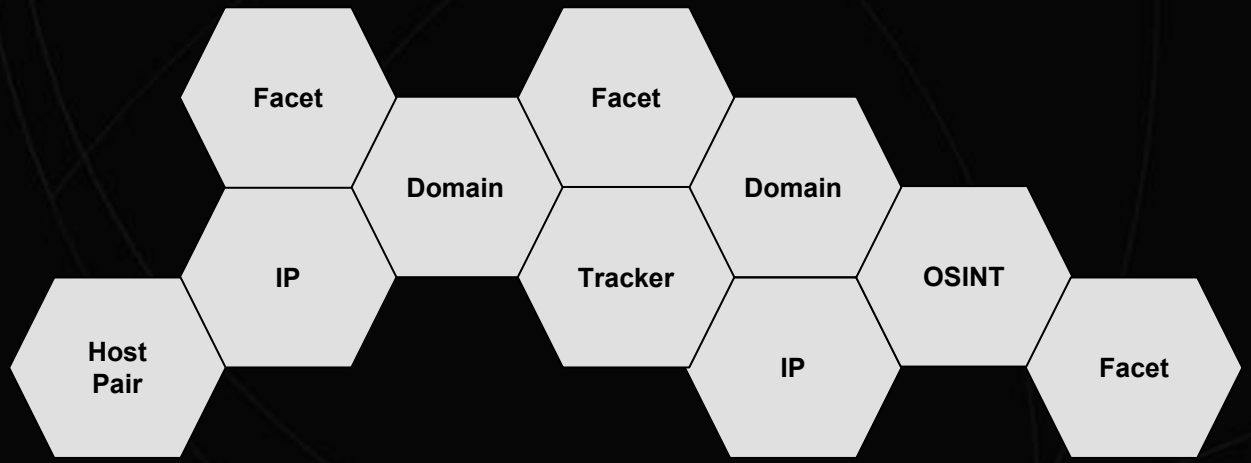

antivirus.safetynote.xyz
topLevelRedirect
topLevelRedirect
redirect
topLevelRedirect
unknown
29
adobe.om
104.239.139.5
redirect
redirect
redirect
redirect
redirect
redirect
redirect
redirect
topLevelRedirect
www.getfiledow.com
redirects.flingguru.com
track.mycontractr.com
verify.info4security.org
tracking.secret-formula.co
trkur.com
track.mcwtg400.com
xfu.perfectsslads.com
30-day-change.com
tracking.best-options.co
oron.com
wegotmedia.com
rackspace_hosting
riskiq
blacklist
phishing
malware

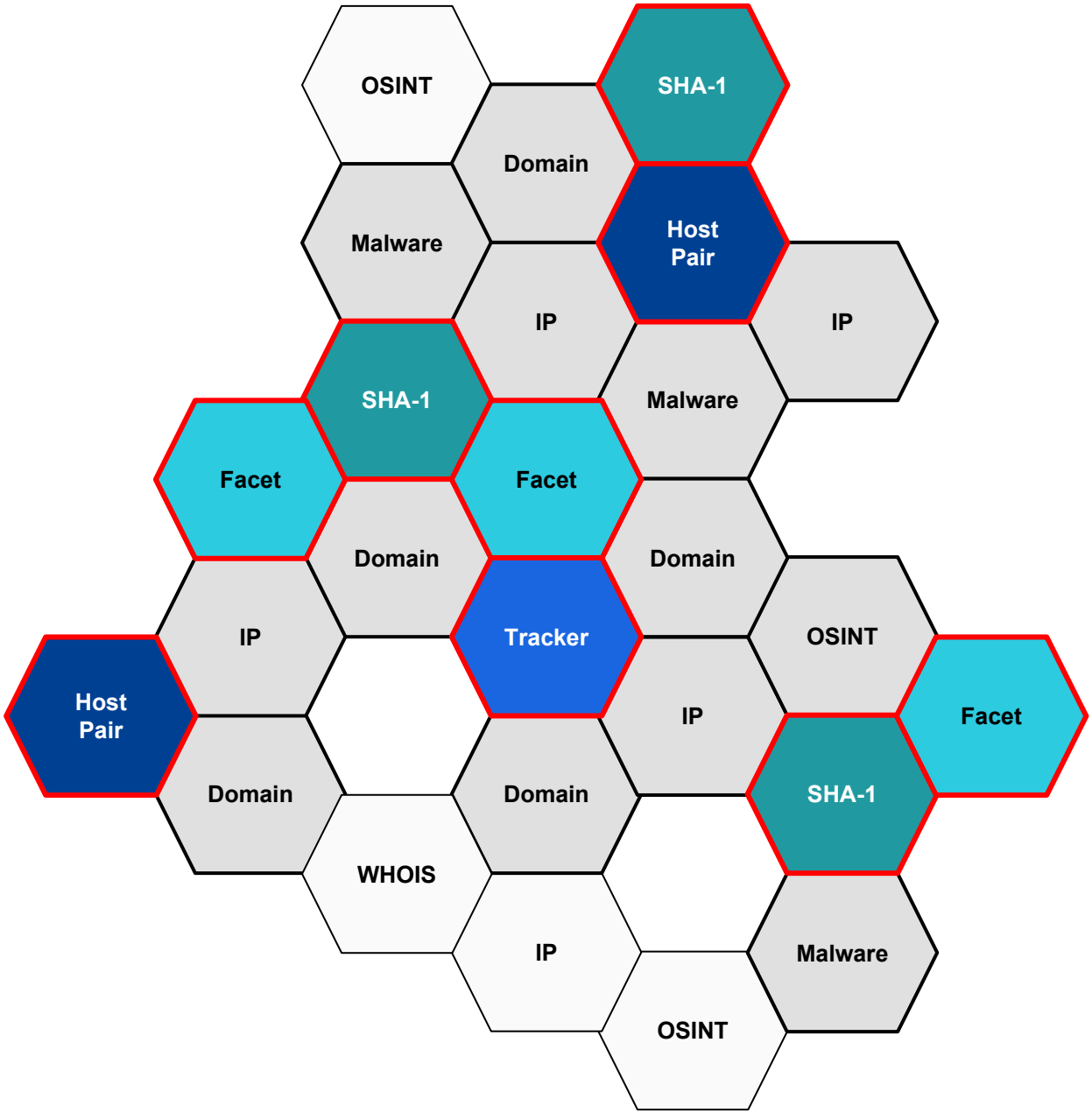
Caveats and Considerations

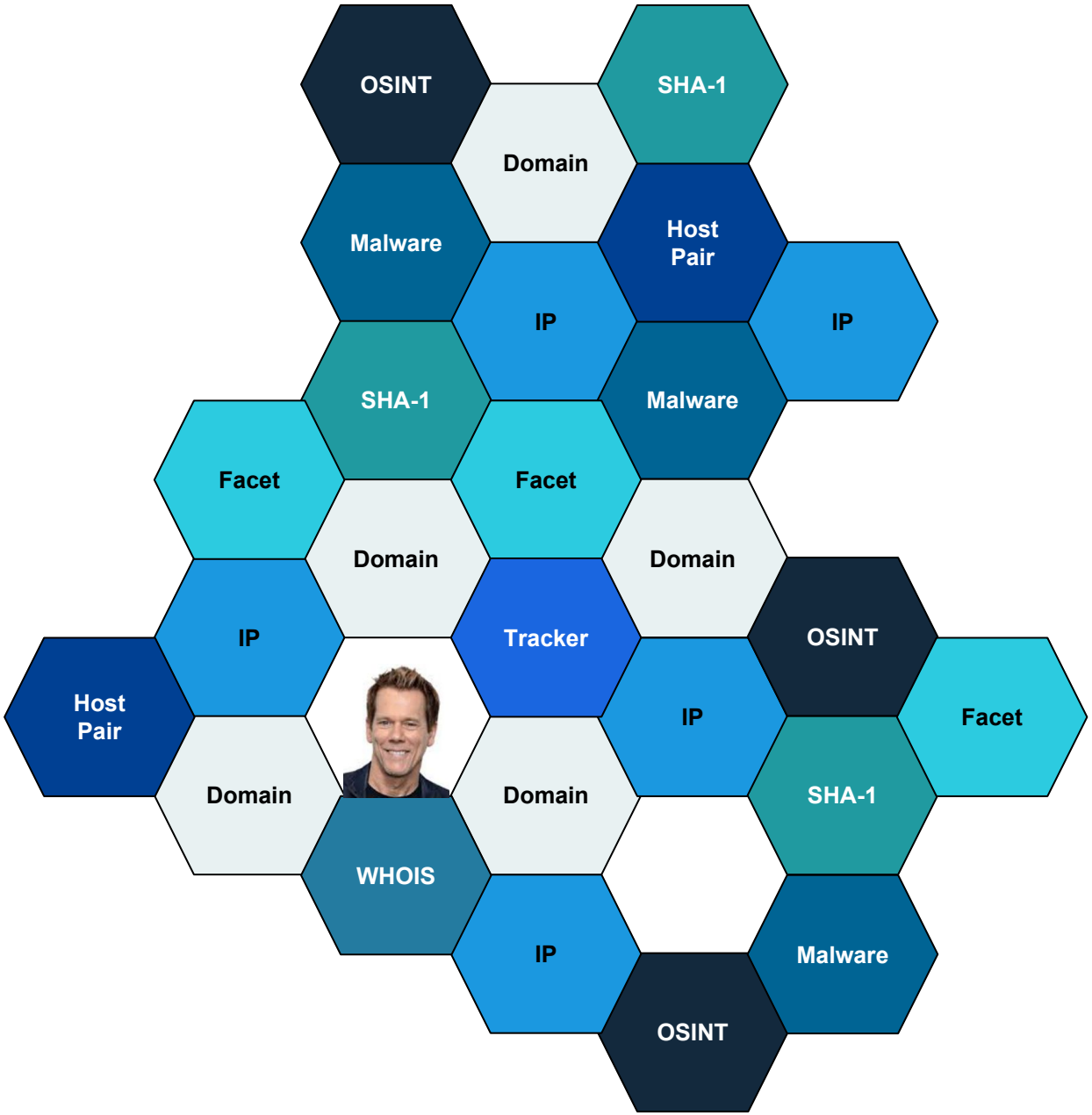
- Modern websites are highly dependent
  - Increased volume of data
- Frequency is a good determining factor for surfacing interesting pairs
- Cause between pairs and the relationship

Building Stronger Connections

- Primary benefits of enhanced datasets
  - Support existing common dataset links
  - Fills gaps in data coverage
  - Provides even more context
  - Allows for more automated analysis




OSINT
SHA-1
Domain
Malware
Host Pair
IP
IP
SHA-1
Malware
Facet
Facet
Domain
Domain
IP
Tracker
OSINT
Host Pair
IP
Facet
Domain
Domain
SHA-1
WHOIS
IP
Malware
OSINT


OSINT
SHA-1
Domain
Malware
Host Pair
IP
IP
SHA-1
Malware
Facet
Facet
Domain
Domain
IP
Tracker
OSINT
Host Pair
IP
Facet
Domain
Domain
SHA-1
WHOIS
IP
Malware
OSINT

Wrap-up

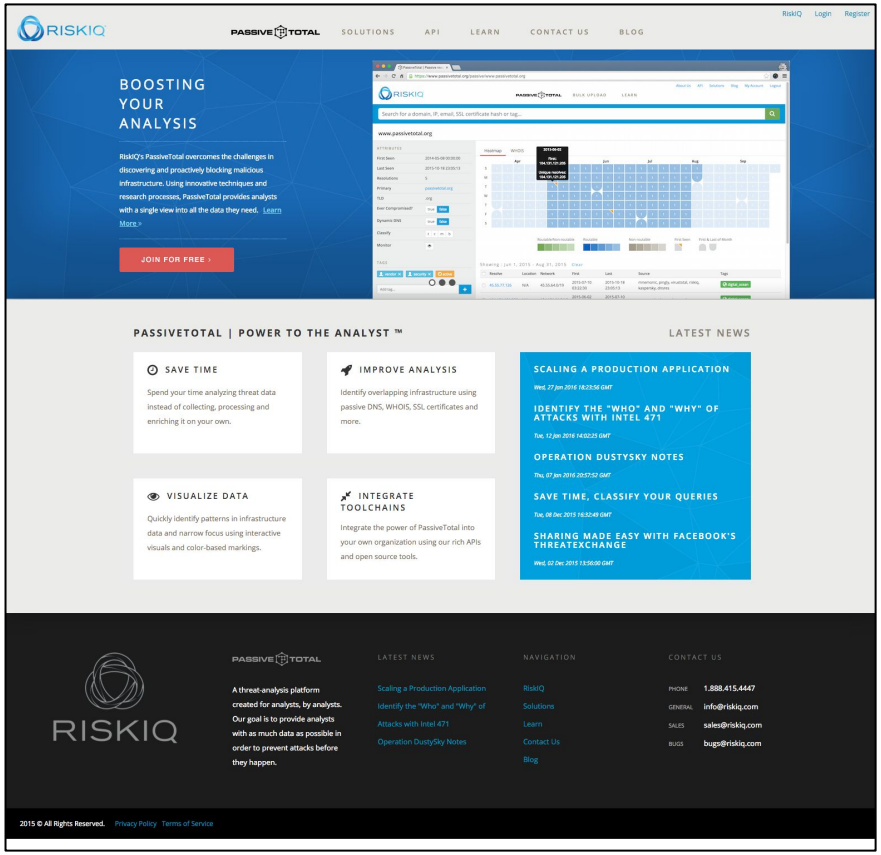
Honorable Mentions

- SSH keys
  - Finds load-balanced servers, overlaps heavy with SSL certificates, no major breakthroughs
- Content hashing
  - Locate related pages based on structure or use of specific libraries or dependency files
  - Promising detecting phishing pages, modified libraries or unique code
- Web components
  - Host, content and dependency changes over time
  - Useful for plotting out evolutions, but not operational for correlation yet

Test Driving the Data

Data available inside of PassiveTotal through the web, API and other integrations

- Allows for user feedback and collaboration
- Overlays all datasets in one place
- Pivot on numerous facets and data points
- Integrate our data within your environment



RISKIQ

Questions?