Catching the Cyber Criminals

Revealing Malicious Infrastructures with OpenDNS

Dragan Novakovic
Cisco Systems
Abstract

Cyber Criminals are increasingly exploiting the Internet to build agile and resilient infrastructures, and consequently to protect themselves from being exposed and taken over.

The Internet is an open system, meaning that the information to expose those infrastructures is available somewhere. The challenge is that fragments of data broken up and spread across the web are not immediately visible.

Connecting the dots, being able to analyze a diverse set of information made of billions of pieces of discrete data allows us to build maps that reveal where malicious infrastructure is hidden and where attacks are being staged. This turns the tables on traditional security with a new approach where the defender takes the upper hand on the attacker, being able to pivot through criminal infrastructure.
Ransomware Kill Chain with DNS

Elements of the infrastructure are involved in each phase

User Clicks a Link or Malvertising

Exploit Kit

Malicious Infrastructure

Ransomware Payload

Email w/ Malicious Attachment

DNS

DNS

DNS

Encrpyption Key C2 Infrastructure
Why leverage DNS to Detect and Block Threats

Most attacker C2 is initiated via DNS lookups with some non-Web callbacks

15% of C2 bypasses Web ports 80 & 443

Lancope Research (now part of Cisco)
- millions of unique malware samples from small office LANs over 2 years

91% of C2 can be blocked at the DNS layer

Cisco AMP Threat Grid Research
- millions of unique malware samples submitted to sandbox over 6 months

NOTE1: Visual Investigations of Botnet Command and Control Behavior
- malware reached out to 150,000 C2 servers over 100,000 TCP/UDP ports
- malware often used 866 (TCP) & 1018 (UDP) “well known” ports, whereas legitimate traffic used 166 (TCP) & 19 (UDP) ports

NOTE2: 2016 Cisco Annual Security Report
- 9% had IP connections only and/or legitimate DNS requests
- 91% had IP connections, which were preceded by malicious DNS lookups
- very few had no IP connections
DNS is *Used by Every Device* on Your Network

**ANY OWNER**
network’s DHCP tells every connected device where to point DNS

**ANY TOPOLOGY**
no matter how your LAN or WAN is set up, it simply works

**ANY OPERATING SYS**
Win, Mac, iOS, Android, Linux, custom app servers, and even IoT
Global Network Built Into the Fabric Of the Internet

- **ZERO** added latency
  - peer w/top 500 ISPs & CDNs

- **100%** uptime since 2006
  - 400+ Gbps capacity, DDoS protection & global fail-over

- **2%** worldwide activity
  - globally-shared DNS cache
Gather Intelligence At the DNS Level

**Request Patterns**
- Compromised systems
- Command & control callbacks
- Malware & phishing attempts
- Algorithm-generated domains
- Domain co-occurrences
- Newly registered domains

**Authoritative Logs**
- Newly staged infrastructures
- Malicious domains, IPs, ASNs
- DNS hijacking
- Fast flux domains
- Related domains

Any Device

Recursive DNS

Global Cache of DNS Responses

Authoritative DNS

root
dom.com.
domain.com.
Some Security Graph Metrics

GLOBAL NETWORK
- 90B+ DNS requests/day
- 65M+ biz & home users
- 100% uptime
- Any port, protocol, app

UNIQUE ANALYTICS
- security research team
- automated classification
- BGP peer relationships
- 3D visualization engine

SECURITY GRAPHS
- > 10 TB/day
- ~46M nodes per day
- ~174M edges per day

© 2016 Cisco and/or its affiliates. All rights reserved. Cisco Public
What does OpenDNS Provide

- **Umbrella (Enforcement)**
  - 208.67.222.222

- **Investigate (Intelligence)**
  - Domain, IP, ASN, EMAIL, HASH

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>IDENTITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>MALWARE</td>
<td>INTERNAL IP</td>
</tr>
<tr>
<td>C2 CALLBACK</td>
<td>HOSTNAME</td>
</tr>
<tr>
<td>PHISHING</td>
<td>AD USER</td>
</tr>
<tr>
<td>CUSTOM (API)</td>
<td>HOSTNAME</td>
</tr>
</tbody>
</table>

OpenDNS API

STATUS & SCORES
CO-OCCURRENCES
RELATIONSHIPS
ATTRIBUTIONS
PATTERNS & GEOS

CONSOLE

SIEM, etc.
Traditional Domain Reputation Techniques Are No Longer Effective

- Domain Reputation is not effective on identifying certain groups of threats such as Exploit Kits or Domain Shadowing
  - Malicious domains move quickly from IP to IP
  - Legitimate domains may be compromised to distribute malware
  - Malware can use DGA/Domain Shadowing
- Conceived for an Internet of 10 years ago
One Domain to Rule Them All!

“FAST FLUX”

@23.4.24.1
@44.6.11.8
@34.4.2.110
@129.3.6.3

bad.com?

CALLBACK

DOMINO GENERATION ALGORITHM

@34.4.2.110
@12.3.2.1
@8.2.130.3

rnd.net?
rnd.com?
rnd.biz?

CALLBACK

DOMINO SHADOWING

@129.3.6.3
@23.4.24.1
hjacklegitdomain.com

decg
dojamg

EK LANDING PAGE
Who Says That a Crystal Ball Is the Only Way to Predict Cyber Attacks?
A Diamond (And a Bunch of Math) Can Help!
Making Sense of Data

Raw Data
- Collection
- Storage
- Persistence

Data Science
- Analytics
- Statistics
- Machine Learning

Information
- Indicators
- Insights
- Stories

Collect
Analyze
Predict
How Security Classification Works

Collect
Millions of data points (DNS Requests) per second

Analyze
Applying statistical models and data science

Predict
Pivot through malicious Infrastructures and identify where the attacks are staged.

- a.ru
- b.com
- 7.7.1.3
- e.net
- p.com/jpg
- 5.9.0.1
Predictive Detectors Used by OpenDNS

- SecureRank
- Co-Occurrences
- NLPRank
- DGA Detectors
- Spike Detectors
- Predictive IP Space Monitoring
SecureRank

- Abstract DNS traffic in a bipartite graph
- Color the graph with different shades of “red” to indicate bad domains, and “green” for good ones.
- There are clusters of ‘red’ separated from “green” zones with few intra links.
- Domains requested by known infected clients but never requested by clean ones are most likely to be bad.
- SecureRank2 is designed to identify these domains
Assigning a Score to Malicious Domains

\[ SR_{\text{Domain}} = \sum \frac{SR_{\text{Client}}}{L_{\text{Client}}} \]

\[ SR_{\text{Client}} = \sum \frac{SR_{\text{Domain}}}{L_{\text{Domain}}} \]

\[ SR_{C}(A) = SR_{D}(D_1) + SR_{D}(D_2) + SR_{D}(D_3) + \frac{SR_{D}(D_4)}{2} \]

\[ SR_{C}(B) = \frac{SR_{D}(D_4)}{2} + SR_{D}(D_5) + SR_{D}(D_6) \]

\[ SR_{D}(D_4) = \frac{SR_{C}(A)}{4} + \frac{SR_{C}(B)}{3} \]

https://labs.opendns.com/2013/03/28/secure-rank-a-large-scale-discovery-algorithm-for-predictive-detection/
The Algorithm in Action

**Link Analysis**
- March through global DNS query data and map the requestor-requestee pairs as a graph.

**Initialize**
- Negative ranks to known blacklisted domains and positive ranks to known whitelisted domains.

**Iteration**
- Run The Algorithm through different iterations

**Final Rank**
- Final ranks are generated when the ranks converge after a number of iterations.
Co-Occurrences

• Sequence of DNS requests to domains that co-occur within seconds of each other across a statistically significant number of streams.

• For a domain, being a co-occurrence is not necessarily a bad thing.

• But what if one of the domains involved is part of a malicious campaign?
Co Occurrences can be correlated with more “traditional” Techniques

CO-OCCURRENCES
domain-to-domain request sequences via recursive DNS

PASSIVE DNS & WHOIS
present & past relationships for domains-to-IP/nameserver/email via authoritative DNS & DNS registrars

INFRASTRUCTURES
domain-to-IP-to-AS relationships via graphing BGP routing data
NLPRank

Identifies malicious domain-squatting and targeted C2 or phishing domains

1. Read APT reports
2. Patterns in domains used in attacks
   - Domain spoofing used to obfuscate
   - Often saw brand names and terms like “update”
   - Examples: update-java[.]net, adobe-update[.]net
3. Checked data & confirmed intuition
   - Dictionary & company names merged
   - Change small # of characters to obfuscate
   - Domains hosted on ASNs unassociated w/company
   - Different webpage fingerprints
4. Built model and continue to tune
   - Detects fraudulent brand domains:
     - 1linkedin.net
     - 2linkedin.com

NLP = natural language processing
NLPRank Detections: DarkHotel

- adobeupdates[.]com
  - This domain is currently in the OpenDNS Security Labs block list
  - Classifier prediction: suspicious
  - OpenDNS Security Graph Score: 64

- microsoft-xpupdate[.]com
  - This domain is currently in the OpenDNS Security Labs block list
  - This domain has a suspicious ASN score
  - This domain has a suspicious prefix score
  - Classifier prediction: suspicious
  - OpenDNS Security Graph Score: 97
NLPRank Detections: Carbanak

- **update-java[.]net**

  - One or more of the IP addresses that this domain resolves to are currently blocked by OpenDNS
  - This domain is currently in the OpenDNS Security Labs block list
  - This domain has a suspicious ASN score

- **adobe-update[.]net**

  - This domain is currently in the OpenDNS Security Labs block list

  ![Graph showing DNS queries over time](image)
DGA Detection
Identifies malicious domain-squatting and targeted C2 or phishing domains

“N-gram” analysis
Do sets of adjacent letters match normal language patterns?

Entropy analysis
Does the probability distribution of letters appear random?

yfrscsddkkdl.com
qgmcoqesagomme.org
iyyxtxxdeypk.com
diiqngijkpop.ru
SPRANK

SPRANK detects domains showing as a sudden surge, or a spike, in DNS queries

Spike Detection
Domain History Filter
Qtype Filter
Domain Records Filter
Exploit Kits
Fake Software
Phishing
DGA
Spam Forums

More exploit kits, fake software, Phishing, etc.
Expand the Intelligence Graph by pivoting around IP, Prefix, ASN, hoster, registrant email to catch more malware domains

© 2016 Cisco and/or its affiliates. All rights reserved. Cisco Public
What Does a Malicious Connection Sounds Like?

What if we could model the traffic spikes as sound waves and identifies “spike behavior” typical of domains used for malware campaigns such as exploit kits, DGAs, fake software, phishing, etc…

Example of An Exploit Kit

Example of a DGA
Example of a DGA
Spike Detection

- New Series of threats such as Exploit Kits or Domain Shadowing make many of the classical domain reputation or IP reputation methods ineffective.
- Spike defined as a jump in traffic over a two hour window.
- Use predetermined threshold. Helps filter out Google, Facebook, etc.
- Use a MapReduce algorithm to calculate domains that spike.
- Output 50-100k domains each hour.
Domain History Filter

- Past query history is used to help remove benign domains and focus in Exact Domain Match ones.

- Allows to eliminate all domains with more than X consecutive non-zero hours of traffic.

- Based on current EK domains traffic patterns, only keep domains that feature Y consecutive most recent non-zero hours of traffic.
Query Type Filter

• Look at past history, DNS Query types, all existing DNS records of a domain, unique IPs, unique resolvers, etc.

• Partition based on Query types Distribution:

  ✓ 1 – A Record
  ✓ 15 – MX Record
  ✓ 16 – TXT Record
  ✓ 99 – SPF Record
  ✓ 255 – ANY Record
Domain Records Filter

• Check for all DNS records available for a domain: the existence/non-existence of certain records helps narrow down the purpose of a domain.

• Partition based on DNS records:
  • A
  • MX
  • TXT
  • CNAME
  • NS, specific name servers, indicative of compromise or malware
Empirical Data on the Model Efficacy

On Average, only 16% of security vendors catch the domains identified by SPRank.

Of the 200 domains, observed in a one hour period, 70 of the compromised domains had not been identified by any other vendor.

SPRank has a 100% success rate of discovering malicious domains before other security vendors (tested hourly against VirusTotal).

https://blog.opendns.com/2015/11/19/opendns-cracks-predictive-security/
Predictive IP Space Monitoring

Predictive IP Space Monitoring is used to further drill into associated indicators by analyzing 8 different recorded hosting patterns:

- Compromised domains, i.e. “domain shadowing”
- Domain shadowing on multiple hosting IPs
- Sibling peripheral ASNs and bulk malware IP setup
- Leaf ASNs
- Offshore registration and diversification of IP space
- Rogue ASN and affiliated hosters
- Abuse of large hosting providers
- Shady hosts within larger hosting providers
Expanding The Selection

Predictive IP Space Monitoring expands the selection of SPRank, to determine which **domains will be the source of future malicious activity**.

For 1 malicious domain identified by SPRank, Predictive IP Space Monitoring predicted 340 Additional domains

https://blog.opendns.com/2015/11/19/opendns-cracks-predictive-security/
Pivoting Through the Attack Infrastructure with Just one Piece of Information (1/2)

**Alerts and risk scores**
Summarise the suspicious activity identified for the domain

- This domain has a suspicious A3N score
- This domain is associated with the following attack: APT & C
- Geo distance between hosts serving this domain is fairly high

**Global Requests Patterns**
Shows an abnormal spike in traffic, which highlights when the attack launched

**Domain Tagging**
Shows history of when the malware was associated with malware or botnet activity

<table>
<thead>
<tr>
<th>Period</th>
<th>Category</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sep 23, 2015 - Current</td>
<td>Botnet</td>
<td></td>
</tr>
</tbody>
</table>

**IP Geography Analysis**
Reveals the domain is hosted by IP addresses on different networks in more than 20 countries, which, for instance, is unusual for legitimate country code top-level domains.

**WHOIS Record Data**
Shows the domain was recently created and registered by someone who used the same email address to register other malicious domains
Pivoting Through the Attack Infrastructure with Just one Piece of Information (2/2)

Mappings of IP prefixes and ASNs
Highlights where the domain is hosted and confirm it’s a “bad neighbor” of many other malicious domains. Pivot on the IP or ASN for more details.

Anomaly Detection
Identifies that this is a fast flux domain, a technique used to hide malware sites behind IPs that are constantly changing.

Related Domains and Co-Occurrences
Identify other domains that were queried with a high statistical frequency right before or after this one and are likely related to the same attack.

Passive DNS Data
Provides insight into the history of the mapping between domains and IPs: this domain was associated with different IPs when detected the first time.

Named Threat Attribution
Confirms that the domain was associated with a particular malware family or botnet C&C.

Starting from a single piece of data, it is possible to quickly investigate the domain leveraging a single, correlated source and speed up incident response.

## Passive DNS Data
Provides insight into the history of the mapping between domains and IPs: this domain was associated with different IPs when detected the first time.

### First seen | Last seen | IPs

### Related Domains and Co-Occurrences
Identify other domains that were queried with a high statistical frequency right before or after this one and are likely related to the same attack.

- www.donnelcompras.es (-63.33) andecomm.com.br (31,03) corporates.dowconvenie.it (23,65)
- www.donnelcompras.es (-13) andecomm.com.br (9) 123contactform.com (7)
- corporates.dowconvenie.it (6) www.donnelcompras.mx (5) forms.dowconvenie.it (3)
- cdn.ubenda.com (3)

## Mappings of IP prefixes and ASNs
Highlights where the domain is hosted and confirm it’s a “bad neighbor” of many other malicious domains. Pivot on the IP or ASN for more details.

### Current Information
- AS: 123456789 (AS Name: Network Domain)
- IP Prefix: 192.168.1.0/24
- Domain: mymaliciousdomain.com

### Historical Data
- Date: 2023-01-25
- Previous AS: 897654321 (AS Name: Other Network Domain)

## Anomaly Detection
Identifies that this is a fast flux domain, a technique used to hide malware sites behind IPs that are constantly changing.

### First seen | Last seen | IPs

## Named Threat Attribution
Confirms that the domain was associated with a particular malware family or botnet C&C.

### This domain is associated with the following attack: ZBot Fast Flux Botnet

## CO-OCCURRENCES
- www.donnelcompras.es (-63.33) andecomm.com.br (31,03) corporates.dowconvenie.it (23,65)
- www.donnelcompras.es (-13) andecomm.com.br (9) 123contactform.com (7)
- corporates.dowconvenie.it (6) www.donnelcompras.mx (5) forms.dowconvenie.it (3)
- cdn.ubenda.com (3)

## Related Domains
- www.donnelcompras.es (-63.33) andecomm.com.br (31,03) corporates.dowconvenie.it (23,65)
- www.donnelcompras.es (-13) andecomm.com.br (9) 123contactform.com (7)
- corporates.dowconvenie.it (6) www.donnelcompras.mx (5) forms.dowconvenie.it (3)
- cdn.ubenda.com (3)
Visualizing Data with OpenGraphiti

- OpenGraphiti, is the Open Source interactive data visualization engine developed by OpenDNS.

- Used by security analysts and researchers, it pairs visualization and Big Data to create 3D representations of threats.

- The basic concept is that information is processed more efficiently when it is presented in visual rather than text form.

- OpenGraphiti can uncover sophisticated behaviors and relationships associated with cyber-attacks.
Using Semantic Networks to Visualize Threats

- Graph = Set of Nodes
- Node = Concept, Edge = Relationship
- Agents populate the graph
- A semantic network can be represented as a graph connecting any kind of information by any kind of relationship
- They can be used to model nearly everything and can be applied to a wide range of problems
Our View of the Internet
providing visibility into global Internet activity (e.g. BGP, AS, WHOIS, DNS)
Public DNS providers struggle to resolve websites amid the @Dyn DDoS. OpenDNS is holding up. Here’s servers trying to resolve Twitter

<table>
<thead>
<tr>
<th>Agent</th>
<th>IP Address</th>
<th>AS</th>
<th>Response Time</th>
<th>Status</th>
<th>Location</th>
<th>Network</th>
</tr>
</thead>
<tbody>
<tr>
<td>183-sootlinux</td>
<td>208.67.222.222</td>
<td>AS36692</td>
<td>14 ms</td>
<td>✔</td>
<td>US - Fremont</td>
<td>Abovenet Comm (AS17029)</td>
</tr>
<tr>
<td>74.207.242.5</td>
<td>Linode, LLC</td>
<td>AS63949</td>
<td>64 ms</td>
<td>✗</td>
<td>US - Fremont</td>
<td></td>
</tr>
<tr>
<td>45.33.58.84</td>
<td>Linode, LLC</td>
<td>AS63949</td>
<td>66 ms</td>
<td>✗</td>
<td>US - Fremont</td>
<td></td>
</tr>
<tr>
<td>8.8.8.8</td>
<td>Google Inc.</td>
<td>AS15169</td>
<td>2 ms</td>
<td>✔</td>
<td>US - Fremont</td>
<td></td>
</tr>
<tr>
<td>209.18.47.62</td>
<td>Time Warner</td>
<td>AS7843</td>
<td>13 ms</td>
<td>✗</td>
<td>US - Fremont</td>
<td>Time Warner C (AS10796)</td>
</tr>
<tr>
<td>208.67.222.222</td>
<td>OpenDNS, LLC</td>
<td>AS36692</td>
<td>26 ms</td>
<td>✔</td>
<td>US - Cleveland</td>
<td>Time Warner C (AS10796)</td>
</tr>
<tr>
<td>209.18.47.61</td>
<td>Time Warner</td>
<td>AS7843</td>
<td>22 ms</td>
<td>✗</td>
<td>US - Cleveland</td>
<td>Time Warner C (AS10796)</td>
</tr>
<tr>
<td>8.8.8.8</td>
<td>Google Inc.</td>
<td>AS15169</td>
<td>37 ms</td>
<td>✗</td>
<td>US - Cleveland</td>
<td>Time Warner C (AS10796)</td>
</tr>
<tr>
<td>24.196.64.53</td>
<td>Charter Commu</td>
<td>AS20115</td>
<td>42 ms</td>
<td>✗</td>
<td>US - St. Louis</td>
<td>Charter Commu (AS20115)</td>
</tr>
<tr>
<td>208.67.222.222</td>
<td>OpenDNS, LLC</td>
<td>AS36692</td>
<td>40 ms</td>
<td>✔</td>
<td>US - Cleveland</td>
<td>Charter Commu (AS20115)</td>
</tr>
<tr>
<td>8.8.8.8</td>
<td>Google Inc.</td>
<td>AS15169</td>
<td>37 ms</td>
<td>✗</td>
<td>US - Cleveland</td>
<td>Charter Commu (AS20115)</td>
</tr>
<tr>
<td>68.94.157.1</td>
<td>AT&amp;T Services</td>
<td>AS7018</td>
<td>21 ms</td>
<td>✗</td>
<td>US - Cleveland</td>
<td>AT&amp;T Services (AS7018)</td>
</tr>
<tr>
<td>68.94.156.1</td>
<td>AT&amp;T Services</td>
<td>AS7018</td>
<td>960 ms</td>
<td>✗</td>
<td>US - Cleveland</td>
<td>AT&amp;T Services (AS7018)</td>
</tr>
</tbody>
</table>
Predict and Prevent Attacks Before They Happen

• With its **90+ Billion** DNS requests analyzed per day OpenDNS has a comprehensive and privileged view of the Internet

• The analysis of this massive and diverse dataset allows to build models and detectors able to identify where attacks are staged.

• Starting from a single piece of information it is possible to pivot through the malicious infrastructure, **exposing** attackers and **predicting** their moves before they happen

• On the other hand, the Internet is not unlimited, so there are zones more prone to be exploited by criminals, or even recycled.
Start your Free Trial Now

https://signup.opendns.com/freetrial/

www.opendns.com
labs.opendns.com
opengraphiti.com
Thank you

Dragan Novakovic
Cisco Systems
dnovakov@cisco.com