Walk Softly and Carry 26 Trillion Sticks

Andrew Hay
Director of Research
ahay@opendns.com
About Andrew Hay

- Director of Research at @OpenDNSLabs
- Recovering Chief Evangelist (@CloudPassage)
- Embittered Sr. (Industry) Analyst (@451Research)
- Academia Escapee (@ulethbridge)
- Former Island Dweller (Capital G Bank)
- Former Herder of Devs (Q1 Labs, Inc.)
- Nokia Firewall Guy
- Author, Blogger, BBQ Enthusiast, and Broken Rugby Prop
Agenda

Overview

Dyre Botnet Research & The ‘Pentagram Network’

DGA Infrastructure Identification

Detecting APT Campaigns With NLP

Research on the Internet of Things (RIoT)
Overview

OpenDNS sees a lot of data

Total Activity
Number of DNS requests per day in billions

89,300,761,598,965 DNS requests since July 10, 2006.
Let’s Quantify “Lots”...

NUMBERS WE ARE PROUD OF

>2% of Internet requests
65 million daily active users
70 billion+ daily DNS requests
Let’s Quantify “Lots”...

24 datacenters around the world...more to come
What This Presentation Is About

- The malware research and IR community has long spent their expertise **DISSECTING** malware binaries, **EXAMINING** footprints on victims’ systems, and **HUNTING** for attribution.

- Instead of studying the **SYMPTOMS** (malware), however, at OpenDNS we study the **BEHAVIOR OF VICTIMS/TARGETS** (patients) and the **BEHAVIOR OF ACTORS/CAMPAIGNS/INFRASTRUCTURE** (diseases).

- This presentation **WAS** going to be about our 2014 findings...
  - But we’ve already done so much in 2015 that the focus has shifted.
Agenda

- Overview
- Dyre Botnet Research & The ‘Pentagram Network’
- DGA Infrastructure Identification
- Detecting APT Campaigns With NLP
- Research on the Internet of Things (RIoT)
Dyre Botnet Research

- Typically distributed by the **CUTWAIL** botnet
  - Via spam emails
  - With links to Dropbox or Cubby files

- **HARVESTS CREDENTIALS** through traffic interception
  - Primarily targets online banking websites for ACH/wire fraud

- When graphed, we noticed the emergence of **UNIQUE SHAPES**
The Dyre ‘Pentagram Network’
The Dyre ‘Pentagram Network’

- **INITIAL QUERY** to domain (red)
- **CO-OCCURRENCES WITH DOMAINS** in a ‘pentagram network’ (white)
- Finally **DIRECTED OUT** of ‘pentagram network’ (blue)
Linear Queries & Time Compression

TIME-1

RED

TIME-2

WHITE

TIME-3

BLUE
The Dyre ‘Pentagram Network’

- Repeatable observations with **DYRE BOTNET VARIANTS**
  - E.g. The Dyre ‘Blade Network’
- Want to expand to other botnets
- Still a work in progress...
Agenda

- Overview
- Dyre Botnet Research & The ‘Pentagram Network’
- DGA Infrastructure Identification
- Detecting APT Campaigns With NLP
- Research on the Internet of Things (RIoT)
Exploring DGAs

- Domain Generation Algorithm (DGA)
- Used to **GENERATE DOMAINS PROGRAMMATICALLY**
- Typically rely on a **SEED** of some sort
  - Date, time, keyword, etc.
  - Allows for the registration of domains that no human would ever type
- Not just used for malware
  - Also used for marketing campaigns
DGA-based Malware C2 Domain Examples

- Ramnit
  - lyxbotuappfreadkfk[.]com

- Tinba
  - eersrjoffvo[.]com, rlhuplxoghh[.]com, vhtsiililikr[.]com

- Emotet
  - ywnjdkgrvivotium[.]eu

- Expiro
  - bpu1ilh[.]mazxceo[.]info
Regular Chain
Irregular Chain
Irregular Circuit
Different Topologies?

- Consider **CO-OCCURRENCE TIME WINDOW**
- Some clients are **NOISIER** than others
- Diversity in domain lookups of clients
- Nature of the DGA algorithm
  - Frequency, redundancy, etc.
Agenda

- Overview
- Dyre Botnet Research & The ‘Pentagram Network’
- DGA Infrastructure Identification
- Detecting APT Campaigns With NLP
- Research on the Internet of Things (RIoT)
Detecting APT Campaigns With NLP

“How do you detect Evil™ in DNS records using lexical features of FQDNs and VALIDATE results?”

- Created a fraud detection system using **NATURAL LANGUAGE PROCESSING (NLP)** techniques and traffic features to identify domain-squatting & brand spoofing in DNS/HTTP
  - a technique commonly used by phishing and APT C2 servers

- Called **NLPRANK** (patent pending)
  - United States ODNS=-1009US0; 62/167,178

- Created by OpenDNS Security Labs researcher **JEREMIAH O’CONNOR**
Targeted Attacks

- From a psychological perspective, if you were a high-profile exec for company **WHAT KIND OF LINKS WOULD YOU CLICK ON?** What are your interests?
- Commodity phishing follows the same psychology
- Topics of interest:
  - $$$, Bank Account/CCs, Financial
  - News
  - Security/Software updates
  - Social Network
High Level Overview of Process

- ASN Filtering
- Defining Malicious Language Within FQDNs
- HTML Content Analysis
High Level Overview of Process

- Domains exhibiting fraudulent behavior are observed to be hosted on ASNs **NOT ASSOCIATED WITH THE COMPANY**

- **E.g. Carbanak**
  - `adobe-update[.]net`
    - ASN 44050, PIN-AS Petersburg Internet Network LLC in Russia
  - `update-java[.]net`
    - ASN 44050, PIN-AS Petersburg Internet Network LLC in Russia
High Level Overview of Process

1. CyrusOne LLC
2. Unified Layer
3. OVH OVH SAS,FR
4. GoDaddy.com
5. HostDime.com
7. Hostinger International Limited
8. Hetzner Online AG,DE
9. Liquid Web
10. Cloudie Limited

ASN Filtering
High Level Overview of Process

- Run basic **TEXT METRICS** on the data, **GAIN INTUITIONS** about the data and **EXTRACT** important words/substrings in APT FQDN datasets
- APT domains **EXHIBIT SIMILAR LEXICAL FEATURES** to commodity phishing domains
- Important look at **WORD CO-OCCURRENCES** (bigrams, trigrams, etc.)
High Level Overview of Process

- From APT data sets **EXTRACTED WORDS FROM DICTIONARY AND APPLIED STEMMING** looking at word stats:
  - Top counts: mail, news, soft, serv, updat, game, online, auto, port, host, free, login, link, secur, micro, support, yahoo

- Words **OFTEN APPEAR WITH EACH OTHER** (a.k.a. Bigram Collocations)
  - adobe-update
  - update-java[.]com
  - Idea: *brandname + ad-action word [.] tld*
High Level Overview of Process

- **Dark Hotel (Kaspersky):**
  - adobeupdates[.]com
  - adobeplugs[.]net
  - adoberegister[.]flashserv[.]net
  - microsoft-xpupdate[.]com

- **Carbanak (Kaspersky):**
  - update-java[.]net
  - adobe-update[.]net

- **APT 1 Domains (Mandiant):**
  - gmailboxes[.]com
  - microsoft-update-info[.]com
  - firefoxupdate[.]com
High Level Overview of Process

- Creating a **MALICIOUS LANGUAGE** derived from lexical features of FQDNs from APT/Phishing data sets
- Built **CORPUS** of domains similar to examples in previous slide
- Create **CUSTOM DICTIONARIES**
  - Brandname Dictionary
    - E.g. google, gmail, paypal, yahoo, bankofamerica, wells Fargo
  - Custom set of stemmed common malicious words
    - E.g. secur, updat, install, etc.
High Level Overview of Process

- Recreate the researcher’s mind
  - Visit site in Tor browser
  - Researcher processes information on site, looks for clues, gains summary
  - Makes decision whether site is legit/malicious

- Specifically for Phishing Sites:
  - What makes people fall for this?
    - Site will be near copy of legitimate site it’s intending to spoof
  - **HOW CAN WE AUTOMATE THIS PROCESS?**
  - Can we apply document similarity algorithms?
High Level Overview of Process

▪ Examples from Apple phishing page
  - Links
    - https://id.apple.com/IDMSAccount/myAccount.html?appIdKey=45571f444c4f547116bfd052461b0b3ab1bc2b445a72138157ea8c5c82fed623&action=register&language=US-EN
  - Images
    - <img alt="" src=https://www.chase.com/etc/designs/chasecomhomepage/images/homepage_background_1px.jpg/>
High Level Overview of Process

HTML Content Analysis
Whitelisting

ASN Filter

Popularity Check

GET Request to URL/Retrieve Content

Edit-distance check/dictionary check

Block List

Auto-Tag Brand/Topic In Phishtank

Email Daily Results

Top N Similar Documents

TF-IDF Weighting

Counts of words on page

Latent Semantic Analysis

Compare Cosine Similarity

- Acquiring Data  - Filtering  - NLP  - Output

URL Feed (HTTP/PT)  

AS-N Filter
Success Identifying Different Types of Attacks

- **Success in TRAINING:**
  - Detecting
  - Careto
  - APT Domains Darkhotel/Carbanak/APT1 etc.
  - AJAX Hacking Group/Flying Kitten infostealer C&C
  - Operation Pawn Storm
  - Operation Saffron Rose

- **Success on LIVE DATA:**
  - Various Exploit Kits
  - Fast-Flux domains
  - And new stuff...
Agenda

Overview

Dyre Botnet Research & The ‘Pentagram Network’

DGA Infrastructure Identification

Detecting APT Campaigns With NLP

Research on the Internet of Things (RIoT)
OpenDNS IoT Research Report

- First-ever **DATA-DRIVEN RESEARCH** about IoT
- Before: mainly hypothetical scenarios, opinion-based, FUD
- Goal of research:
  - **USE A DATA-DRIVEN METHODOLOGY**
  - **EXPLORE THE PREVALENCE OF IOT DEVICES IN ENTERPRISE NETWORKS**
  - **UNDERSTAND THE POTENTIAL SECURITY RISKS & VULNERABILITIES**
DNS Data

- Used ANONYMIZED DNS DATA from the OpenDNS Global network (which sees traffic from 10K+ ENTERPRISES WORLDWIDE)
- All data was meticulously SANITIZED so as not to reveal client IP addresses or individual company names
- Examined BACKEND INFRASTRUCTURE IOT DEVICES CONNECT TO for updating, storing data, etc.
  - Used 3rd party tools to see which domains were secure & which contained known vulnerabilities
Overview

• The findings were **METICULOUSLY CALCULATED AND RECALCULATED**

• Our findings uncover the extent to which **IOT PERMEATES NEARLY EVERY MAJOR MARKET VERTICAL**
  
  • including energy, healthcare, education, consumer electronics, manufacturing, defense, government, financial institutions, and retail
About The Research: Methodology

• **SMALL SUBSET** of frequently accessed IoT domains chosen for the first stage

• Include a number of **POPULAR PRODUCT MANUFACTURERS, IOT ENABLEMENT PLATFORMS, AND NEW-TO-MARKET IOT STARTUPS**...
  - spanning both commercial product lines and personal electronics intended for home use
GeoIP Location of Client FQDN Queries

Client FQDN Queries For Complete Data Population (n = 561,816)

Client FQDN Queries For Enterprise Customer Data Population (n = 68,044)
### Vertical Industries Identified (56 total)

- Accounting, HR, and Admin
- Advertising, Communications, and PR
- Aerospace, Weapons, Defense Contractors
- Agriculture
- Apparel
- Biotechnology
- Business Services
- Chemicals
- Churches and Houses of Worship
- Communications
- Computer Software
- Construction and Architecture
- Consulting
- Consumer Products
- Distributor and Wholesaler
- Education
- Education: K-12
- Electronics
- Energy/Utilities
- Engineering
- Entertainment
- Environmental
- Financial Services
- Food & Beverage
- Government
- Health Insurance
- Healthcare
- Higher Education
- Hospitality
- Insurance
- ISP
- IT Services
- Legal Services
- Leisure
- Libraries
- Machinery
- Managed Service Providers
- Manufacturing
- Media
- Mining/Raw Materials
- Non-Profit
- Not for Profit
- Oil and Gas
- Other
- Pharmaceutical manufacturers
- Professional Services
- Property Management
- Real Estate
- Recreation
- Restaurants
- Retail
- Services
- Technology
- Technology Reseller
- Telecommunications
- Transportation, Shipping, and Logistics

There were **56 DISTINCT INDUSTRY VERTICALS IDENTIFIED** in our data set in addition to an “Other” category for industries that our business intelligence team didn’t have a 1:1 mapping for.
## Top 30 Queried FQDNs

<table>
<thead>
<tr>
<th>DOMAIN</th>
<th>QUERY COUNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>gld.push.samsungosp.com.</td>
<td>6342</td>
</tr>
<tr>
<td>api.samsungosp.com.</td>
<td>6028</td>
</tr>
<tr>
<td>us-odc.samsungapps.com.</td>
<td>5227</td>
</tr>
<tr>
<td>vas.samsungapps.com.</td>
<td>4306</td>
</tr>
<tr>
<td>eu2-scloud-proxy.ssp.samsungosp.com.</td>
<td>4228</td>
</tr>
<tr>
<td>iphone-api.fitbit.com.</td>
<td>3853</td>
</tr>
<tr>
<td>iphone-client.fitbit.com.</td>
<td>3500</td>
</tr>
<tr>
<td>hub.samsungapps.com.</td>
<td>3066</td>
</tr>
<tr>
<td>android-api.fitbit.com.</td>
<td>2867</td>
</tr>
<tr>
<td>desktop-static.fitbit.com.</td>
<td>2677</td>
</tr>
<tr>
<td>android-client.fitbit.com.</td>
<td>2587</td>
</tr>
<tr>
<td>iedc.fitbit.com.</td>
<td>2513</td>
</tr>
<tr>
<td>auth.samsungosp.com.</td>
<td>2347</td>
</tr>
<tr>
<td>sc-auth.samsungosp.com.</td>
<td>1624</td>
</tr>
<tr>
<td>mkt-odc.samsungapps.com.</td>
<td>1413</td>
</tr>
<tr>
<td>desktop-client.fitbit.com.</td>
<td>1397</td>
</tr>
<tr>
<td>client.fitbit.com.</td>
<td>1314</td>
</tr>
<tr>
<td>updates.logitech.com.</td>
<td>1243</td>
</tr>
<tr>
<td>hub-odc.samsungapps.com.</td>
<td>850</td>
</tr>
<tr>
<td>nexusapi.dropcam.com.</td>
<td>729</td>
</tr>
<tr>
<td>weather.nest.com.</td>
<td>457</td>
</tr>
<tr>
<td>ca-odc.samsungapps.com.</td>
<td>436</td>
</tr>
<tr>
<td>uk-odc.samsungapps.com.</td>
<td>426</td>
</tr>
<tr>
<td>img.samsungapps.com.</td>
<td>389</td>
</tr>
<tr>
<td>wearable.samsungapps.com.</td>
<td>301</td>
</tr>
<tr>
<td>bwhome.logitech.com.</td>
<td>299</td>
</tr>
<tr>
<td>bwsnotif01.logitech.com.</td>
<td>291</td>
</tr>
<tr>
<td>bwsupdate01.logitech.com.</td>
<td>291</td>
</tr>
<tr>
<td>brinks.axeda.com.</td>
<td>274</td>
</tr>
<tr>
<td>bwsregister01.logitech.com.</td>
<td>221</td>
</tr>
</tbody>
</table>
Top 5 IoT Infrastructure Hosting Providers

<table>
<thead>
<tr>
<th>HOSTING PROVIDER</th>
<th>TOTAL QUERY COUNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon</td>
<td>23,181</td>
</tr>
<tr>
<td>SoftLayer Technologies</td>
<td>15,992</td>
</tr>
<tr>
<td>Verizon Business</td>
<td>11,127</td>
</tr>
<tr>
<td>Fastly</td>
<td>5,208</td>
</tr>
<tr>
<td>E.I. du Pont de Nemours and Co.</td>
<td>1,634</td>
</tr>
</tbody>
</table>
### Top 5 IoT Infrastructure Hosting Providers

<table>
<thead>
<tr>
<th>Hosting Provider</th>
<th>Total Query Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon</td>
<td>23,181</td>
</tr>
<tr>
<td>SoftLayer Technologies</td>
<td>15,992</td>
</tr>
<tr>
<td>Verizon Business</td>
<td>11,127</td>
</tr>
<tr>
<td>Fastly</td>
<td>5,208</td>
</tr>
<tr>
<td>E.I. du Pont de Nemours and Co.</td>
<td>1,634</td>
</tr>
</tbody>
</table>

Unsurprisingly, Amazon Web Services (AWS) has the highest count of DNS queries for IoT-related products.
## Top 5 IoT Infrastructure Hosting Providers

<table>
<thead>
<tr>
<th>HOSTING PROVIDER</th>
<th>TOTAL QUERY COUNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon</td>
<td>23,181</td>
</tr>
<tr>
<td>SoftLayer Technologies</td>
<td>15,992</td>
</tr>
<tr>
<td>Verizon Business</td>
<td>11,127</td>
</tr>
<tr>
<td>Fastly</td>
<td>5,208</td>
</tr>
<tr>
<td>E.I. du Pont de Nemours and Co.</td>
<td>1,634</td>
</tr>
</tbody>
</table>

Fastly received for vendors like Fitbit and Withings. This shows that IoT vendors are taking the time to architect resiliency into their respective service platforms ahead of major service disruptions.
## Top 5 IoT Infrastructure Hosting Providers

<table>
<thead>
<tr>
<th>HOSTING PROVIDER</th>
<th>TOTAL QUERY COUNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon</td>
<td>23,181</td>
</tr>
<tr>
<td>SoftLayer Technologies</td>
<td>15,992</td>
</tr>
<tr>
<td>Verizon Business</td>
<td>11,127</td>
</tr>
<tr>
<td>Fastly</td>
<td>5,208</td>
</tr>
<tr>
<td>E.I. du Pont de Nemours and Co.</td>
<td>1,634</td>
</tr>
</tbody>
</table>

The majority of these queries were related to Samsung mobile platforms, but we also observed Dropcam and IoT platform provider Thingworx related queries.
Top 5 Countries Hosting IoT Infrastructure

<table>
<thead>
<tr>
<th>Country</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>555</td>
</tr>
<tr>
<td>DK</td>
<td>1029</td>
</tr>
<tr>
<td>GB</td>
<td>2669</td>
</tr>
<tr>
<td>CA</td>
<td>2817</td>
</tr>
<tr>
<td>US</td>
<td>56,957</td>
</tr>
</tbody>
</table>
A number of the ASNs used to host the vendor’s IoT infrastructure have also been **IDENTIFIED BY OPENDNS SECURITY LABS TO BE HOSTING MALICIOUS DOMAINS.**
Notable Observed Devices

- Nest thermostats
- Dropcam video cameras
- Fitbit fitness devices
- Western Digital “My Cloud” hard drives
- Logitech devices
- Samsung
  - Mobile devices (e.g. tablets, phones)
  - Smart Televisions
Notable Observed Devices and Infrastructure

- **Alarm Systems**
  - Honeywell Security Alarmnet

- **IoT management platforms**
  - mnubo
  - Axeda (a PTC company)
  - Arrayent
  - Thingworx (a PTC company)

- **HVAC**
  - Ecobee
Notable Observations: FREAK

- 184 unique FQDNs were found to be susceptible to CVE-2015-0204 (a.k.a. FREAK)

- An **ALARMING NUMBER** of organizations were observed querying these vulnerable systems
  - Energy (42)
  - Media (27)
  - Utilities (24)
  - Healthcare (23)
  - Education (51)
  - Not For Profit (35)
<table>
<thead>
<tr>
<th>INDUSTRY VERTICALS</th>
<th>CVE-2015-0204 VULNERABLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accounting/HR/Admin</td>
<td>17</td>
</tr>
<tr>
<td>Biotechnology</td>
<td>2</td>
</tr>
<tr>
<td>Computer Software</td>
<td>20</td>
</tr>
<tr>
<td>Construction/Architecture</td>
<td>2</td>
</tr>
<tr>
<td>Education: College</td>
<td>4</td>
</tr>
<tr>
<td>Education: K-12</td>
<td>47</td>
</tr>
<tr>
<td>Energy</td>
<td>42</td>
</tr>
<tr>
<td>Financial Services</td>
<td>9</td>
</tr>
<tr>
<td>Food &amp; Beverage</td>
<td>14</td>
</tr>
<tr>
<td>Government</td>
<td>2</td>
</tr>
<tr>
<td>Healthcare</td>
<td>23</td>
</tr>
<tr>
<td>Hospitality</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INDUSTRY VERTICALS</th>
<th>CVE-2015-0204 VULNERABLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Libraries</td>
<td>1</td>
</tr>
<tr>
<td>Logistics / Transportation</td>
<td>1</td>
</tr>
<tr>
<td>Managed Service Providers</td>
<td>2</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>1</td>
</tr>
<tr>
<td>Media</td>
<td>27</td>
</tr>
<tr>
<td>Not For Profit</td>
<td>35</td>
</tr>
<tr>
<td>Retail</td>
<td>18</td>
</tr>
<tr>
<td>Technology</td>
<td>1</td>
</tr>
<tr>
<td>Telecommunications/ISP’s</td>
<td>1</td>
</tr>
<tr>
<td>Transportation/Shipping/Logistics</td>
<td>2</td>
</tr>
<tr>
<td>Utilities</td>
<td>24</td>
</tr>
</tbody>
</table>
Notable Observations: Western Digital

- Data shows that **NOT ALL WD2GO.COM DOMAINS ARE VULNERABLE**, however.
- Of the 70 unique MyCloud storage endpoints, **ONLY 30** were found to be vulnerable to CVE-2015-0204.
Notable Observations: Samsung Smart TVs

- Smart televisions **FREQUENTLY DEPLOYED IN BOARDROOMS**, replacing overhead projectors
- Offices also use them in common areas as active signage for advertising or employee/location information
- To focus our research on Samsung Smart TV-related domains, we decided to obtain a TV (model **UN32H5203AF**) and test
- Just how talkative are they...
Notable Observations: Samsung Smart TVs

- Nothing directly malicious was found in the examination of the Samsung Smart TV’s network communications.
- The fact that a Smart TV does so ALMOST EVERY MINUTE IT’S POWERED ON—even without user interaction—is CONCERNING.
- It makes the use of these devices much easier to determine from outside a corporate network.
Takeaways...

- Organize your data in graphs to quickly spot relationships and anomalies
  - Great for botnet fingerprinting

- OpenGraphiti is a free, Open Source, and awesome data visualization tool
  - Used to visualize any relational data as an interactive 2D or 3D model
Takeaways...

- NLPRank development continues to iterate
- Exploiting a technique commonly used by phishing and APT C2 servers
- Patent pending
  - United States ODNS=-1009US0; 62/167,178
- Created by OpenDNS Security Labs researcher Jeremiah O’Connor
- Full blog post here:
  - https://labs.opendns.com/2015/03/05/nlp-apt-dns/
Takeaways...

- Confirm the presence of new IoT devices in your environment
- Spend time researching the devices to understand the risks
- Put some pressure on device manufacturers and platform vendors for better documentation
- How can you gain visibility into IoT devices in your environment?
  - DNS is one way...
- The full report can be downloaded at:
  - https://opendns.com/iot
Questions?

Andrew Hay, Director of Research
ahay@opendns.com
labs.opendns.com