THE SECURITY WOLF OF WALL STREET: FIGHTING CRIME WITH HIGH-FREQUENCY CLASSIFICATION & NLP
Jeremiah O’Connor
Thibault Reuille
OpenDNS
$ WHOIS THIBAULT

• Manager of Development in Research Team at OpenDNS.
• Creator of OpenGraphiti.
• Creator of the Avalanche project
• Focus: Data Visualization, 3D Graphics, Graph Theory and Real-time systems.
$ WHOIS JEREMIAH

- Mad Scientist at OpenDNS/Cisco Labs
- M.S. in Computer Science from University of San Francisco
- Previously worked at Mandiant (IR/DNS Research), Evernote (AppSec/IR), Uber (Data Science)
- Career Goals: Solve interesting problems (Networking/Security, Bioinformatics, GPS Tracking, Video Games, etc.)
- Proud SFSPCA Pitbull Puppy owner
INTRODUCTION
Data Science & Network Security

Big Security Data

DNS Traffic:
• ~80B DNS requests per day

HTTP Traffic:
• ~10.1M requests per day

Daily Tasks:
• Detection Algorithms, Security Data Analysis,
• Distributed Systems, Big Data Engineering, Data Viz
REAL-TIME!
THE AVALANCHE PROJECT
HIGH FREQUENCY TRADING vs TRAFFIC CLASSIFICATION

THE WOLF OF WALL STREET

- HistoricaL DATABASE
- BACKTESTING
- LOG Aggregation
- ML Training
- Stock exchange
- Quant server
- STRATEGIES
- Portfolio / Risk Management
- Execution
- Resolvers
- Blocking Whitelisting
- Domain Tagging
- Models & Classifiers
- Predicted Impact on Users
WHAT IS AVALANCHE?

OVERVIEW AND TECHNICAL DETAILS

Open source project:
- [http://github.com/ThibaultReuille/avalanche](http://github.com/ThibaultReuille/avalanche)

“Real-time” data processing framework

Modular, parallel and distributed design

Written with Python and ZeroMQ

Platform for RT OpenDNS models (Private):
- NLP-Rank
- DNS Tunnelling
- DGA classifier
- Others...
AVALANCHE PIPELINE

DIVIDE AND CONQUER

PLUGIN

PLUGIN RACK
AVALANCHE DESIGN

DIVIDE AND CONQUER

INPUT QUEUE -> NODE (PLUGIN) -> OUTPUT QUEUE

ZeroMQ Socket

ZeroMQ Socket
import json
import plugins.base

class Plugin1(plugins.base.Plugin):
    def __init__(self, info):
        # NOTE: The info argument contains the full node definition
        # written in the pipeline configuration file.
        pass

    def process_message(self, message):
        # NOTE: Here we can process the message, add fields, remove, etc.
        # Returning None drops the message from the pipeline.
        return message

class Plugin2(plugins.base.Plugin):
    def __init__(self, info):
        # NOTE: The info argument contains the full node definition
        # written in the pipeline configuration file.
        pass

    def run(self, node):
        # NOTE: Each node runs on its own thread/process,
        # Here we enter our infinite loop.
        while True:

            # NOTE: Read incoming data sent to our node
            data = node.input.recv()

            # NOTE: Parse it as a JSON message
            message = json.loads(data)

            # NOTE: This template plugin doesn't do anything except using a passthru filter.
            # This is where the processing would actually happen in a real processor.
            # You can send whatever data you like in the output stream. That can be a modified
            # version of the incoming messages or any other message of your creation.

            # NOTE: Send it back through the pipeline
            node.output.send_json(message)

if __name__ == "__main__":
    print("Please import this file!")
RUN AVALANCHE

$ ./avalanche.py path/to/my_pipeline.json 10000

Things you get for free:
- Modularity
- Multi-Threading
- A library of plugins ready-to-use
- Reusability & collaboration
- An insanely fast messaging system
THE RESEARCH PIPELINE
AVALANCHE CLUSTER

HIGH LEVEL VIEW

RESOLVER  AMAZON S3  AVALANCHE  PRODUCTION
8 Amazon instances
Master distributes work
  • Round-robin
  • “Fire and forget”
Slaves process the chunks
Results are stored in central database
4 Avalanche Pipeline (Processes)
Monitored with DaemonTools
Results are cross-checked and filtered
TRAFFIC SPEED vs AVALANCHE PIPELINE

NUMBERS DON'T LIE

<table>
<thead>
<tr>
<th>QUERIES / CHUNK</th>
<th>AUTHLOGS (AMS.M1)</th>
<th>QUERYLOGS (AMS.M1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noon (UTC)</td>
<td>564 752</td>
<td>6 147 997</td>
</tr>
<tr>
<td>Midnight (UTC)</td>
<td>412 050</td>
<td>3 315 157</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>QUERIES / SECOND</th>
<th>AUTHLOGS (AMS.M1)</th>
<th>QUERYLOGS (AMS.M1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noon (UTC)</td>
<td>941.25</td>
<td>10246.66</td>
</tr>
<tr>
<td>Midnight (UTC)</td>
<td>686.75</td>
<td>5525.26</td>
</tr>
</tbody>
</table>

Avalanche Benchmark:
- ~30000 messages per second ⇔ 1 message every 33 microseconds
- 3 times faster than AMS.m1 query logs at peak time
GRAPH-ORIENTED DATA MINING
EXPLORING DATA SEEDS...

- Domain
- URL
- IP
- ASN
- Hash
- Email
- Regex

SEED

- Investigate: Scores, Co-occurrences ...
- Maxmind GeoIP: Country Info, ASNs ...
- VirusTotal: Malware URLs, Vendor Info...
- Shodan: Banner Info ...
- HTTP: HTML Content, Certificate, Links ...
- Custom Models: Grid Computing, ML classification ...
DATA EXPLORATION:
BREADTH FIRST TRAVERSAL
DISTRIBUTED BREADTH FIRST TRAVERSAL
LAMBDA MINING

Functional Graph Exploration

Rule Based / Thresholds / Topology based ...

Subgraph Isomorphism Detection

Profiles for specific use cases

Automated Smart Data Mining
PHISHING DETECTION WITH NLP-RANK
Welcome To Paypal.
To resolve your case, what we need is a Confirmation of Your Account to
Confirm Your Account, Please Log In

Email address

Password

Log In
Details for hpaypal.co.uk

Classifier prediction: benign
OpenDNS Security Graph Score: +54

DNS queries

WHOIS record data

Registrar Name: GoDaddy.com, LLP. [Tag = GODADDY] IANAID: 146
Last retrieved February 9, 2016

Created: February 9, 2016
Updated: February 9, 2016
Expires: February 9, 2017
PURPOSE:

Overview of our NLPRank:

• Fraud detection system using NLP/ML techniques and traffic features to identify domain-squatting/brand spoofing in DNS (a technique commonly used by phishing and APT CnCs).
HUMAN-COMPUTER INTERACTION

Attack psychology:
What kind of links do people typically click on? What are people typically interested in?

Topics of interest:
• $$$, Bank Account/CCs, Financial
• News
• Security/Software updates
• Social Network
Acquiring Data
Filtering
NLP
Output

Input Feed
(DNS/HTTP)
HEURISTIC #1
ASN FILTERING
Acquiring Data  Filtering  NLP  Output

Input Feed (DNS/HTTP)
ASN FILTER + WHITELISTING

- Text processing is resource intensive, initial step to remove as much traffic as possible

- Autonomous System Number == neighborhood/zipcode on the internet

- Domains exhibiting fraudulent behavior hosted on ASNs that are not associated with the company they’re spoofing

- Authlogs come in → Enricher plugin will look up ASN of server IP and add to logs
  • Create mapping of Brand Names to their legitimate ASNs
  • Lookup domains/IPS as they come in
HEURISTIC #2
DEFINING MALICIOUS LANGUAGE OF INTERNET FRAUD
Acquiring Data

Filtering

NLP

Output

Input Feed (DNS/HTTP) → Whitelisting → ASN Filter → Popularity Check
Input Feed (DNS/HTTP) -> Whitelisting -> ASN Filter -> Popularity Check

- Substring Edit-distance/Regex/Custom Dictionary check
BUILDING INTUITIONS

From APT/Phishing data sets extracted stemmed English words:
mail, news, soft, serv, updat, game, online, auto, port, host, free, login, link, secur, micro, support, yahoo

Bigram Collocations:
• Words that often occur with each other in FQDN/URL
• Idea: brandname + ad-action word [.] tld
Phishing Samples

http://wwelllssssfffarrgo.webzdarma.cz.html
http://dandraghicescu.ro/dbox/dpbx/dpbx/
http://school76.irkutsk.ru/language/Wellsfargo/online.htm
http://createcrafts.ph/bankofamerica.com.update.login.in.info/de17792ab89754c6b0a58d767a6985fc/
http://wellsfargoonline.pfwv.com.br/wellsfargo/
http://www.cityroo.com/sarasoa/wellsfargo/wellsfargo-online.php
APT Samples

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

Carbanak (Kasperksy):
• update-java[.]net
• adobe-update[.]net

APT 1 Domains (Mandiant):
• gmailboxes[.]com
• microsoft-update-info[.]com
• firefoxupdate[.]com
NLP ON FQDN

Creating a “malicious language” of FQDNs/URLs derived from lexical features of APT/Phishing data sets

Built corpus of domains similar to examples in previous slide

Create custom dictionaries

• Brandname Dictionary
  Ex. gmail, paypal, yahoo, bankofamerica, wells Fargo

• Custom set of stemmed common malicious words
  Ex. secur, updat, install, verif, etc.

Reason for stemming example: updat -> firefoxupdata[.]com (APT1)

Apply Edit-Distance/Automata Theory on substrings to build spam language
Edit-Distance Overview

- Shortest Path, Dynamic Programming algorithm checking similarity between 2 strings (think spellcheck!)

- How many “edits” it takes to turn one string into the other

- Widely used in many fields, ex. bioinformatics for DNA sequence alignment

**Example Edits:**

- google.com --&gt; g00g1e.com 3 edits
- dropbox.com --&gt; dr0pb0x.com 2 edits
- bankofamerica.com --&gt; bank0fam3r1ca.com 3 edits
HEURISTIC #3
HTML CONTENT MINING
Acquiring Data

Filtering

NLP

Output

Input Feed (DNS/HTTP) → Whitelisting → ASN Filter → Popularity Check

Edit distance/Regex/Custom Dictionary check
Acquiring Data

Input Feed (DNS/HTTP) → Whitelisting → ASN Filter → Popularity Check

Filtering

Counts of words on page → Check for Form Fields on Page → Get Request to URL/Retrieve Content

NLP

TF-IDF

Latent Semantic Analysis

Compare Cosine Similarity To Corpus Docs

Output

Edit distance/Regex/Custom Dictionary check
RECREATING RESEARCHER’S MIND

Typical method for reviewing classifier results:
1. Visit site in Tor browser
2. Researcher processes information on site, looks for clues, gains summary
3. Makes decision whether site is legit/malicious

Specifically for Phishing Sites:

Human-Computer Interaction:
• What makes people fall for this?
  • Site will be near copy of legitimate site

How can we automate this process?
Document similarity algorithms?
“Unsupervised learning is the future. It’s all about the features.”

--Michal Sofka, the Cisco CTA Cognitive Threat Analytics Lead
UNSUPERVISED FOR DETECTION

- Knowledge Discovery Algorithms
- Using topic modeling techniques to gain summary of website
- Great for building recommender systems (i.e. PhishTank)
- Used as features for a classifier
BUILDING CORPUS

Built collection of HTML Content of Phishing pages ex. WellsFargo, Paypal, Amazon, Apple, Bank of America

Only focused on big name brands (for now)

Data collection, although at times tedious, become very intimate with the data

See all kinds of variations of Phishes

90s Paypal vs. 2000s Paypal vs. 2015 Paypal

Christian Mingle Phishing?
BUILDING MODEL: TFIDF

Input: Word Count Vector From Terms in HTML Document (Query), Word Count Matrix over a collection (Corpus)

TF-IDF - Show how important word is to a collection

Balance between: Frequency of Term and Rarity over all documents

Term-Frequency: # of times term t, appears in the document d

- Term Relevance does not increase proportional with term-frequency

Inverse-Document Frequency: the # of documents that contain term t

TFIDF - tf-weight * idf-weight

TFIDF - Increases with number of occurrences within a document, and rarity of term over all documents

Ex. TFIDF – Top 10 TFIDF Scores for Shakespeare’s Macbeth:

$$w_{t,d} = (1 + \log_{10} tf_{t,d}) \times \log_{10}(N/df_t)$$
BUILDING MODEL: LSA/LSI

Input: X, count matrix (or TFIDF), where m (rows) is number of terms, and n (columns) is number of documents

- Pick a value k, which represents the number of topics/concepts/dimensions

Process: Decompose X into 3 matrices, U, S, V(T)

U = m x k matrix, where m = terms, k = concepts

S = k x k diagonal matrix. Elements are amount of variation captured from each concept.

V(transpose) = k x n matrix, where k = concepts, n = documents

\[ X \approx USV^T \]
COSINE DISTANCE BETWEEN VECTORS

Cosine distance between two vectors:

\[ \text{In[1]} := \cosineDistance\{a, b, c\}, \{x, y, z\} \]

\[ \text{Out[1]} = 1 - \frac{ax + by + cz}{\sqrt{Abs[a]^2 + Abs[b]^2 + Abs[c]^2} \sqrt{Abs[x]^2 + Abs[y]^2 + Abs[z]^2}} \]
Acquiring Data

Input Feed (DNS/HTTP) → Whitelisting → ASN Filter → Popularity Check

Filtering

Counts of words on page → Check for Form Fields on Page → Get Request to URL/Retrieve Content

NLP

TF-IDF

Latent Semantic Analysis

Edit distance/Regex/Custom Dictionary check

Output

Compare Cosine Similarity To Corpus Docs
Acquiring Data

Input Feed (DNS/HTTP) → Whitelisting → ASN Filter → Popularity Check

Filtering

Whitelisting → Counts of words on page → Check for Form Fields on Page

NLP

Get Request to URL/Retrieve Content → Edit distance/Regex/Custom Dictionary check

TF-IDF

Counts of words on page → Check for Form Fields on Page → Get Request to URL/Retrieve Content

Latent Semantic Analysis

TF-IDF → Counts of words on page

Top N Similar Documents

Compare Cosine Similarity To Corpus Docs → Top N Similar Documents

Edit distance/Regex/Custom Dictionary check → Block List → Auto-Tag Brand/Topic in Phishtank

Email Daily Results

Build Training Sets, Periodically Retain Corpus/Fetch Legit Brand Sites

OpenDNS Security Labs
RESULTS
AUTO-LABELING BRAND RESULTS:

Sample Output (Document Handle, Document (Cosine) Similarity Score, Brand/FQDN of URL):

Input URL/Query: WellsFargo/fitac.com.tr.html
(61, 0.99899197) WellsFargo/wellsfargo.com.html
(62, 0.99890876) WellsFargo/usam.edu.sv.html
(60, 0.9984659) WellsFargo/school76.irkutsk.ru.html
(59, 0.98146677) WellsFargo/theweddingcollection.gg.html
(63, 0.97453147) WellsFargo/exin.ba.html
K-Fold Cross Validation

600 HTML Documents From Phishtank - 80% (480 Documents), 20% (120 Documents) Split

Run 1:
Accuracy: 0.9941176470588236
Precision: 0.95

Run 2:
Accuracy: 0.9950980392156863
Precision: 0.9583333333333334

Run 3:
Accuracy: 990196078431373
Precision: 0.9916666666666667

Run 4:
Accuracy: 0.9921568627450981
Precision: 0.9333333333333333

Run 5:
Accuracy: 0.9901960784313726
Precision: 0.9166666666666666
DEDICATED SAMPLES
Apple ID

Manage your Apple account

Apple ID

Password

[ ] Remember me
2016-03-30T15:56:03.956Z df0f13dc73e14f59[.]apple-confirmation[.]online 0.9116500020027161 Apple/appleaccount.identity-confirmation.com.html
Login to your account

Email address

Password

Login

Forgot your email address or password?

Open a free account

Payment All-in-one.
Pick a card, any, or a bank account or apply credit with our service. You spend your money wish.

Simple. And usually free.
You can open a free PayPal account and choose method of payment: you pay no commission transactions when you make purchases.
One account. All of Google.

Sign in to continue to Gmail

Email

Password

Sign in

Stay signed in

Need help?

Create an account
### Domains Associated with s.miloshevich@yandex.ru

<table>
<thead>
<tr>
<th>Domain Name</th>
<th>Security Categories</th>
<th>Content Categories</th>
<th>Last Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>alert-login-gmail.com</td>
<td>Malware, Phishing</td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>suporteng.com</td>
<td>Malware</td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>docsautentification.com</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>g00glemail.com</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>googledraive.com</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>googlsupport.com</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>membrana52.com</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>pwdrecover.com</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
</tbody>
</table>

Showing 8 of 8 results
security-appleinc.com  0.9436904191970825,  Apple/www.lcloudid-ds.top.html

DETAILS FOR SECURITY-APPLEINC.COM
This domain is currently in the OpenDNS Security Labs block list
This domain is associated with the following type of threat: Phishing
Classifier prediction: suspicious  OpenDNS Security Graph Score: -83

WHOIS RECORD DATA
Registrar Name: ZhuHai NaiSiNiKe Information Technology Co., Ltd.  IANAID: 1619  Last retrieved February 5, 2016
Created: January 8, 2016  Updated: January 16, 2016  Expires: January 8, 2017

Email Address  Associated Domains  Email Type  Last Observed
trustedmon@gmail.com  11 Total - 8 malicious  Administrative, Registrant,  Current
### DOMAINS ASSOCIATED WITH TRUSTEDMON@GMAIL.COM

<table>
<thead>
<tr>
<th>Domain Name</th>
<th>Security Categories</th>
<th>Content Categories</th>
<th>Last Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>appleidprivacy.com</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>appleinc-security.com</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>appleinc-support.com</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>icloudprivacy.com</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>inc-appleid.info</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>secure-appleinc.com</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>security-apple.com</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>security-appleinc.com</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
</tbody>
</table>

Showing 8 of 8 results
MALICIOUS DOMAINS HOSTED BY 104.207.132.165

gmail-edit.pw  yahoo-maintain.pw  gmail-safety.pw  yahoo-safety.com  gmail-retry.tk

FEATURES

<table>
<thead>
<tr>
<th>Known domains hosted at this IP</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>LD2 domains count</td>
<td>16</td>
</tr>
<tr>
<td>LD3 domains count</td>
<td>20</td>
</tr>
<tr>
<td>LD2-1 domains count</td>
<td>15</td>
</tr>
<tr>
<td>LD2-2 domains count</td>
<td>20</td>
</tr>
<tr>
<td>LD2 domains diversity</td>
<td>0.8</td>
</tr>
<tr>
<td>LD3 domains diversity</td>
<td>1</td>
</tr>
<tr>
<td>LD2-1 domains diversity</td>
<td>0.75</td>
</tr>
<tr>
<td>LD2-2 domains diversity</td>
<td>1</td>
</tr>
</tbody>
</table>

KNOWN DOMAINS HOSTED BY 104.207.132.165

Yahoo奇摩讓你左右逢源，盡如人意。

無與倫比的 Yahoo奇摩電子信箱、重大地方新聞和國內外新聞、財經、運動、音樂和影視等精采內容。探索網上大千世界，一覽人間五光十色。
Google

One account. All of Google.

Click continue to verify

Password

Recovery Email

Phone number

Continue

Stay signed in

Need help?

One Google Account for everything Google
# Domains Associated with sjwdrum@gmail.com

<table>
<thead>
<tr>
<th>Domain Name</th>
<th>Security Categories</th>
<th>Content Categories</th>
<th>Last Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>dhlpackagacentresecuredlink.info</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>googlesecuredsecuritycentre.info</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>mightyhighband.com</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>realquickverifiedserv.info</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>samuelwelch.com</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>secured-google.com</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>simplefmtservice.com</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>welchtuningsystems.com</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>welchtuningsytems.com</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
</tbody>
</table>
Dare I say predictive??

website coming soon!

Please check back soon to see if the site is available.
COMPROMISED SAMPLES
Personal Identification

Email
Email Address

Password
Password

Personal Informations

First Name
Ex: John

Last Name
Ex: Doe

Date of Birth
Day   Month   Year

All in one pay.
Pick a card, any card... or a bank account. It's your money, you choose how to spend it.

Simple. And usually free.
It's free to sign up for a PayPal account, and we don't charge you a transaction fee when you buy something, no matter how you choose to pay.
STÉPHANIE LASSALLE
Photographie

ENTRER
Français
Get your money’s worth from your tax refund
Open a new savings account today and put your return to work for you.

Rates are low
Now is a great time to buy a home or refinance your mortgage.

Online Bill Pay
Pay multiple bills all at once.

Express your thanks. Thank our troops, and we’ll give $1* to help support them. Join us

Stay in the know
Check balances anytime you want—right from your smartphone.

NEW 20,000 online bonus points offer
Help with your home loan payments

Share website feedback
Emportez Outlook avec vous
Vivez une expérience de messagerie optimale sur appareils iOS, Android et Windows.
En savoir plus
Creating something special for you
FALSE POSITIVES
switched to db phishing
Start date: Wed Mar 02 2016 04:00:00 GMT+0000 (UTC)
End date: Wed Mar 02 2016 10:00:00 GMT+0000 (UTC)
Score threshold: 0.87

7779 freshly marked vertices.
27 candidates.

2016-03-02T04:21:00.552Z  icloud[.]account-id[.]com 0.9970257878303528 Apple/appleidverifysupport.com.html
2016-03-02T04:14:22.609Z  facebook[.]online 0.9935976266860962 Facebook/khansports.net.html
2016-03-02T04:01:19.971Z  bankofamerica[.]com 0.9831417202949524 BankOfAmerica/janwill.co.ke.html
2016-03-02T04:07:35.164Z  bankofamerica[.]com 0.9831417202949524 BankOfAmerica/janwill.co.ke.html
2016-03-02T04:07:47.790Z  bankoramerica[.]com 0.9831417202949524 BankOfAmerica/janwill.co.ke.html
2016-03-02T04:07:09.371Z  bankofamerica[.]com 0.9831417202949524 BankOfAmerica/janwill.co.ke.html
2016-03-02T04:27:51.324Z  wellsafargo[.]com 0.9378566145896912 WellsFargo/wellsfargo.com.html
2016-03-02T04:28:46.840Z  wellsafargo[.]com 0.9378566145896912 WellsFargo/wellsfargo.com.html
2016-03-02T04:27:10.762Z  wellsafargo[.]com 0.9376039505004883 WellsFargo/wellsfargo.com.html
2016-03-02T04:27:28.924Z  wellsafargo[.]com 0.9375162720680237 WellsFargo/wellsfargo.com.html
2016-03-02T04:22:43.132Z  wellsafargo[.]com 0.9375162720680237 WellsFargo/wellsfargo.com.html
2016-03-02T04:27:52.661Z  wellsafargo[.]com 0.9375162720680237 WellsFargo/wellsfargo.com.html
2016-03-02T04:22:16.901Z  wellsafargo[.]com 0.9375162720680237 WellsFargo/wellsfargo.com.html
2016-03-02T04:27:50.260Z  wellsafargobank[.]com 0.9375162720680237 WellsFargo/wellsfargo.com.html
2016-03-02T04:28:40.807Z  www/wellsfargo[.]com 0.9373353123664856 WellsFargo/wellsfargo.com.html
2016-03-02T04:29:44.988Z  wellsafgows[.]com 0.9371532797813416 WellsFargo/wellsfargo.com.html
2016-03-02T04:27:52.220Z  wellsafgows[.]com 0.9369886517524719 WellsFargo/wellsfargo.com.html
2016-03-02T04:29:31.825Z  wellsafgows[.]com 0.9369886517524719 WellsFargo/wellsfargo.com.html
2016-03-02T04:21:15.342Z  wellsafargo[.]com 0.936801016330719 WellsFargo/wellsfargo.com.html
2016-03-02T04:26:44.585Z  wellsafargo[.]com 0.9366189241409302 WellsFargo/wellsfargo.com.html
2016-03-02T04:22:22.710Z  seguridad[.]dsal-apple[.]com 0.8931735754013062 Apple/randoserc.com.html
Мы представляем вам уникальный сервис с помощью которого вы сможете быстро и легко заработать
details for play.googles.re

Classifier prediction: benign
OpenDNS Security Graph Score: +100

DNS queries

WHOIS record data

Registrar Name: TLD Registrar Solutions Ltd  IANAID: 1564

Created: June 24, 2014  Updated: June 24, 2015  Expires: June 24, 2016

<table>
<thead>
<tr>
<th>Email Address</th>
<th>Associated Domains</th>
<th>Email Type</th>
<th>Last Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="mailto:topxaker@gmail.com">topxaker@gmail.com</a></td>
<td>7 Total - 1 malicious</td>
<td>Administrative, Registrant</td>
<td>Current</td>
</tr>
<tr>
<td><a href="mailto:AfnicRoleObject@internet.bs">AfnicRoleObject@internet.bs</a></td>
<td>9 Total - 3 malicious</td>
<td>Technical</td>
<td>Current</td>
</tr>
</tbody>
</table>
### Domains Associated with topxaker@gmail.com

<table>
<thead>
<tr>
<th>Domain Name</th>
<th>Security Categories</th>
<th>Content Categories</th>
<th>Last Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>rr3.us</td>
<td>Drive-by Downloads/Exploits</td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>apple.us</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>apples.re</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>google.re</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>miniopera.us</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>rovio.us</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>skyperu.us</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>swizard.us</td>
<td></td>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>Domain/URL</td>
<td>Score</td>
<td>Alternative Domain/URL</td>
<td>Score</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>-------</td>
<td>-----------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>pyplusers.com</td>
<td>0.9960060715675354</td>
<td>Paypal/bluespruceus.com.html</td>
<td></td>
</tr>
<tr>
<td>googledrave.com</td>
<td>0.9555937051773071</td>
<td>Google/www.rotisseriebuongusto.com.br</td>
<td></td>
</tr>
<tr>
<td>idmappleinc.info</td>
<td>0.9283350110054016</td>
<td>Apple/www.lcloudid-ds.top.html</td>
<td></td>
</tr>
<tr>
<td>supportapples.com</td>
<td>0.9700624942779541</td>
<td>Apple/santuariosdocorporo.com.html</td>
<td></td>
</tr>
<tr>
<td>wellllsfargoservice.goldenhomes.co.za</td>
<td>0.9379003047943115</td>
<td>WellsFargo/wellsfargo.com.html</td>
<td></td>
</tr>
<tr>
<td>facebooksytfa.tr.gg</td>
<td>0.9258965253829956</td>
<td>Facebook/facebook.com.html</td>
<td></td>
</tr>
<tr>
<td>googlecomments.netne.net</td>
<td>0.9551351070404053</td>
<td>Facebook/facebook.com.html</td>
<td></td>
</tr>
<tr>
<td>google-gems.pe.hu</td>
<td>0.9421336650848389</td>
<td>Google/gmailstealer.tk.html</td>
<td></td>
</tr>
<tr>
<td>isdmaapple-step2.com</td>
<td>0.9403073787689209</td>
<td>Apple/carsdrive.com.ar.html</td>
<td></td>
</tr>
<tr>
<td>lindungifacebookmu.tk</td>
<td>0.9336977601051331</td>
<td>Facebook/facebook.com.html</td>
<td></td>
</tr>
<tr>
<td>atfacebooktt451287656709889008723457667985983465309890.notiexpreesht.com</td>
<td>0.9466191530227661</td>
<td>Facebook.com.html</td>
<td></td>
</tr>
<tr>
<td><a href="http://www.ifacebook.pl">www.ifacebook.pl</a></td>
<td>0.9442495107650757</td>
<td>Facebook/facebook.com.html</td>
<td></td>
</tr>
<tr>
<td>apple.idevicetrackers.com</td>
<td>0.9769512414932251</td>
<td>Apple/www.lcloudid-ds.top.html</td>
<td></td>
</tr>
<tr>
<td>mobile-bankofamerica.com.ssnverify.</td>
<td>0.9848669171333313</td>
<td>BankOfAmerica/janwill.co.ke.html</td>
<td></td>
</tr>
<tr>
<td>mobile-bankofamerica.com.ssnverify.user</td>
<td>0.9848620295524597</td>
<td>BankOfAmerica/janwill.co.ke.html</td>
<td></td>
</tr>
<tr>
<td>mobile-bankofamerica.com.ssnverify.usere</td>
<td>0.9848669171333313</td>
<td>BankOfAmerica/janwill.co.ke.html</td>
<td></td>
</tr>
<tr>
<td>free-facebook.gq</td>
<td>0.9977947473526001</td>
<td>Facebook/khansports.net.html</td>
<td></td>
</tr>
<tr>
<td>transaction.paypalcom.verification.global.kimora-fashion.com</td>
<td>0.9959402084350586</td>
<td>Paypal/dinastiaaccessorios.com.br.html</td>
<td></td>
</tr>
<tr>
<td>bankofamerica.com.ssnid6530.us</td>
<td>0.9848620295524597</td>
<td>BankOfAmerica/janwill.co.ke.html</td>
<td></td>
</tr>
<tr>
<td>fac3book2.hostingsiteforfree.com</td>
<td>0.9887204170227051</td>
<td>Facebook/fcebo00k.com.html</td>
<td></td>
</tr>
<tr>
<td>ymfacebook.ml</td>
<td>0.9958401918411255</td>
<td>Facebook/khansports.net.html</td>
<td></td>
</tr>
</tbody>
</table>
STACKING MODELS
DETAILS FOR WWW.ICLOUD-SAFRI.COM

This domain is currently in the OpenDNS Security Labs block list

Classifier prediction: suspicious
OpenDNS Security Graph Score: 100

WHOIS RECORD DATA

Registrar Name: SHANGHAI MEICHENG TECHNOLOGY INFORMATION DEVELOPMENT CO., LTD. IANAID: 1621

Created: October 15, 2015
Updated: October 15, 2015
Expires: October 15, 2016

Email Address: kelvindai1@qq.com
Associated Domains: Greater than 500 Total
Email Type: Administrative, Registrant, Technical
Last Observed: Current
COMBINING DETECTION MODELS

PHISHING, SPIKING, AND BAD HOSTING

SEPTEMBER 14, 2015
BY DHIA MAHJOUB, JEREMIAH O’CONNOR, THIBAULT REUILLE AND THOMAS MATHEW

At OpenDNS Labs we have developed a number of predictive models to hunt down evil on the Internet. We have discussed in previous blogs and conferences our algorithms NLPRank [1][2][3], Spike detector [4][5][6], and malicious IP space/rogue host detectors [7][8](section 14)[9][10][11][12][13][14][15].

In this blog we will discuss how we integrate all of these detection models to improve detection coverage of current threats and walk through a few interesting examples.

PHISHING AND SPIKES

One of the recent samples we have found was a Facebook phishing campaign that was surfaced by our real-time alert system. Our model NLPRank detected the campaign of Facebook phishing sites spoofing Facebook under the second-level domain (2LD) 2nso3s[.]com.

For this particular domain, when visiting the 2LD, 2nso3s[.]com from your browser, you would be directed to a URL that looks like:


As we can see in the path of the URL the next page routes you directly to the legitimate facebook[.]com after they have stolen the entered credentials. We also cross referenced this domain with our crowd-sourced system Phishtank, and found someone from the community submitted one of these hostnames.
OTHER CLUES:

HTTrack – tools used to clone site

```html
<!DOCTYPE HTML><html lang="">

<!-- Mirrored from tools.google.com/dlpage/drive/index.html by HTTrack Website Copier/3.x [XR&CO'2014], Tue, 23 Sep 2014 08:58:40 GMT -->

<!-- Added by HTTrack --><meta http-equiv="content-type" content="text/html;charset=utf-8" /><!-- /Added by HTTrack -->

<head><script type="text/javascript">

function utmx_section(){function utmx(){}
```
INTERESTING RESULTS

Carbanak (banking trojan) came out in February:

2015-01-23 14:52:58 -- a96e74b8-b052-4f42-a517-d7273d4f13e7

NLPRank High-Risk Results (FQDNs)

cdneu.windows8downloadscdn.com
update-java.net
**INTERESTING RESULTS**

symantecupdates.com

<table>
<thead>
<tr>
<th><strong>Whois information</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Registration date</strong></td>
</tr>
<tr>
<td><strong>Registrar name</strong></td>
</tr>
<tr>
<td><strong>Registrant</strong></td>
</tr>
<tr>
<td><strong>Registrant contact address</strong></td>
</tr>
</tbody>
</table>
21,533 DOMAINS???

crowcasinovip.biz mybestbrand.biz mybestbrands.biz huarencheluewangzhi.com icbczay.com boyinbocai5.com
haoyunc3.com bocaiwangzhenqianpingtai.com zuqiuobocaiwangzhan7.com weinisirenyulecheng94.com
xinquanxunwang244.com dfjdh.com yaojiyulecheng9898.com wanbaoluyulecheng94.com xinpujingyule15.com
toabao.com jinbaiyiylulecheng26.com toubakd.com tiantianleyulecheng61.com wangziyulecheng33.com
yezonghuiyulecheng82.com bocwry.com huangguantouzhuwangzhanwangzhi86.com huangguanwangquaomen29.com
haiwangxingylc1664.com yinghuangylc727.com bocaiasd.com changjianggjylc.com jinmaylcoiu.com
yazhougjylc.com huangguanxin2wang32.com benchixsyl.com zhucapiaiosongcaijin.com ceoylcdf.com
zhucesigncaijinwendzhan62.com aomenduchangyouxiyounaxie30.com mengtekaluoylcb.com
guojihuangguanyule40.com huangguantiyupingtai93.com huangguanjianxinwangxinyu37.com
aomenduchangpaixing27.com 500wanylycyu.com dajihuiylc686.com ruifengguojiyy.com makeboluoylcb.com
jincaigjylc.com xindongfangylc869.com aomenduchangzainali50.com wangshangyulekaihusongcaijin.com
huangguanjxwhk.com jinbangylc77.com baijialeqo.com yataigjylc.com baishenggjylcwe.com bocaiqionsgqe.com
wufagjylc.com moerbenylckk.com bogouylcl663.com huangguanjilidwanzhi23.com bojueylcpo.com
bocaiwangzhanq.com taoataao.com bbhunas.com sjzd36.com sjpt63.com bjlk33.com
bajialebishengtouzhuiqiao20.com xijialiansaijifenbang57.com bajialeyule86.com xijiapaiming46.com
aomenbajialechanging76.com bajialeyulepingtai34.com wangshangbajialekaihusongcaijin76.com
ouzhouwudaliainspaiming53.com wudaliainsitedian39.com bajialekaihusong50caijin17.com bajialeguize52.com
zhibobazuqiuqinhou200.com zuqiuobifenqiaonan88.com dejiasaichengbiao88.com zuqiuba85.com mahuitqzzejw83.com
sjzd01.com weixingjianqiang29.com cwanpp.com xingboyulezaixian86.com mwqaipah.com
jiankongpingtairuanjian43.com zhenqianyulechengguangwanz3.com nijdyytyj.com fanheer.com 999coin.com
shengann74.com jackwolfskinsalejp.com zaozhanguq.com bj17788.com ruhejiankongshouji2.com
aomenduchangyingqianliao75.com shoujingweichaxunsuanjian12.com shoujijiantingshebei46.com aomen916.com
shoujikajiatingqi77.com zhenqianyouxipaixing2.com rysevw.com wanzhengqianwangzhan36.com vrcgw.com
fellvinshengannayylecheng20.com duchangyingqianmijue81.com zzvqo.com
wangshangzhenqianzhajinhuyouxi62.com zql2014.com bocwtr.com baijiyulezaixian87.com pacyfjg.com

OpenDNS
SMH
<table>
<thead>
<tr>
<th>Domain Name:</th>
<th>TopSec.COM 107</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registration ID:</td>
<td>1857525015_DOMAIN.COM-VRSN</td>
</tr>
<tr>
<td>Registrar:</td>
<td>Go Daddy, LLC</td>
</tr>
<tr>
<td>Registrar IANA ID:</td>
<td>146</td>
</tr>
<tr>
<td>Registrar Abuse Contact:</td>
<td><a href="mailto:abuse@godaddy.com">abuse@godaddy.com</a></td>
</tr>
<tr>
<td>Registrar Abuse Contact Email:</td>
<td><a href="mailto:abuse@godaddy.com">abuse@godaddy.com</a></td>
</tr>
<tr>
<td>Registrar Abuse Contact Phone:</td>
<td>+1.480-624-2505</td>
</tr>
<tr>
<td>Domain Status:</td>
<td>ok</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Domain Name:</th>
<th>TopSec.COM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registration ID:</td>
<td>1857525015_DOMAIN.COM-VRSN</td>
</tr>
<tr>
<td>Registrar:</td>
<td>Go Daddy, LLC</td>
</tr>
<tr>
<td>Registrar IANA ID:</td>
<td>146</td>
</tr>
<tr>
<td>Registrar Abuse Contact:</td>
<td><a href="mailto:abuse@godaddy.com">abuse@godaddy.com</a></td>
</tr>
<tr>
<td>Registrar Abuse Contact Email:</td>
<td><a href="mailto:abuse@godaddy.com">abuse@godaddy.com</a></td>
</tr>
<tr>
<td>Registrar Abuse Contact Phone:</td>
<td>+1.480-624-2505</td>
</tr>
<tr>
<td>Domain Status:</td>
<td>ok</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Domain Name:</th>
<th>TopSec.COM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registration ID:</td>
<td>1857525015_DOMAIN.COM-VRSN</td>
</tr>
<tr>
<td>Registrar:</td>
<td>Go Daddy, LLC</td>
</tr>
<tr>
<td>Registrar IANA ID:</td>
<td>146</td>
</tr>
<tr>
<td>Registrar Abuse Contact:</td>
<td><a href="mailto:abuse@godaddy.com">abuse@godaddy.com</a></td>
</tr>
<tr>
<td>Registrar Abuse Contact Email:</td>
<td><a href="mailto:abuse@godaddy.com">abuse@godaddy.com</a></td>
</tr>
<tr>
<td>Registrar Abuse Contact Phone:</td>
<td>+1.480-624-2505</td>
</tr>
<tr>
<td>Domain Status:</td>
<td>ok</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Domain Name:</th>
<th>TopSec.COM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registration ID:</td>
<td>1857525015_DOMAIN.COM-VRSN</td>
</tr>
<tr>
<td>Registrar:</td>
<td>Go Daddy, LLC</td>
</tr>
<tr>
<td>Registrar IANA ID:</td>
<td>146</td>
</tr>
<tr>
<td>Registrar Abuse Contact:</td>
<td><a href="mailto:abuse@godaddy.com">abuse@godaddy.com</a></td>
</tr>
<tr>
<td>Registrar Abuse Contact Email:</td>
<td><a href="mailto:abuse@godaddy.com">abuse@godaddy.com</a></td>
</tr>
<tr>
<td>Registrar Abuse Contact Phone:</td>
<td>+1.480-624-2505</td>
</tr>
<tr>
<td>Domain Status:</td>
<td>ok</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Domain Name:</th>
<th>TopSec.COM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registration ID:</td>
<td>1857525015_DOMAIN.COM-VRSN</td>
</tr>
<tr>
<td>Registrar:</td>
<td>Go Daddy, LLC</td>
</tr>
<tr>
<td>Registrar IANA ID:</td>
<td>146</td>
</tr>
<tr>
<td>Registrar Abuse Contact:</td>
<td><a href="mailto:abuse@godaddy.com">abuse@godaddy.com</a></td>
</tr>
<tr>
<td>Registrar Abuse Contact Email:</td>
<td><a href="mailto:abuse@godaddy.com">abuse@godaddy.com</a></td>
</tr>
<tr>
<td>Registrar Abuse Contact Phone:</td>
<td>+1.480-624-2505</td>
</tr>
<tr>
<td>Domain Status:</td>
<td>ok</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Domain Name:</th>
<th>TopSec.COM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registration ID:</td>
<td>1857525015_DOMAIN.COM-VRSN</td>
</tr>
<tr>
<td>Registrar:</td>
<td>Go Daddy, LLC</td>
</tr>
<tr>
<td>Registrar IANA ID:</td>
<td>146</td>
</tr>
<tr>
<td>Registrar Abuse Contact:</td>
<td><a href="mailto:abuse@godaddy.com">abuse@godaddy.com</a></td>
</tr>
<tr>
<td>Registrar Abuse Contact Email:</td>
<td><a href="mailto:abuse@godaddy.com">abuse@godaddy.com</a></td>
</tr>
<tr>
<td>Registrar Abuse Contact Phone:</td>
<td>+1.480-624-2505</td>
</tr>
<tr>
<td>Domain Status:</td>
<td>ok</td>
</tr>
</tbody>
</table>
Future Work

- Reduce validation time on PhishTank → push to community
- Integrate with Cisco Proxy/ HTTP Logs/ Email Corpus
- Building, Testing, Tuning, Iterating
- Data Collection/Building Corpus/More brands
- Detecting Targeted Attacks
THE AVALANCHE PROJECT: WHEN HIGH FREQUENCY TRADING MEETS TRAFFIC CLASSIFICATION

NOVEMBER 5, 2015
BY THIBAULT REUILLE

One of the key challenges for OpenDNS (now part of Cisco) is handling a massive amount of DNS queries and simultaneously running classification models on them as fast as possible. Today, we’re going to talk about Avalanche, a real-time data processing framework currently used in our research cluster.
Special Thanks to:

- OpenDNS Analyst Team, especially Vinny LaRiza and Artsiom Holub

- OpenDNS Marketing Team, Owen Lystrup, Lynne Cox, Kara Drapala (former), Stephen Lynch (former)

- BlackHat Staff
QUESTIONS?

@jmoconnor415
jeremiah@opendns.com
joeoconno@cisco.com

@ThibaultReuille
thibault@opendns.com
treuille@cisco.com