THE SECURITY WOLF OF WALL STREET: FIGHTING CRIME WITH HIGH-FREQUENCY CLASSIFICATION && NLP

Jeremiah O'Connor

Thibault Reuille

OpenDNS

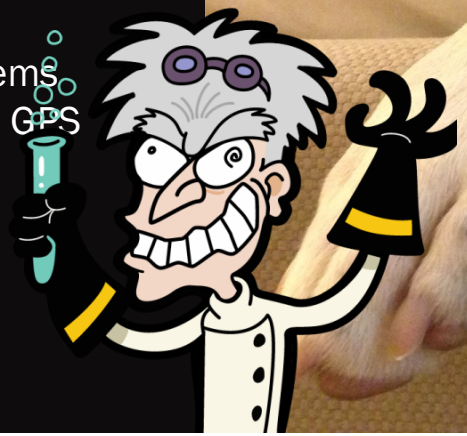
UBM

$ 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



PRESENTATION AGENDA

- INTRODUCTION
- THE AVALANCHE PROJECT & THE RESEARCH PIPELINE
- GRAPH ORIENTED DATA MINING
- FRAUD DETECTION WITH NLP-RANK
- CONCLUSION


INTRODUCTION

OpenDNS Security Labs

Data Science && Network Security

Big Security Data

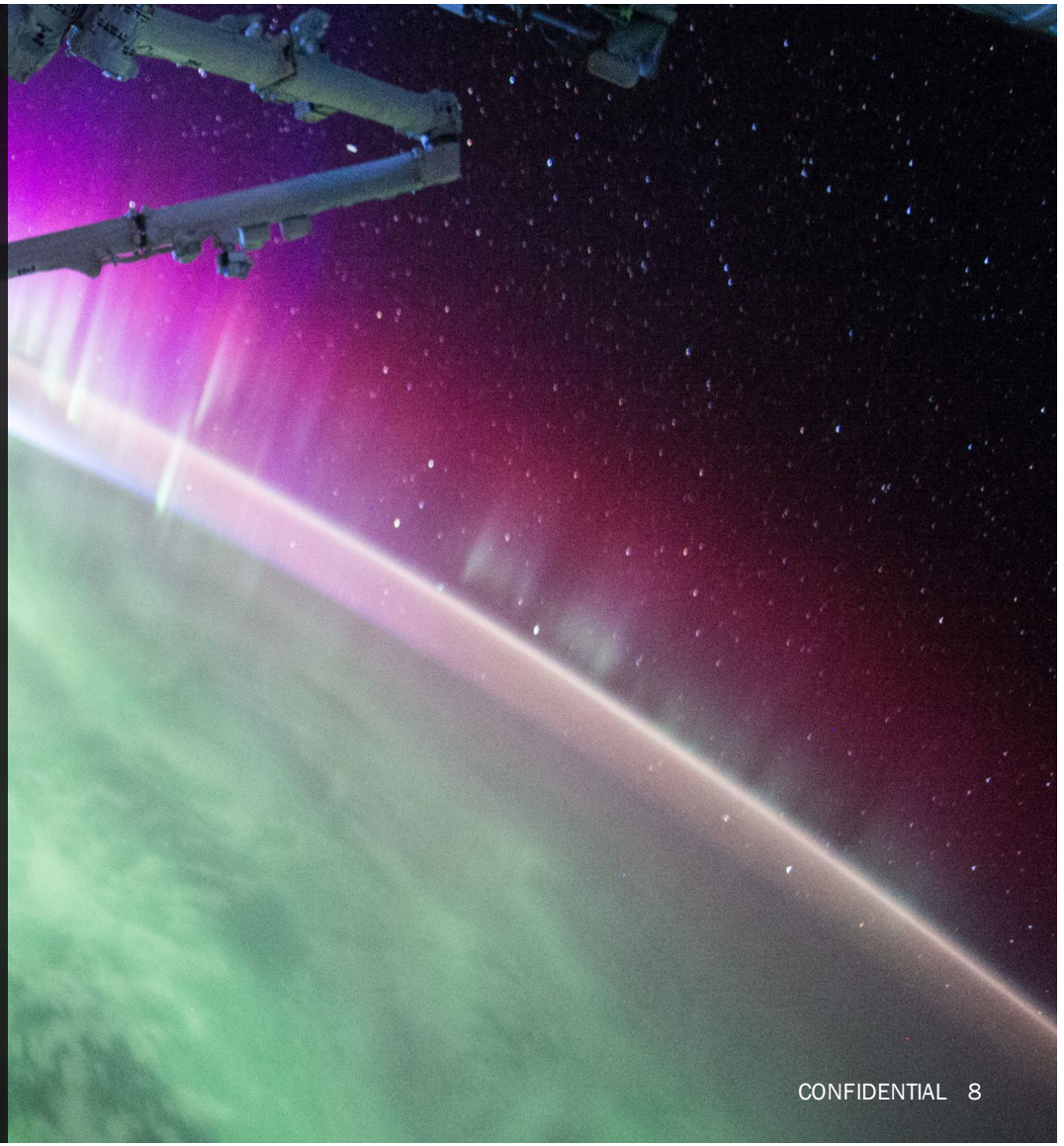
DNS Traffic:
- ~80B DNS requests per day

HTTP Traffic:
- ~10.1M requests per day

Daily Tasks:
- Detection Algorithms, Security Data Analysis,
- Distributed Systems, Big Data Engineering, Data Viz

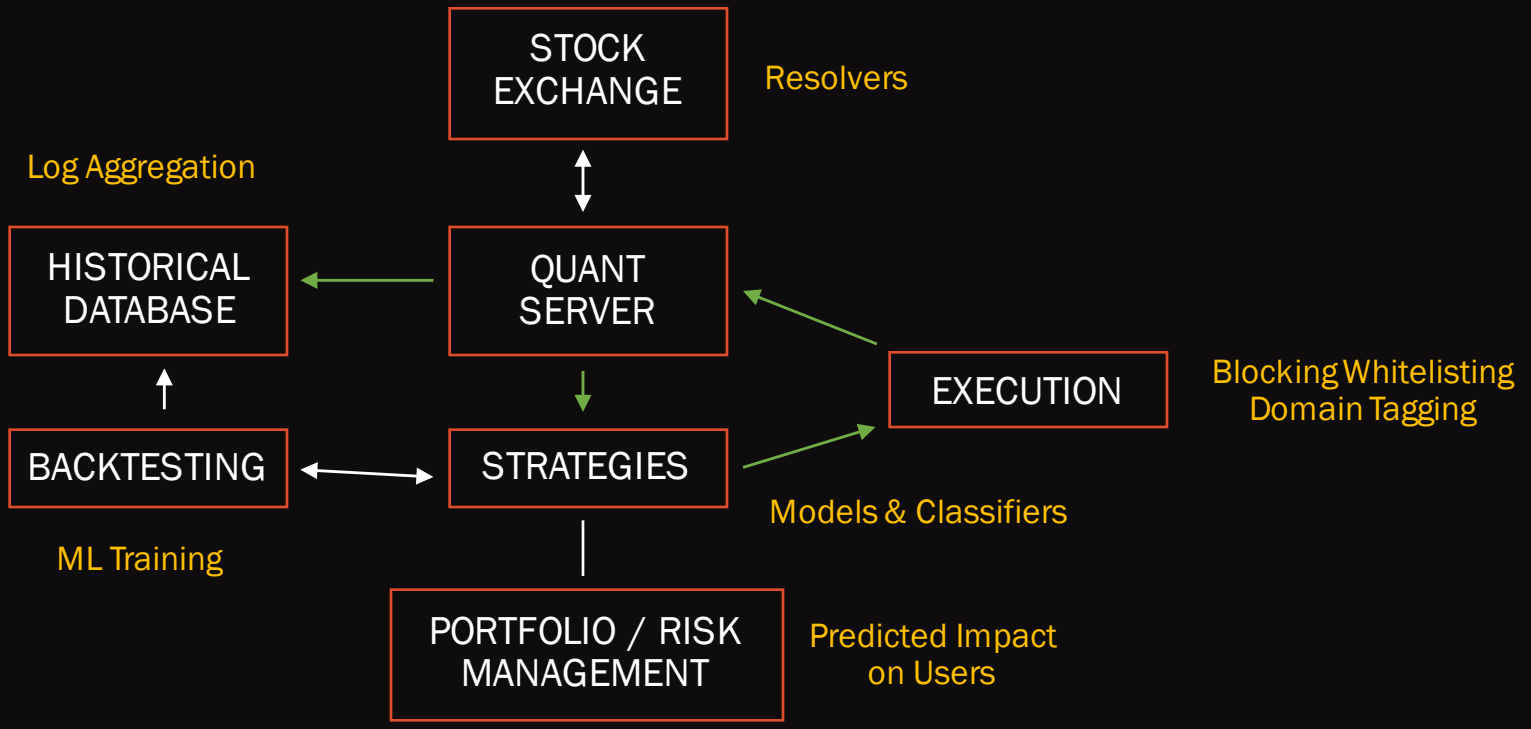
REAL-TIME!

THE AVALANCHE PROJECT



HIGH FREQUENCY TRADING vs TRAFFIC CLASSIFICATION

THE WOLF OF WALL STREET



WHAT IS AVALANCHE?

OVERVIEW AND TECHNICAL DETAILS

Open source project:

- http://github.com/ThibaultReuille/avalanche

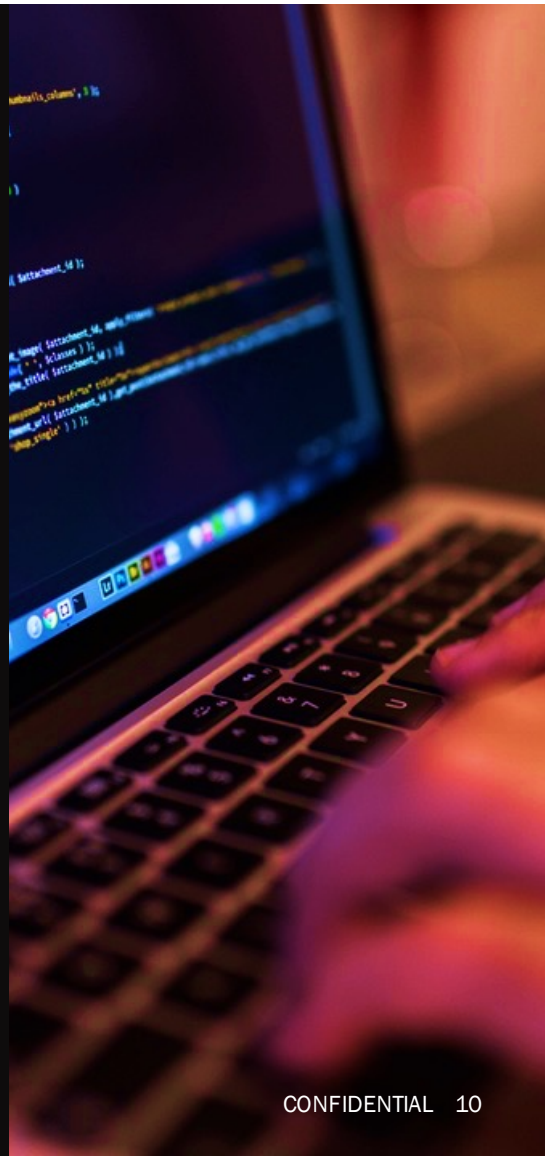
"Real-time" data processing framework

Modular, parallel and distributed design

Written with Python and ZeroMQ

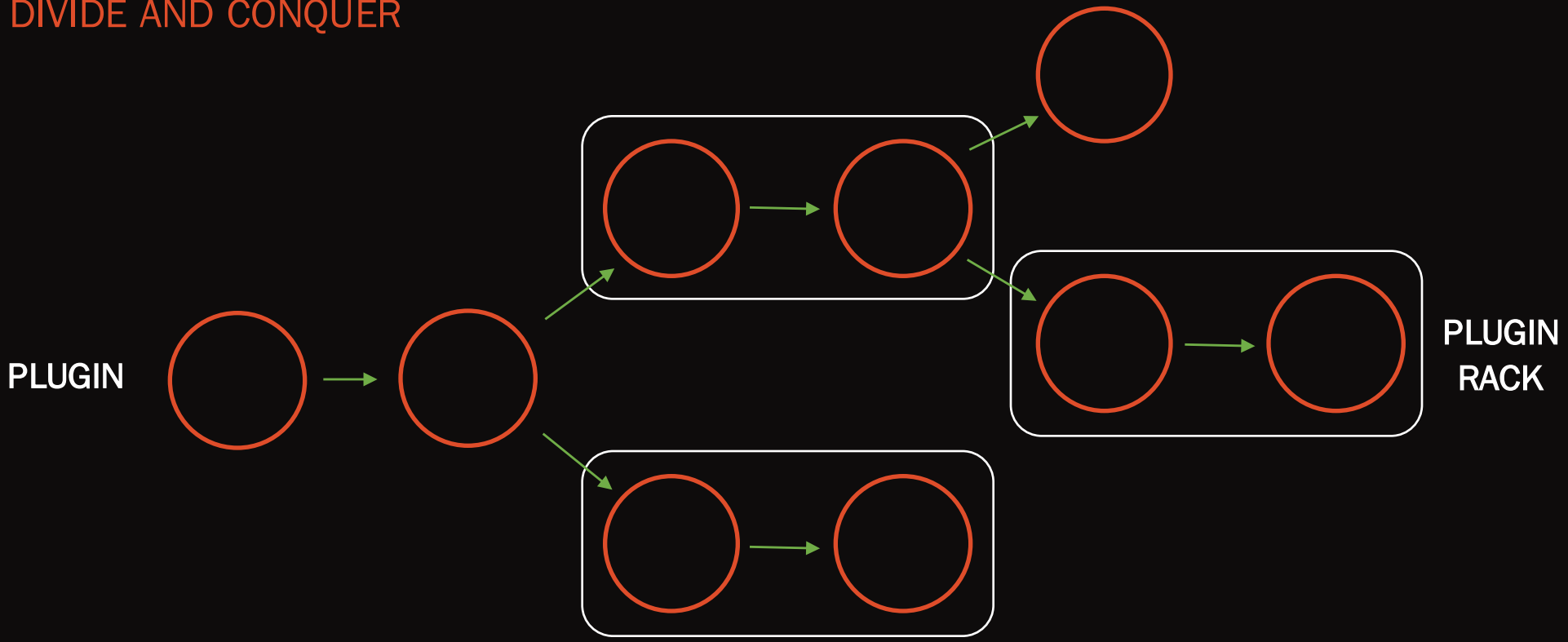
Platform for RT OpenDNS models (Private):

- NLP-Rank
- DNS Tunnelling
- DGA classifier
- Others...



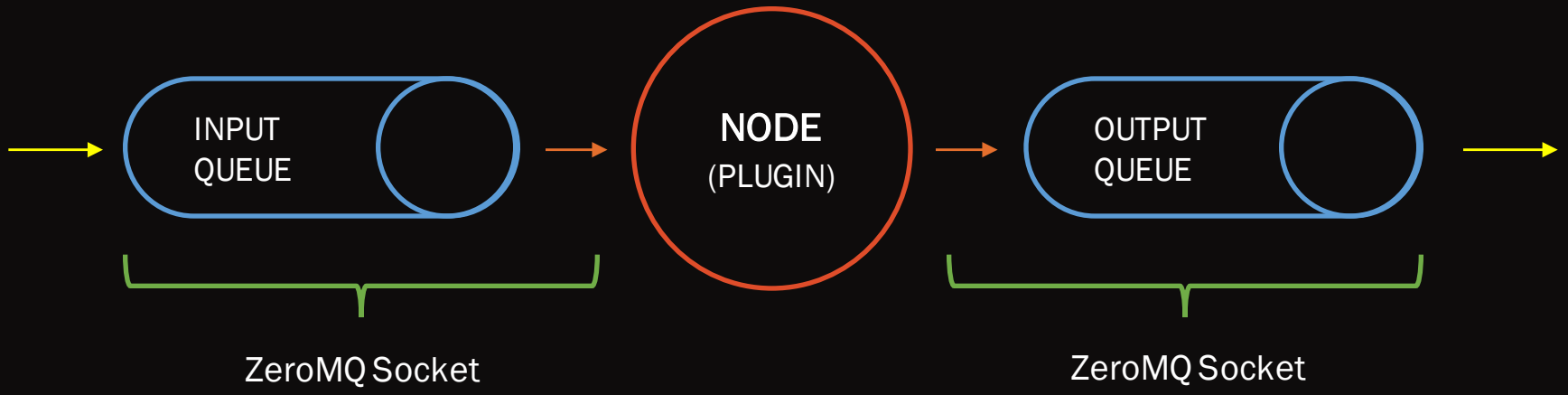
AVALANCHE PIPELINE

DIVIDE AND CONQUER



AVALANCHE DESIGN

DIVIDE AND CONQUER



AVALANCHE NODE

PLUGIN TEMPLATE CODE

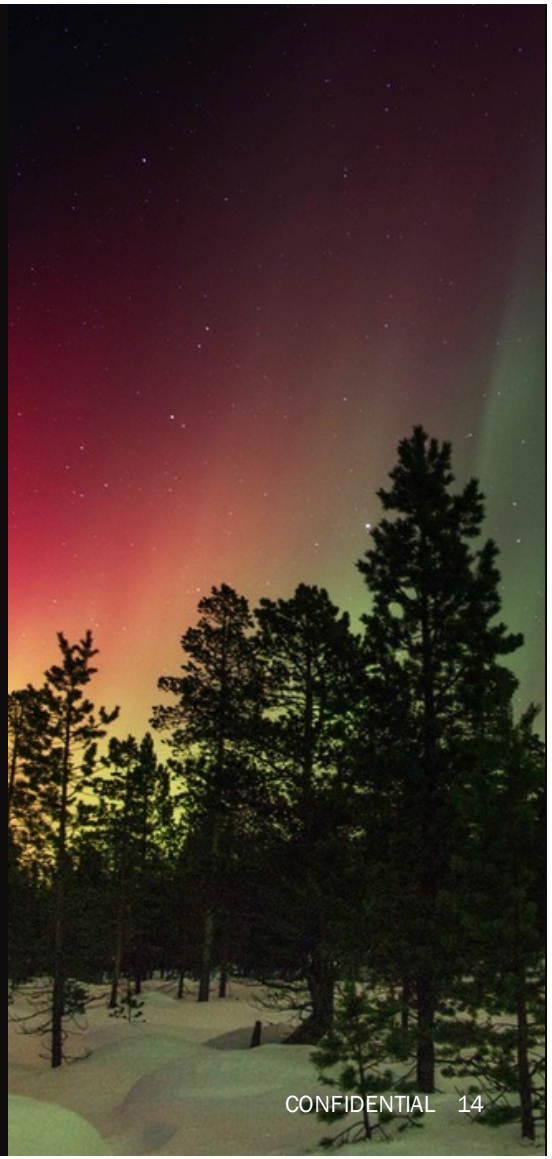
```python
import json
import plugins.base

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 field, remove, etc.
        # Retuning 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 being 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

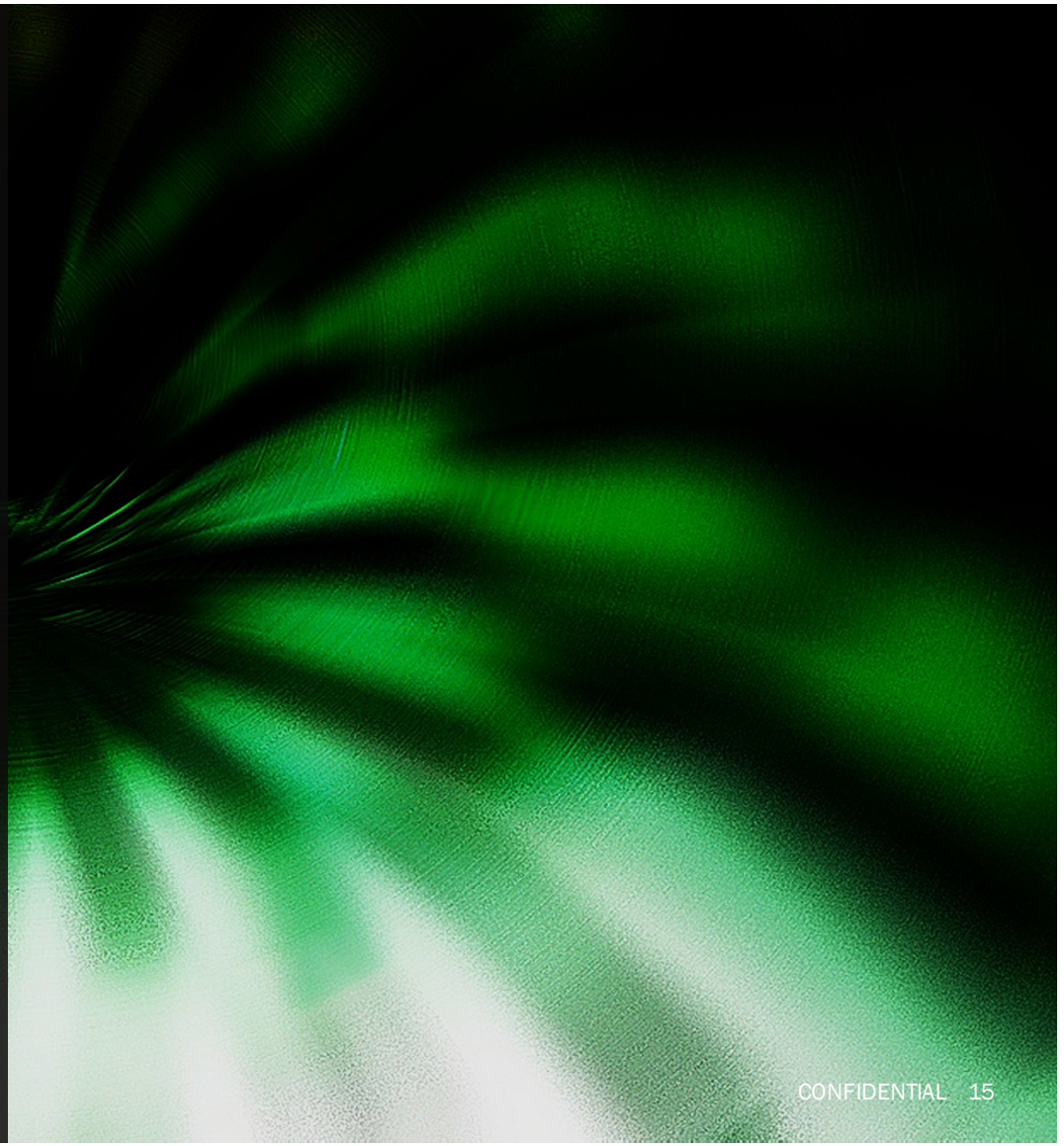
```
$ ./avalanche.py path/to/my_pipeline.json 10000
```

Things you get for free:

- Modularity
- Multi-Threading
- A library of plugins ready-to-use
- Reusability & collaboration
- An insanely fast messaging system

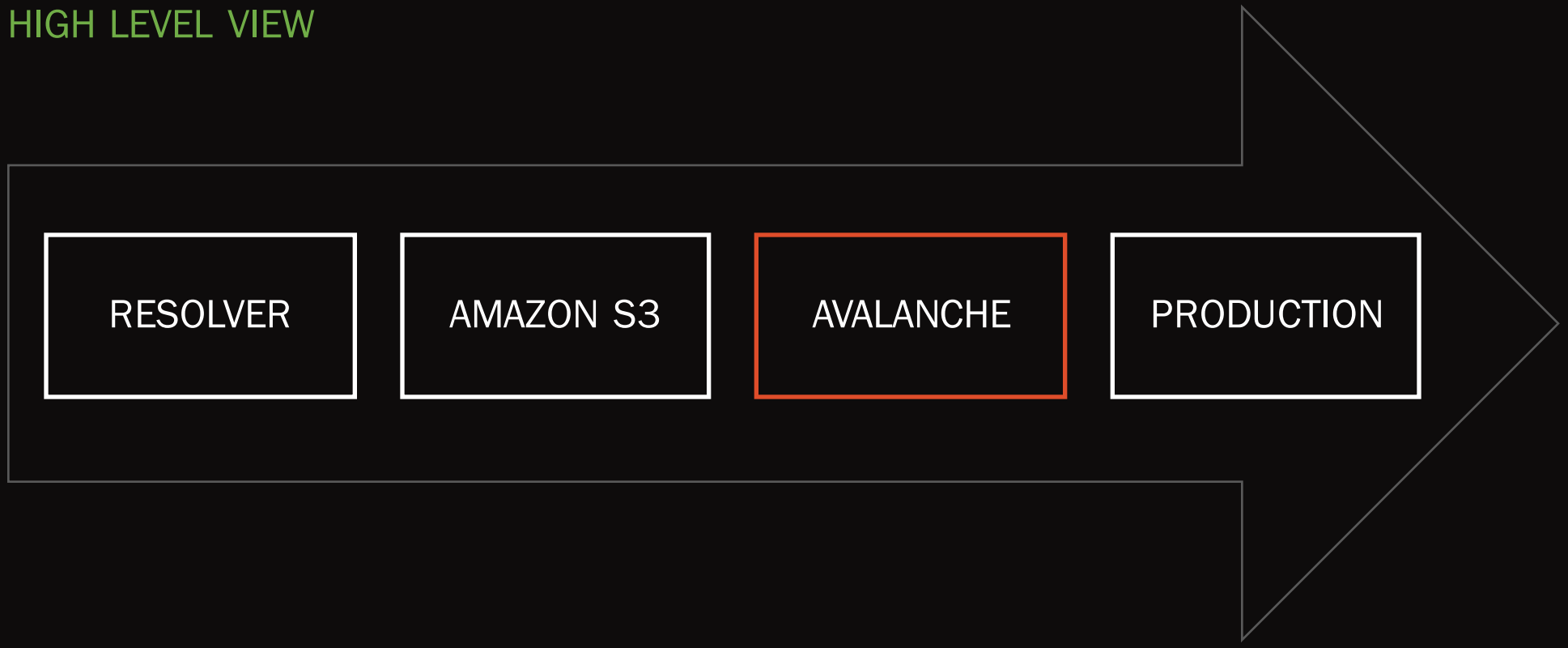


THE RESEARCH PIPELINE



AVALANCHE CLUSTER

HIGH LEVEL VIEW



AVALANCHE CLUSTER

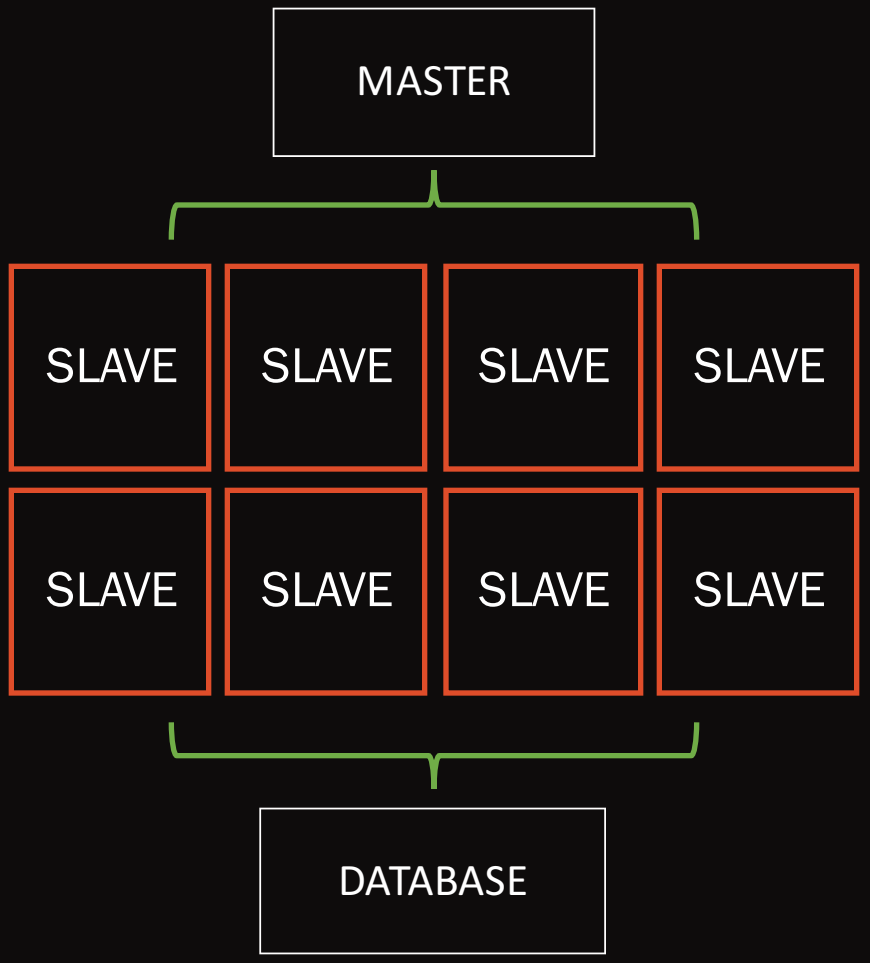
8 Amazon instances

Master distributes work

- Round-robin
- “Fire and forget”

Slaves process the chunks

Results are stored in central database

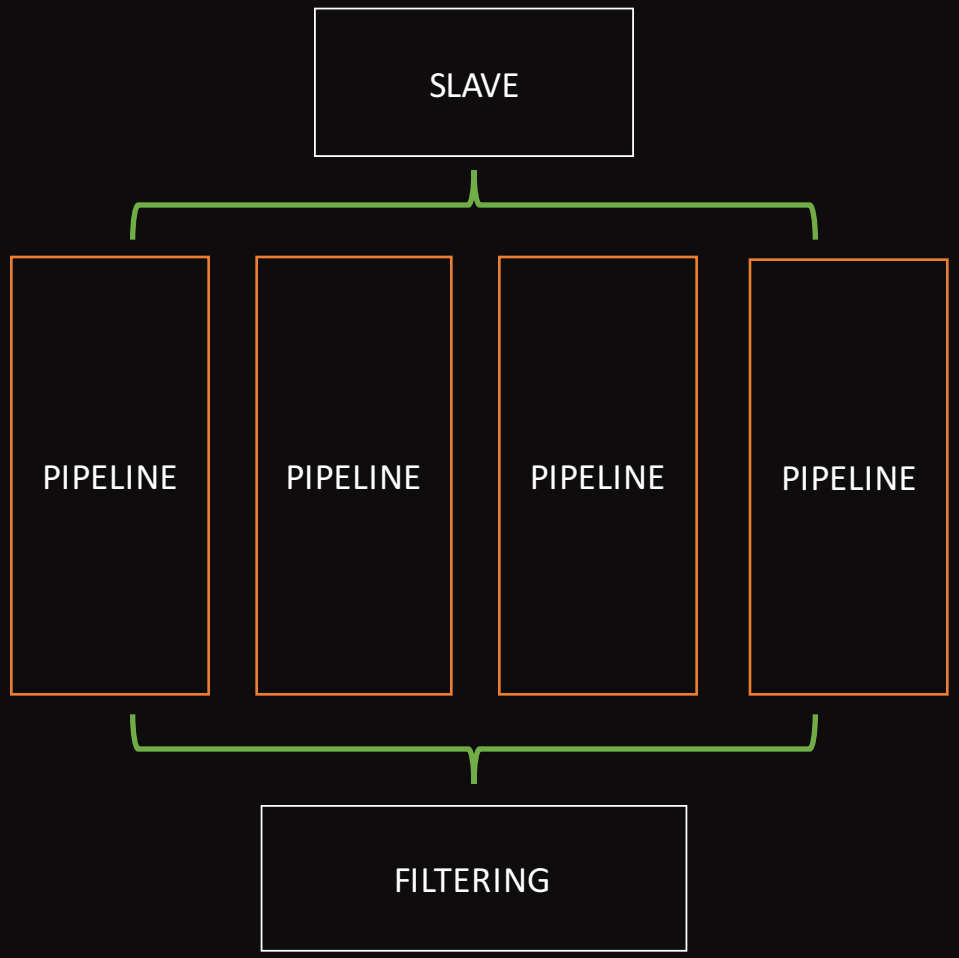


AVALANCHE SLAVE

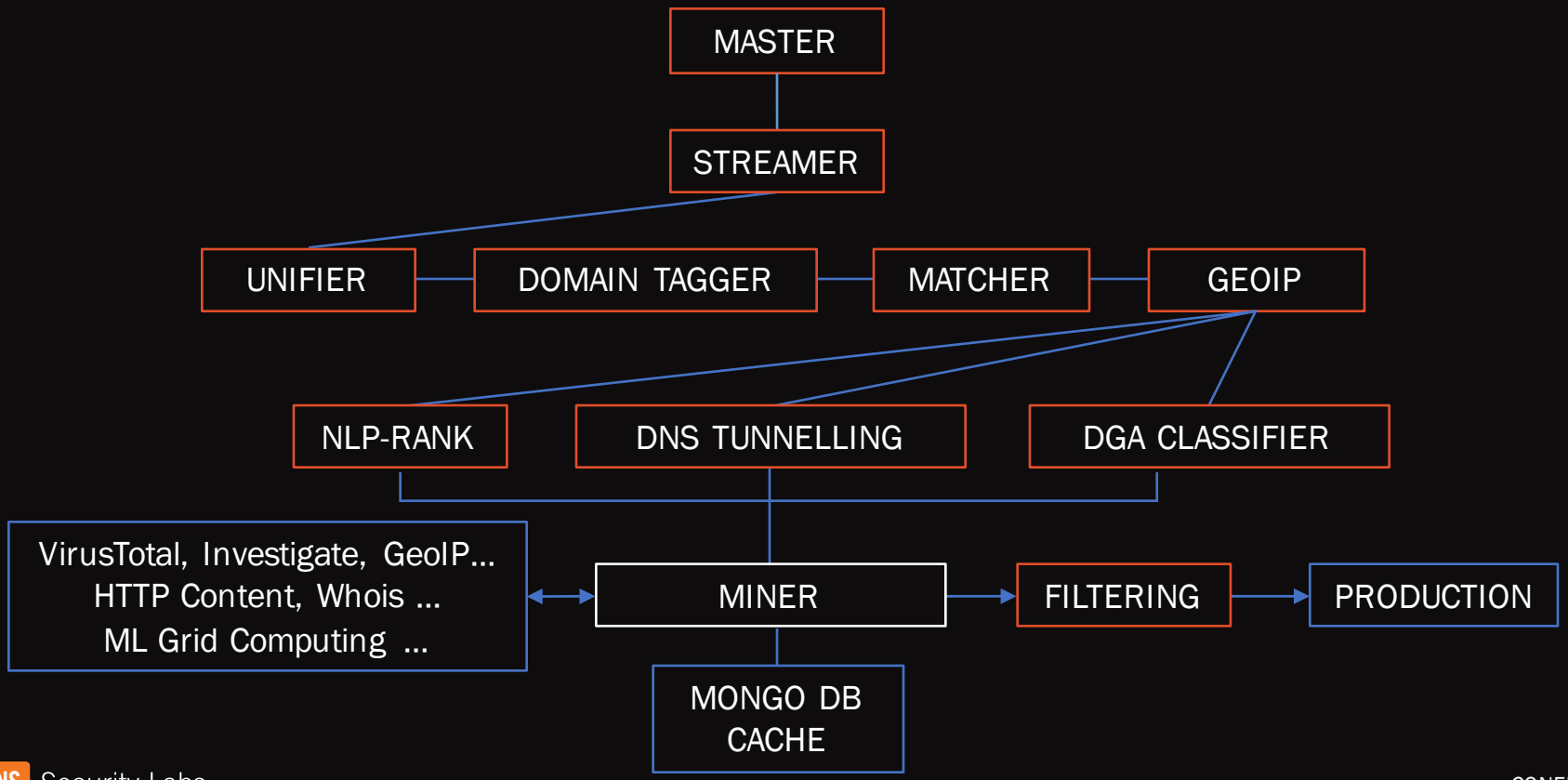
4 Avalanche Pipeline (Processes)

Monitored with DaemonTools

Results are cross-checked and filtered



PIPELINE (Simplified)



TRAFFIC SPEED vs AVALANCHE PIPELINE

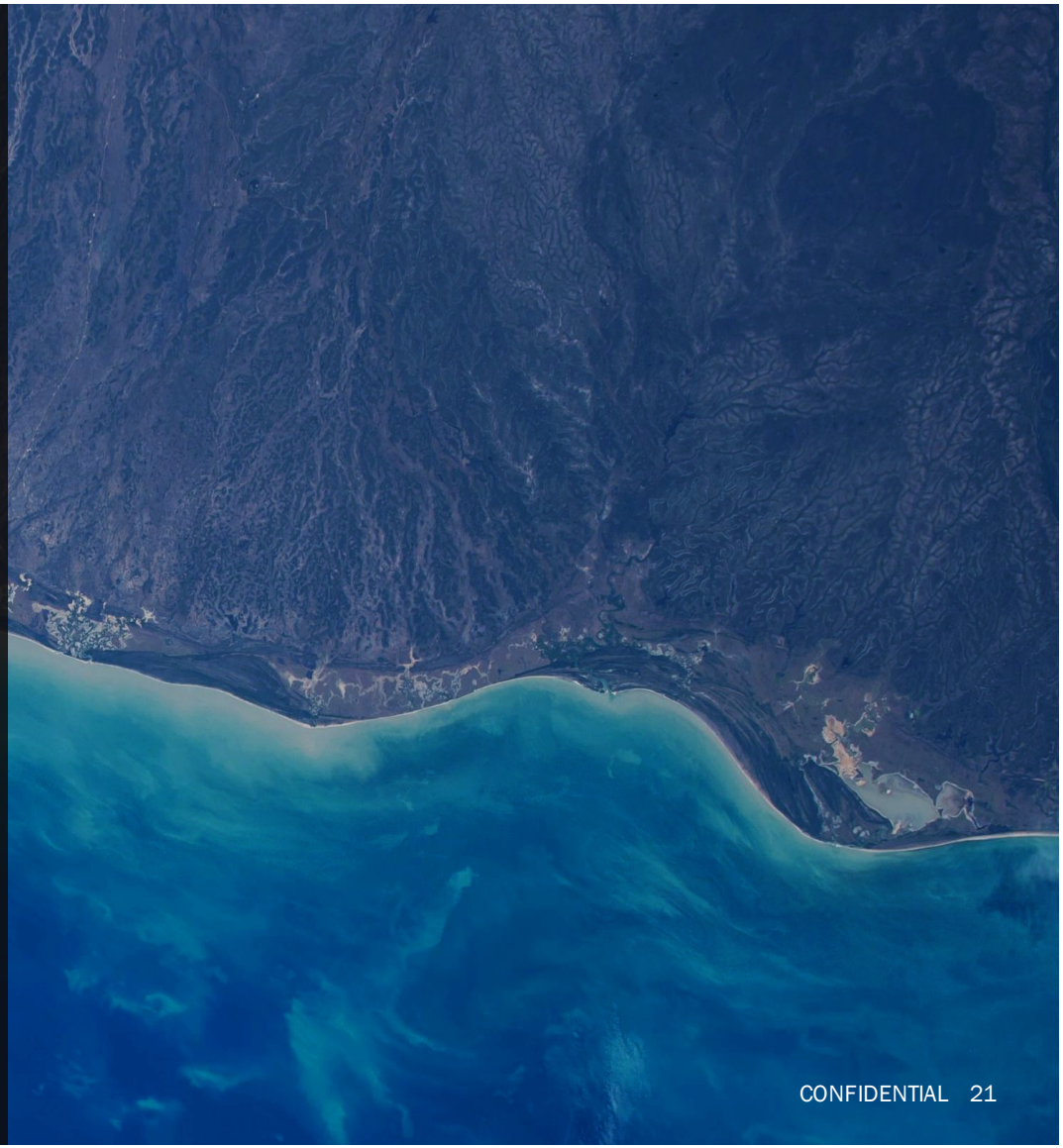
NUMBERS DON'T LIE

QUERIES / CHUNK	AUTHLOGS (AMS.M1)	QUERYLOGS (AMS.M1)
Noon (UTC)	564 752	6 147 997
Midnight (UTC)	412 050	3 315 157
QUERIES / SECOND	**AUTHLOGS (AMS.M1)**	**QUERYLOGS (AMS.M1)**
Noon (UTC)	941.25	10246.66
Midnight (UTC)	686.75	5525.26

Avalanche Benchmark:

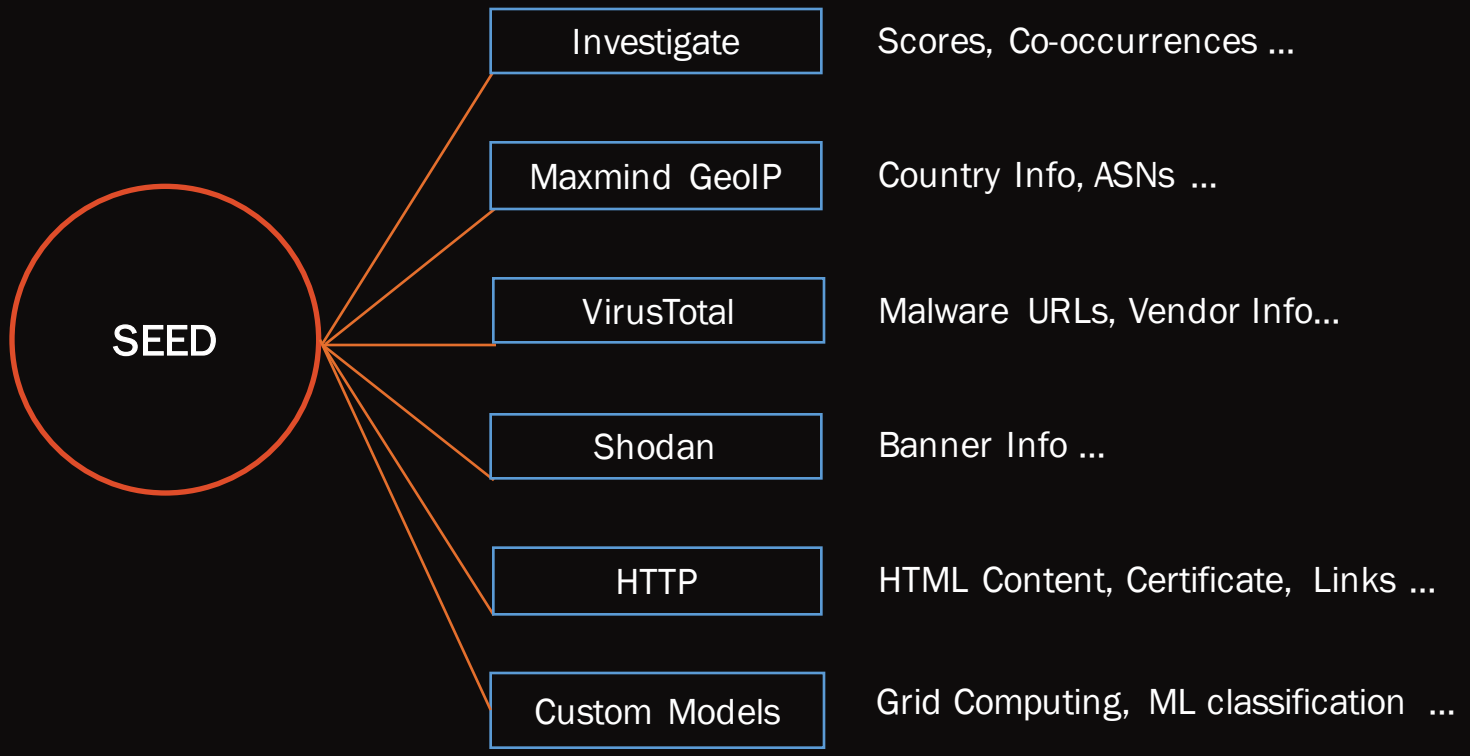
- ~30000 messages per second ⇔ 1 message every 33 microseconds
- 3 times faster than AMS.m1 query logs at peak time

GRAPH-ORIENTED DATA MINING

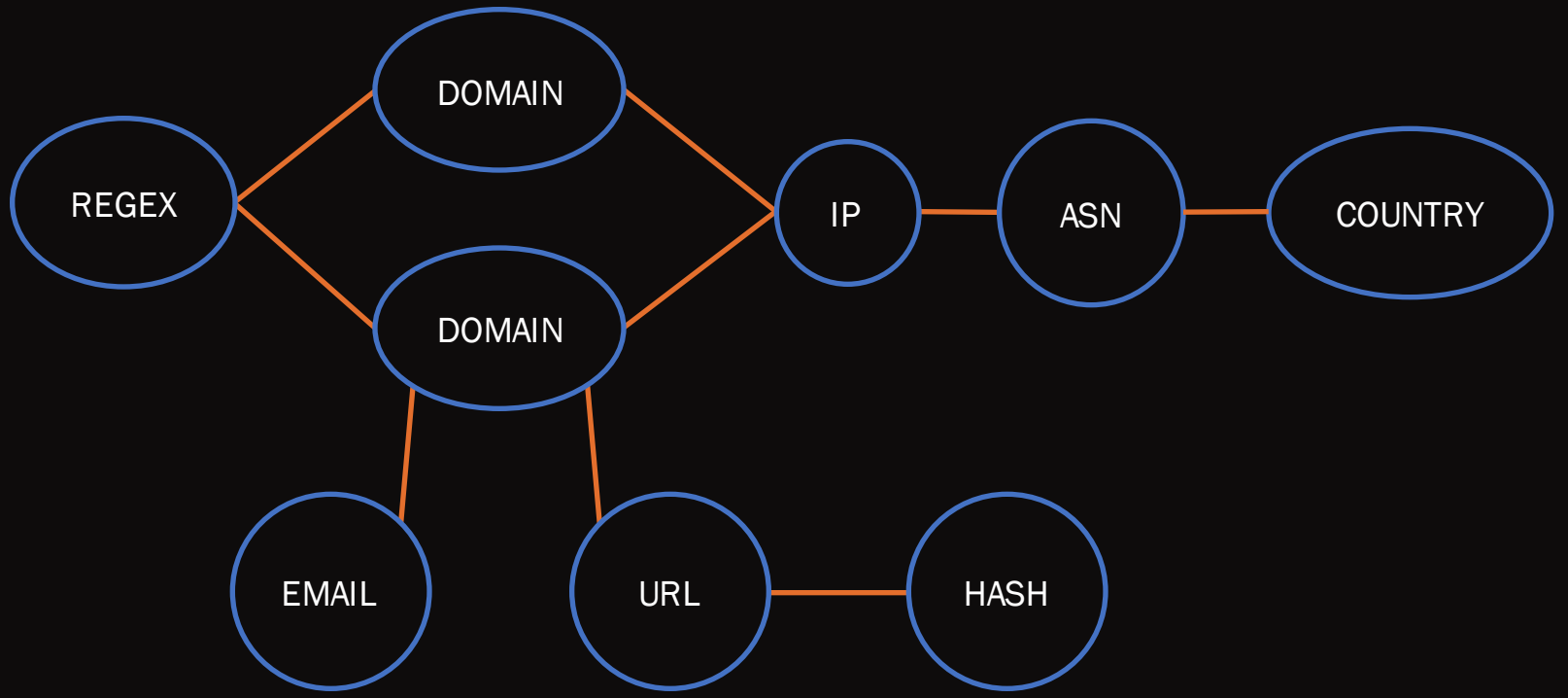


EXPLORING DATA SEEDS...

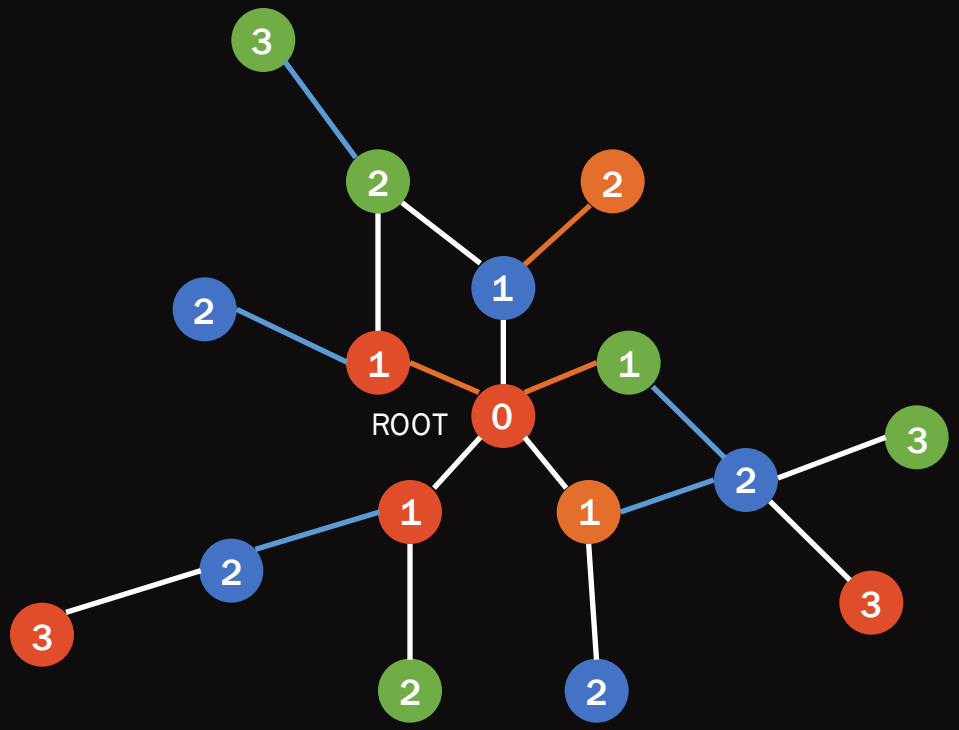
- Domain
- URL
- IP
- ASN
- Hash
- Email
- Regex



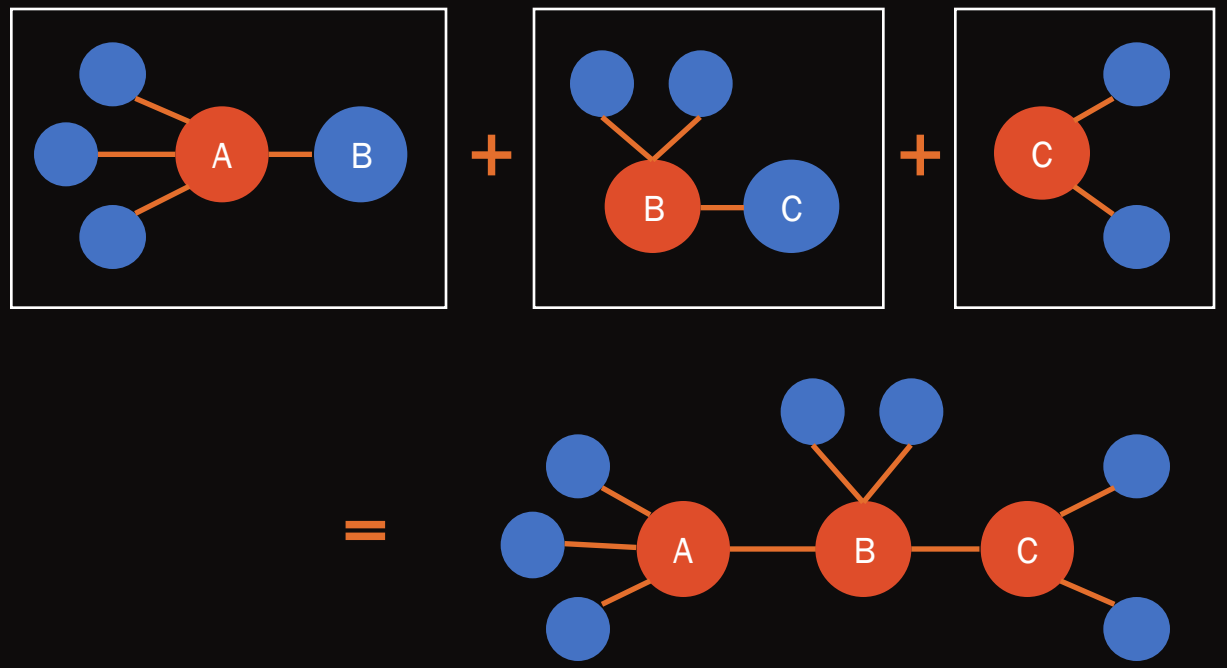
DATA MODELING EXAMPLE



DATA EXPLORATION: BREADTH FIRST TRAVERSAL



DISTRIBUTED BREADTH FIRST TRAVERSAL



LAMBDA MINING

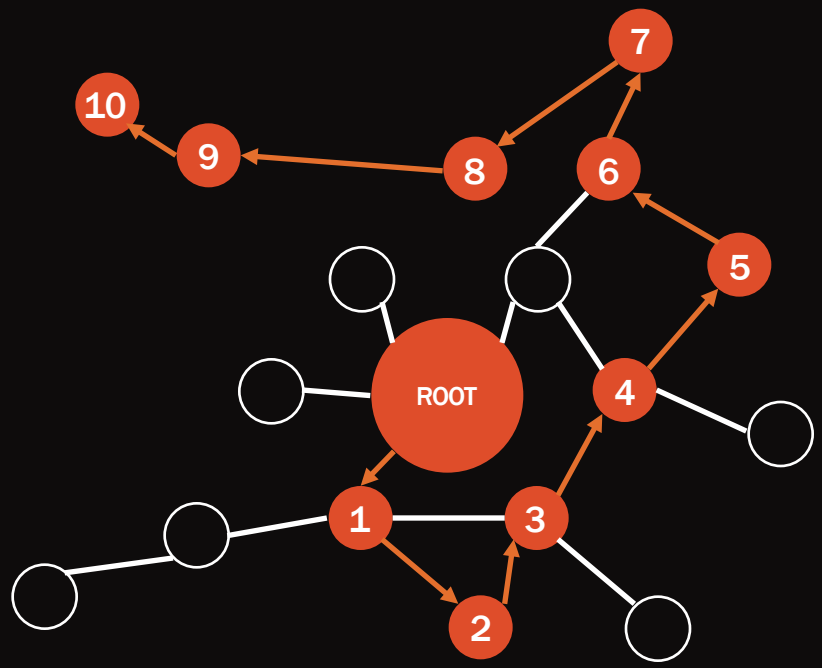
Functional Graph Exploration

Rule Based / Thresholds / Topology based ...

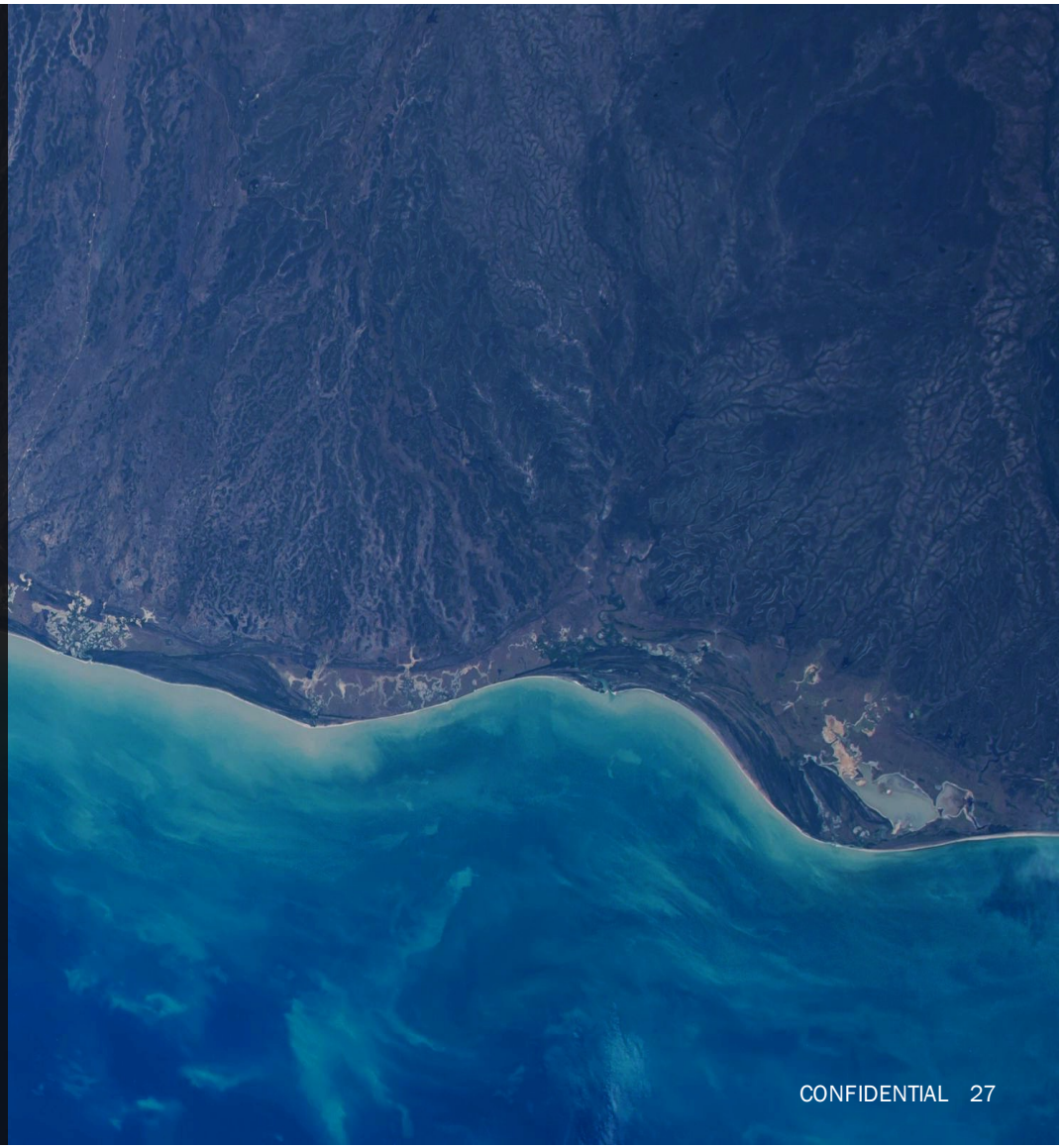
Subgraph Isomorphism Detection

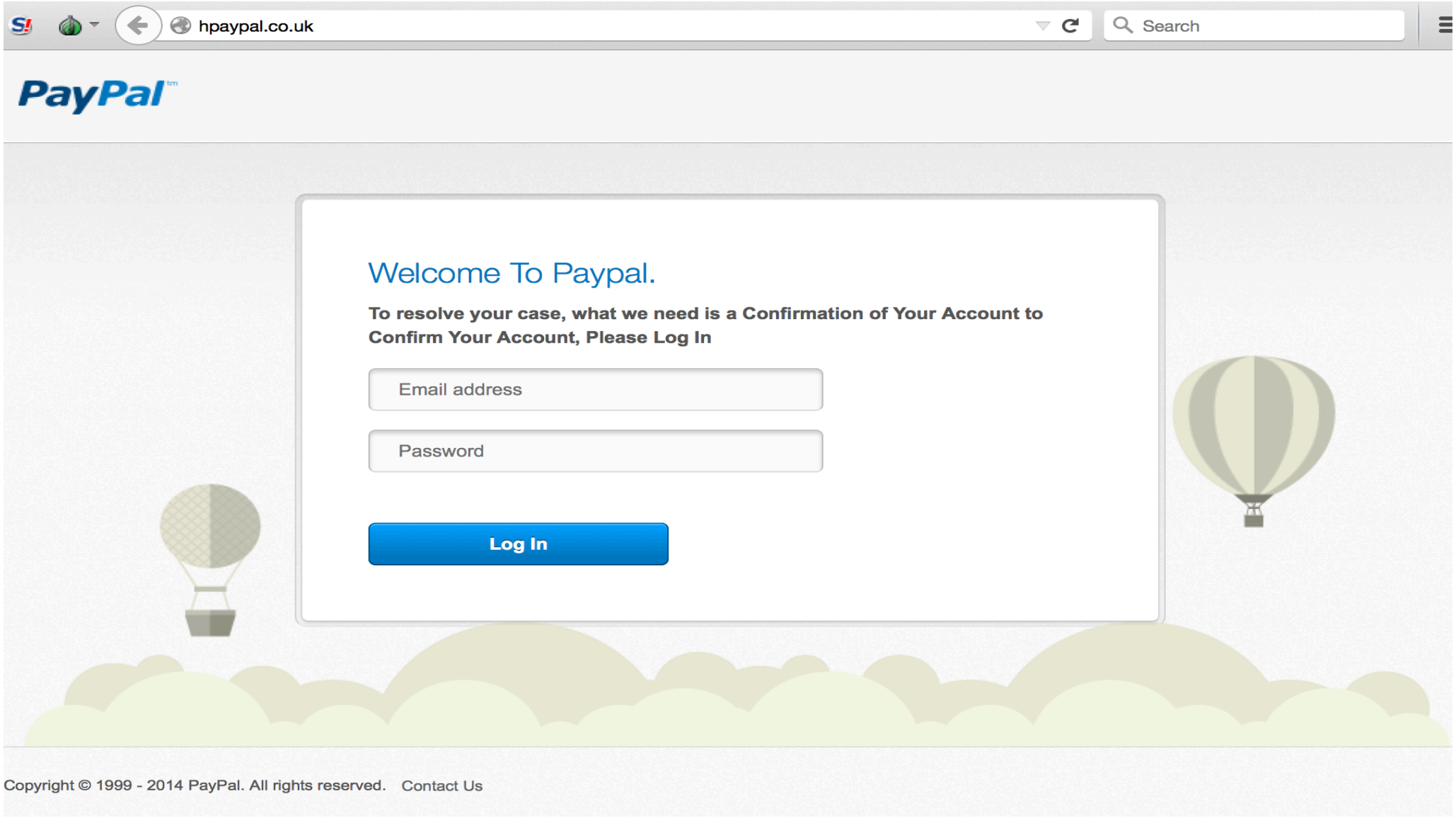
Profiles for specific use cases

Automated Smart Data Mining



PHISHING DETECTION WITH NLP-RANK




hpaypal.co.uk
Search
PayPal
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
Copyright © 1999 - 2014 PayPal. All rights reserved.
Contact Us

2016-02-09T19:10:18.439Z, hpaypal.co.uk, 0.9990674257278442, Paypal/www.paypal-united.com.html

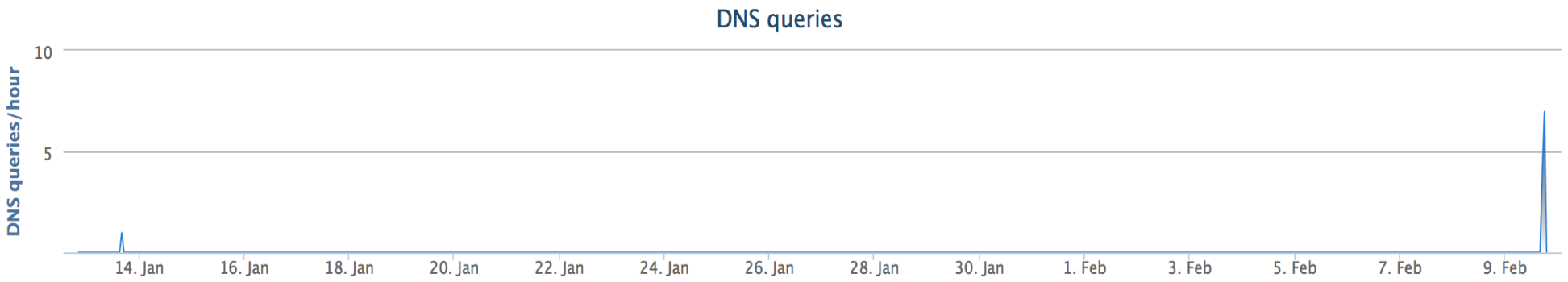
hpaypal.co.uk INVESTIGATE Visualize Back to top

DETAILS FOR HPAYPAL.CO.UK

Search in Google

Download as CSV

Classifier prediction: benign OpenDNS Security Graph Score: +54



WHOIS RECORD DATA

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

Created: February 9, 2016 NEW **Updated:** February 9, 2016 **Expires:** February 9, 2017 Raw data

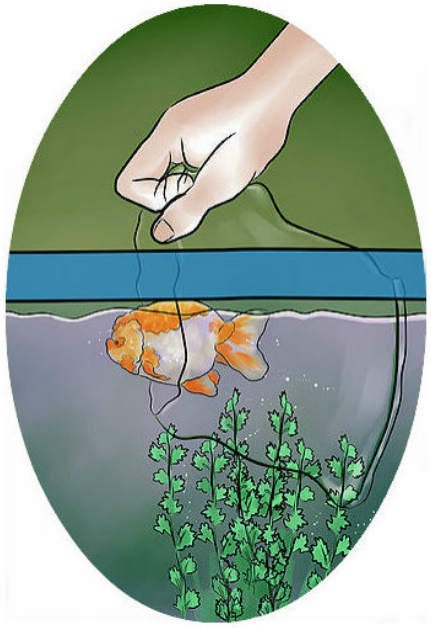
PURPOSE:

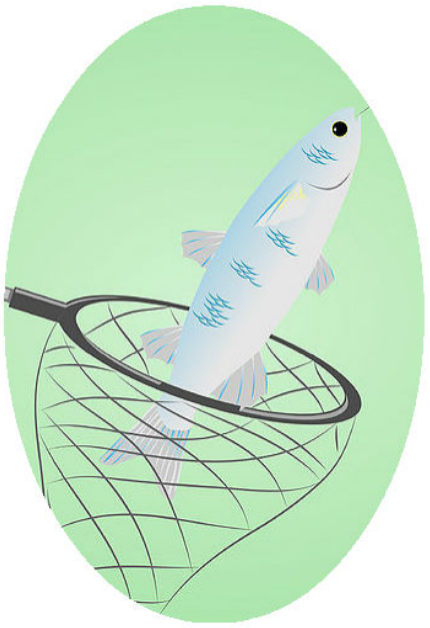
Overview of our NLPRank:

- 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).

PhishTank® Out of the Net, into the Tank.

Submit → Vote! → Categorize → Filter





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

Bank of America

PayPal

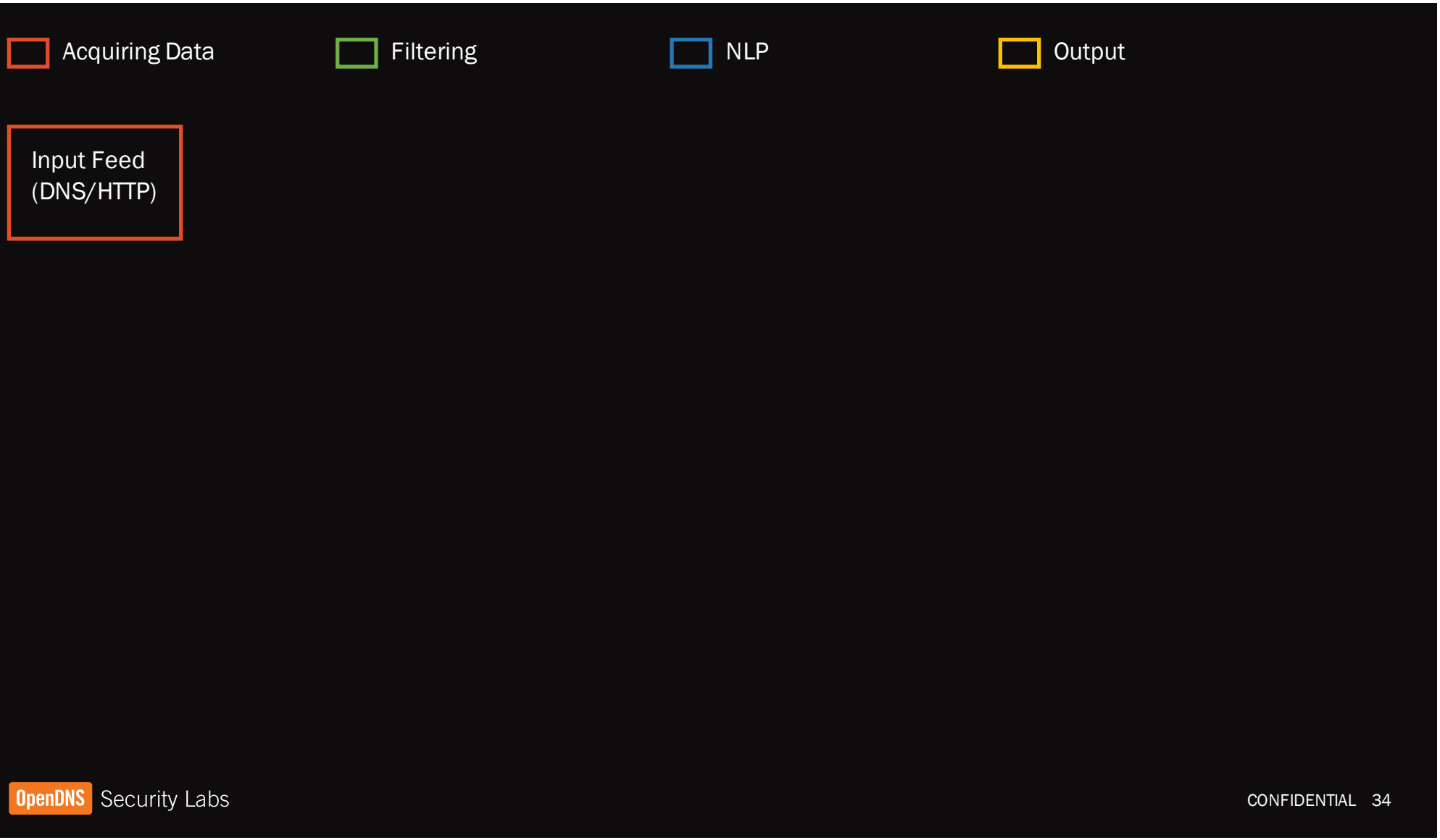
CHASE

Google

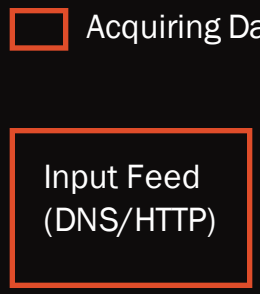
YAHOO!

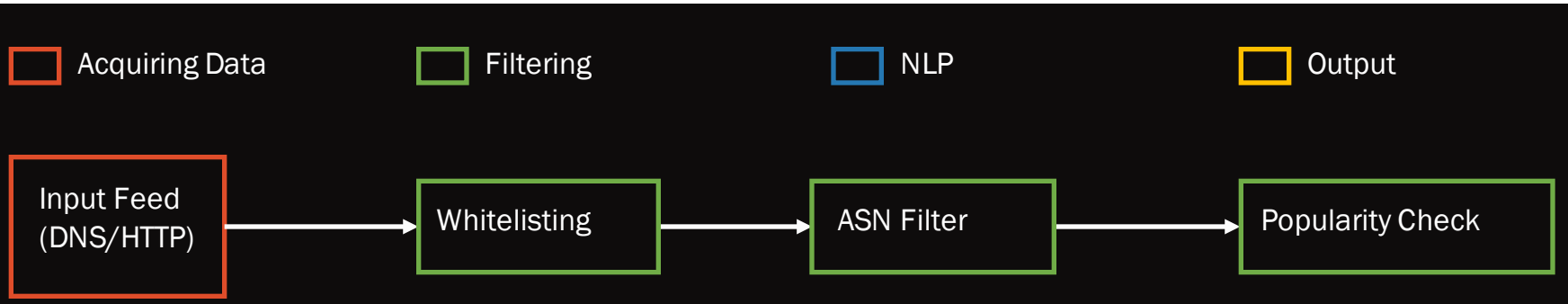
AMERICAN EXPRESS

WELLS FARGO


Acquiring Data
Filtering
NLP
Output
Input Feed
(DNS/HTTP)

HEURISTIC #1
ASN FILTERING


Acquiring Data
Filtering
NLP
Output
Input Feed
(DNS/HTTP)


Acquiring Data
Filtering
NLP
Output
Input Feed (DNS/HTTP)
Whitelisting
ASN Filter
Popularity Check

ASN FILTER + WHITELISTING

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

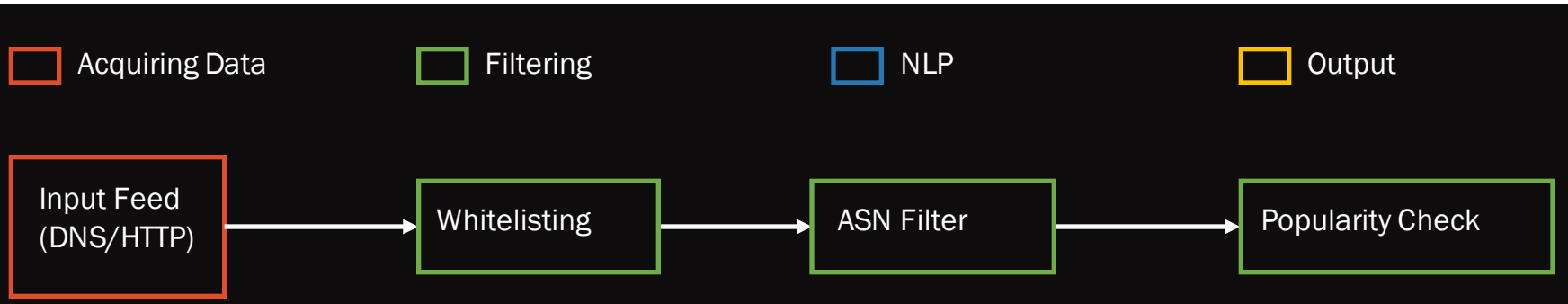
-Autonomous System Number == neighborhood/zipcode on the internet

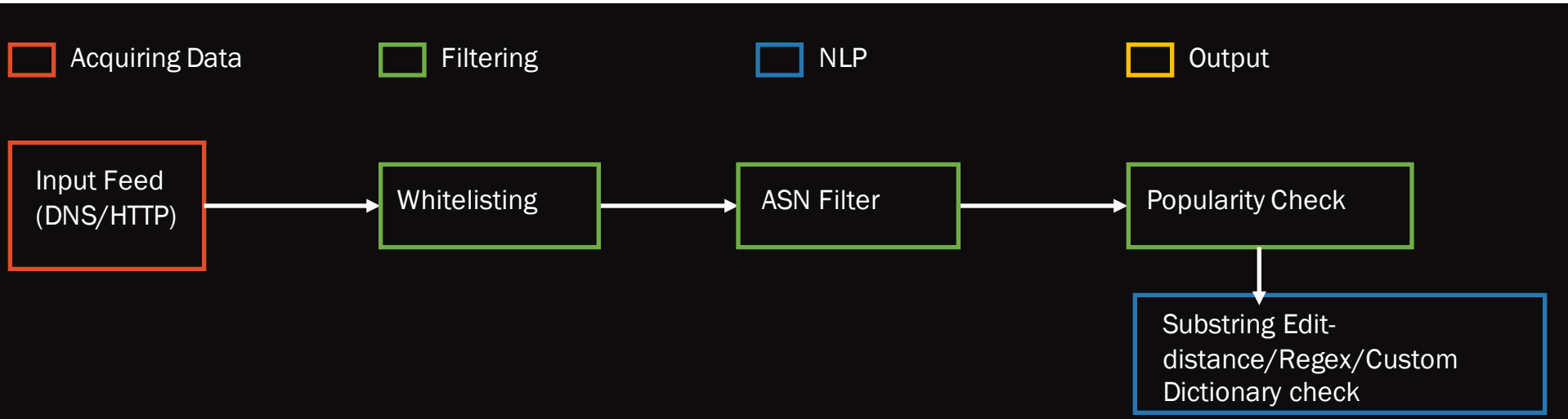
--Domains exhibiting fraudulent behavior hosted on ASNs that are not associated with the company they're spoofing

-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


Acquiring Data
Filtering
NLP
Output
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://www.kingdomhome.com.au/wp-admin/wellsfargo.zip/wellsfargo-online.server/details.html
http://wellsfargoonline.pfwv.com.br/wellsfargo/
http://www.cityroo.com/sarasoa/wellsfargo/wellsfargo-online.php
http://wellsfargo.com.billing.account.updatemyaccount.wellsfrago.com.onlineaccounts.upgrade.online.billing.account.update.nlineaccounts.upgrade.online.billing.account.update.kowafdfsfs.net
http://comosecuraladiabetes.com/wp-admin/js/well.htm

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
- firefoxupdata[.]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, wellsfargo

- 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

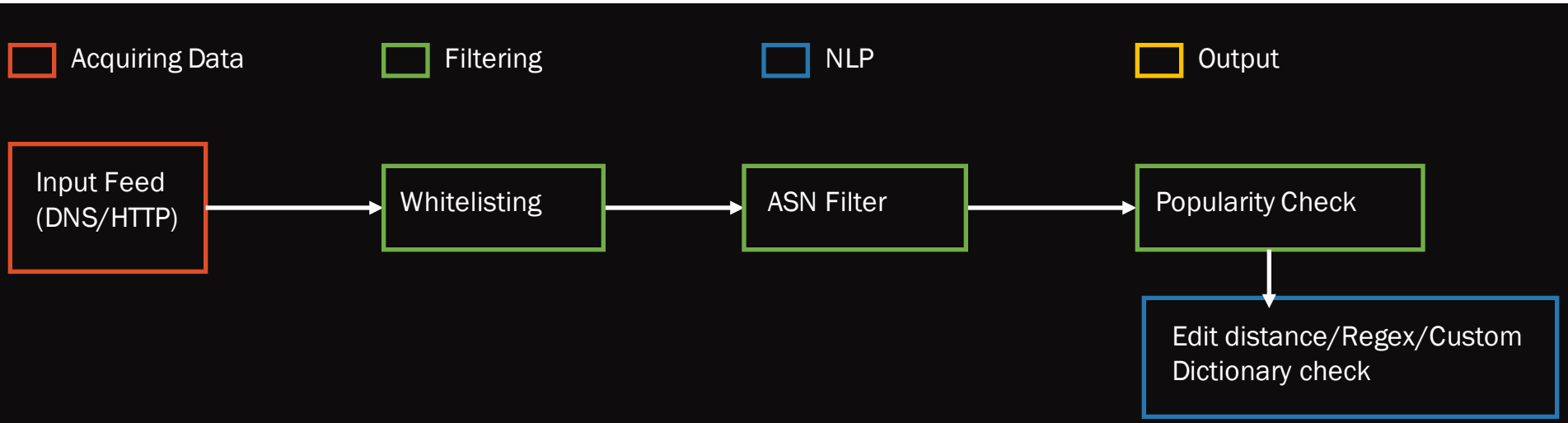
-Shortest Path, Dynamic Programming algorithm checking similarity between 2 strings (think spellcheck!)

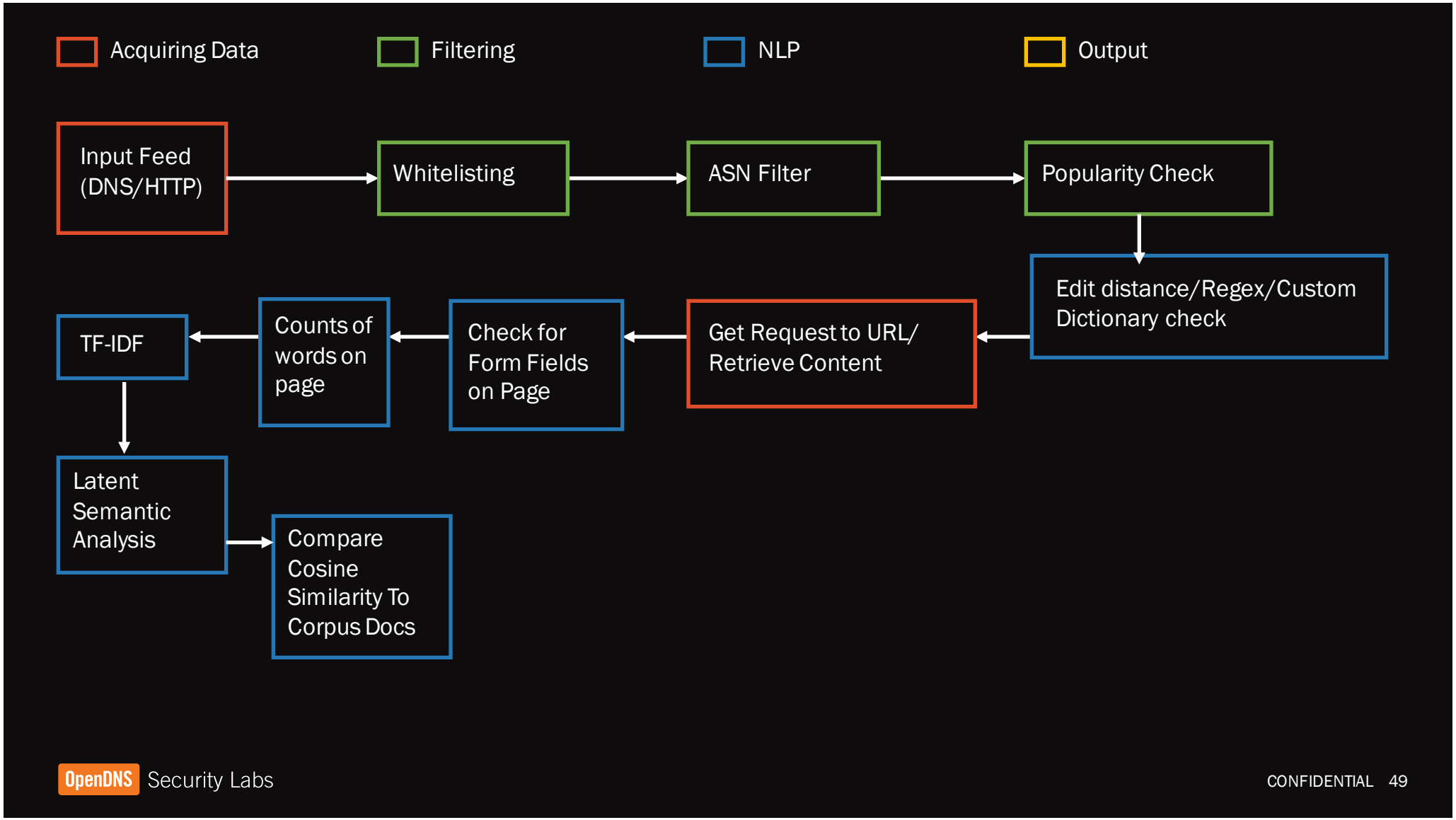
-How many “edits” it takes to turns one string into the other

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

google.com --> g00g1e.com 3 edits
dropbox.com → dr0pb0x.com 2 edits
bankofamerica.com → 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
Filtering
NLP
Output
Input Feed (DNS/HTTP)
Whitelisting
ASN Filter
Popularity Check
Edit distance/Regex/Custom Dictionary check
Get Request to URL/ Retrieve Content
Check for Form Fields on Page
Counts of words on page
TF-IDF
Latent Semantic Analysis
Compare Cosine Similarity To Corpus Docs

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:

1. macbeth 2. macduff 3. murder 4. ladi 5. witch 6. banquo 7. malcolm 8. ill 9. duncan 10. fear

$$w_{t,d} = (1 + \log \mathrm{tf}_{t,d}) \times \log_{10}(N / \mathrm{df}_t)$$

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

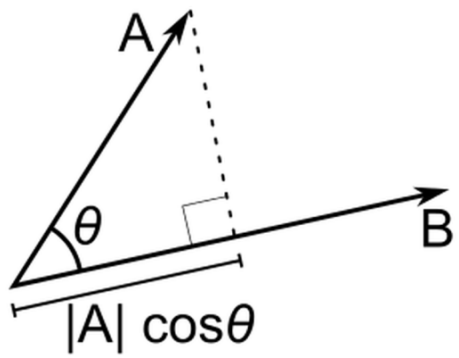
Process: Decompose X into 3 matrices, U, S, V(T)

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

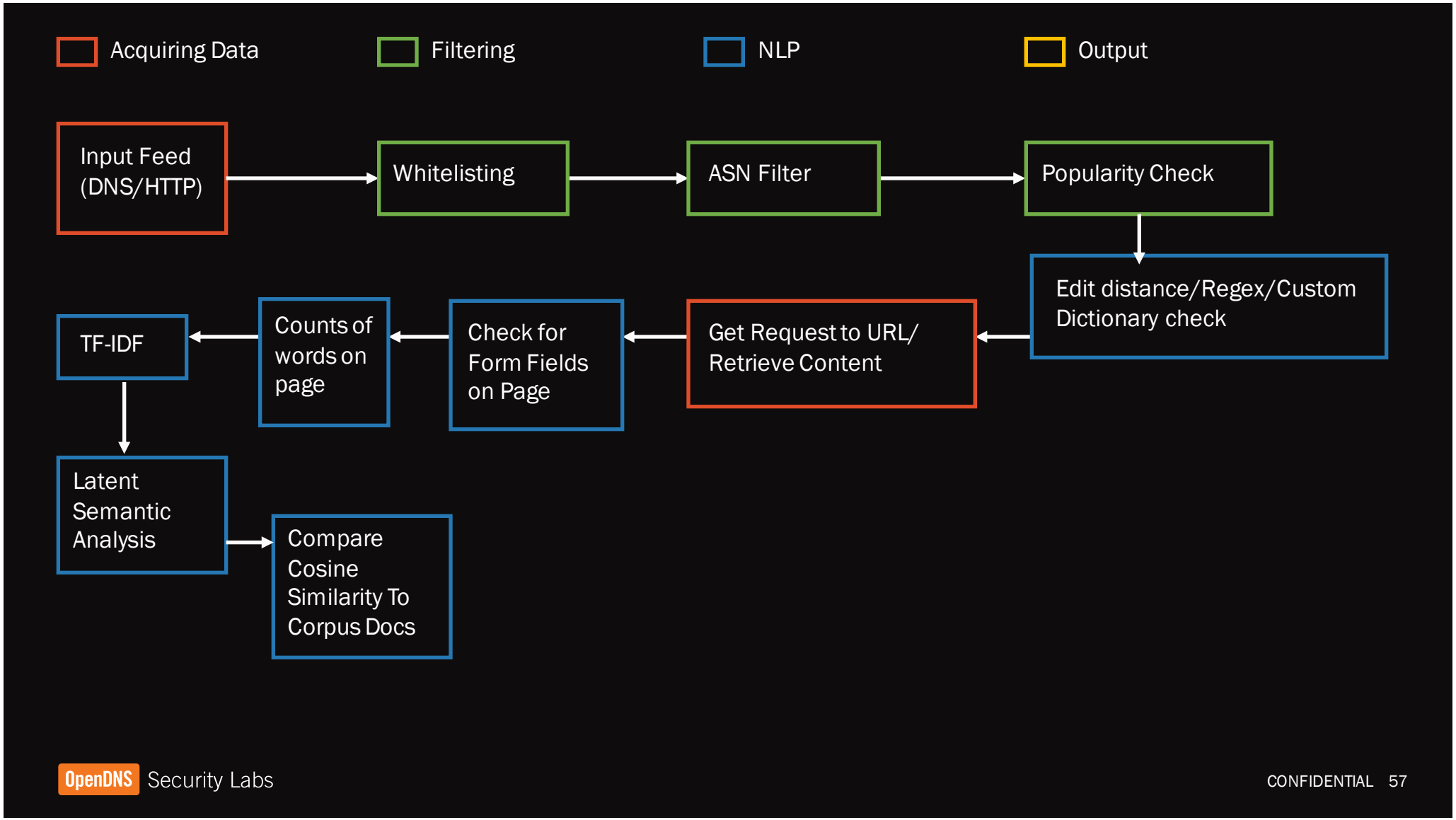
COSINE DISTANCE BETWEEN VECTORS

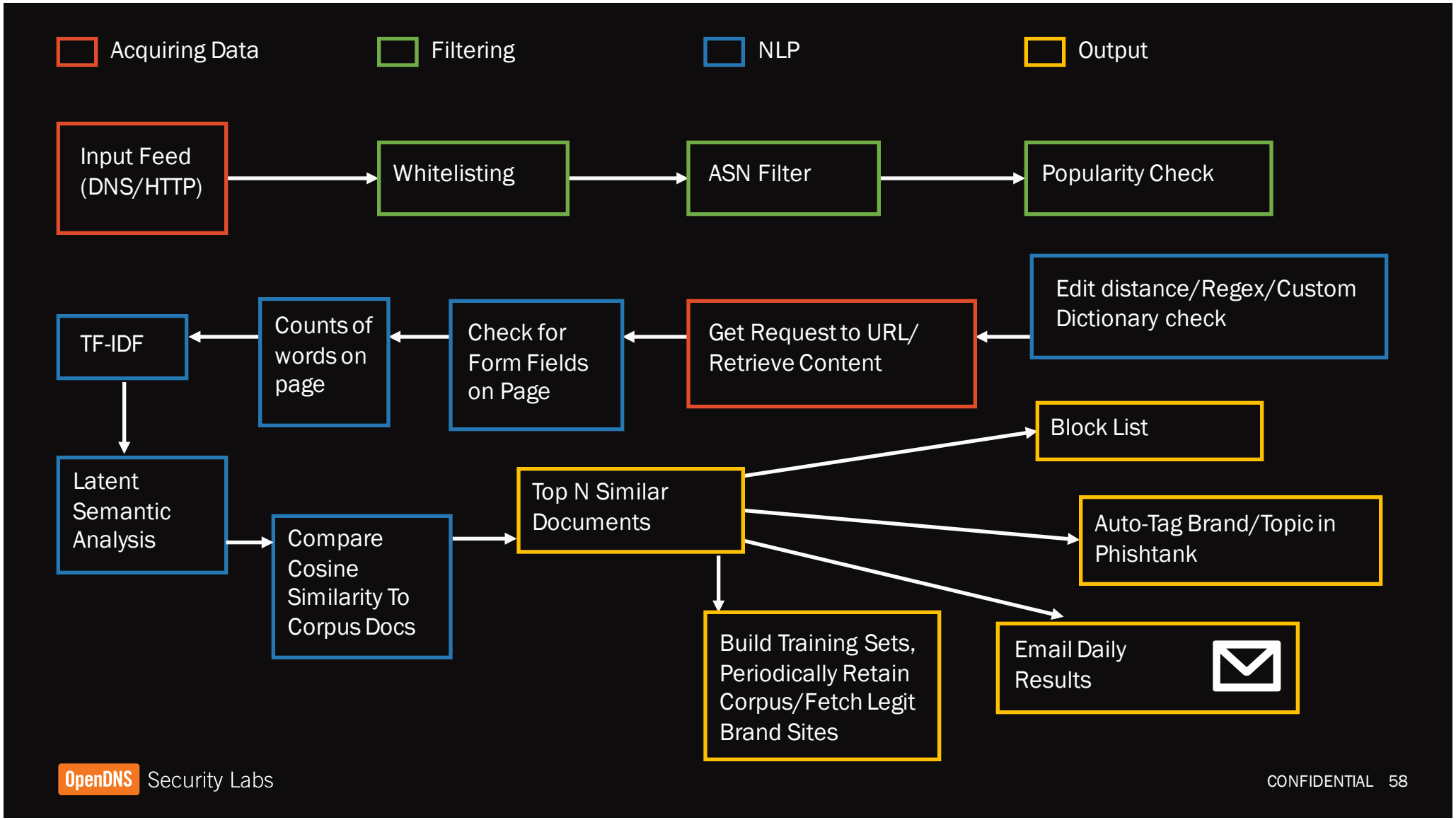


Cosine distance between two vectors:

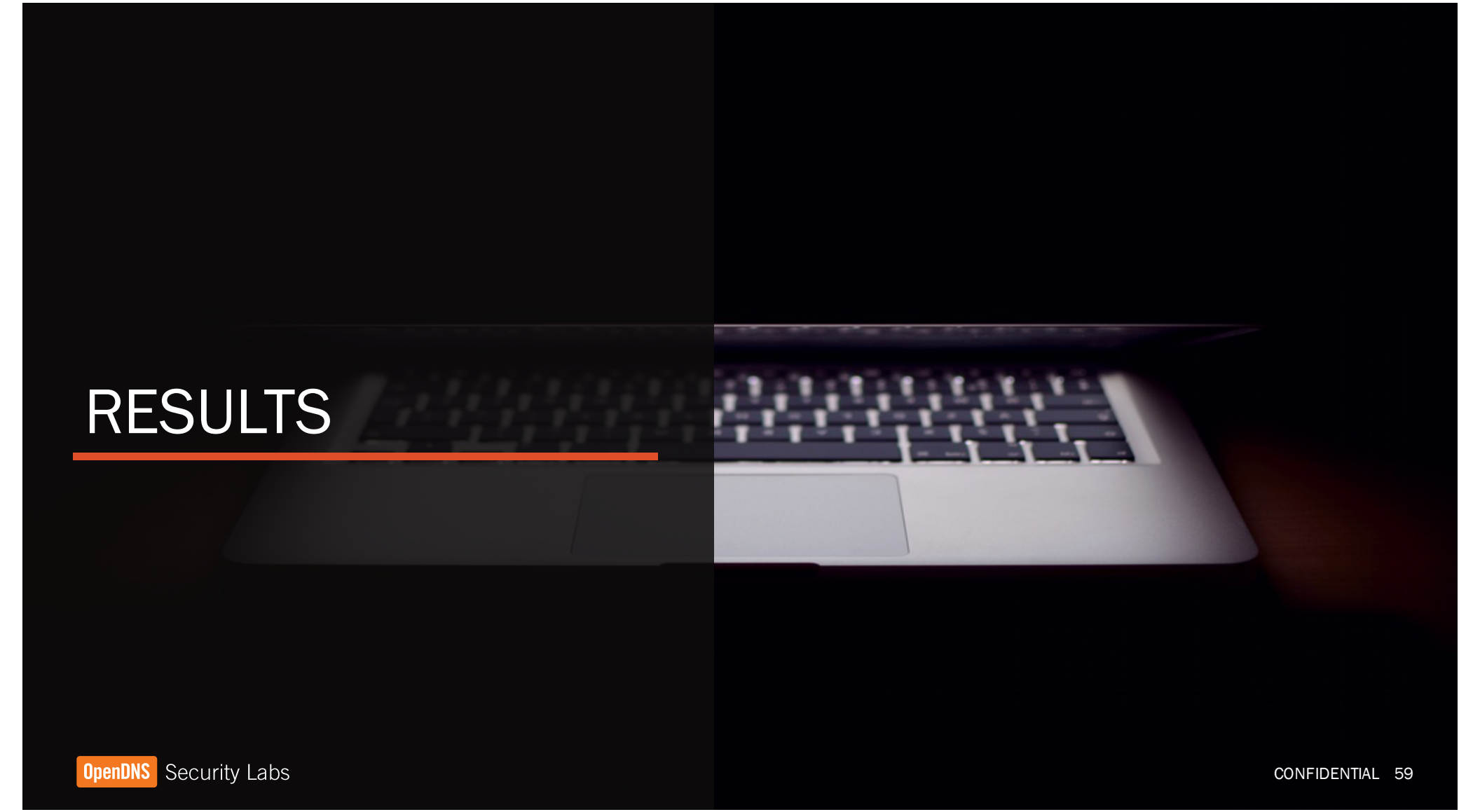
In[1]:= **CosineDistance[{a, b, c}, {x, y, z}]**

Out[1]=
$$1 - \frac{a\,x + b\,y + c\,z}{\sqrt{\text{Abs}[a]^2 + \text{Abs}[b]^2 + \text{Abs}[c]^2}\ \sqrt{\text{Abs}[x]^2 + \text{Abs}[y]^2 + \text{Abs}[z]^2}}$$


Acquiring Data
Filtering
NLP
Output
Input Feed (DNS/HTTP)
Whitelisting
ASN Filter
Popularity Check
Edit distance/Regex/Custom Dictionary check
Get Request to URL/ Retrieve Content
Check for Form Fields on Page
Counts of words on page
TF-IDF
Latent Semantic Analysis
Compare Cosine Similarity To Corpus Docs


Acquiring Data
Filtering
NLP
Output
Input Feed (DNS/HTTP)
Whitelisting
ASN Filter
Popularity Check
Edit distance/Regex/Custom Dictionary check
Get Request to URL/ Retrieve Content
Check for Form Fields on Page
Counts of words on page
TF-IDF
Block List
Latent Semantic Analysis
Compare Cosine Similarity To Corpus Docs
Top N Similar Documents
Auto-Tag Brand/Topic in Phishtank
Email Daily Results
Build Training Sets, Periodically Retain Corpus/Fetch Legit Brand Sites

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: 9990196078431373

Precision: 0.9916666666666667

Run 4:

Accuracy: 0.9921568627450981

Precision: 0.9333333333333333

Run 5:

Accuracy: 0.9901960784313726

Precision: 0.9166666666666666

DEDICATED SAMPLES


3276422.apple-confirmation.online/apple/3210ddbeaa16948a702b6049b8d9a202.ht
iframe name="pendik evden
Mac
iPad
iPhone
Watch
TV
Music
Support
Apple ID
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

https://investigate.opendns.com/domain-view/name/df0f13dc73e14f59.apple-confirmation.online/view

Search

df0f13dc73e14f59.apple-confirmation.online INVESTIGATE Visualize Back to top

DETAILS FOR DF0F13DC73E14F59.APPLE-CONFIRMATION.ONLINE

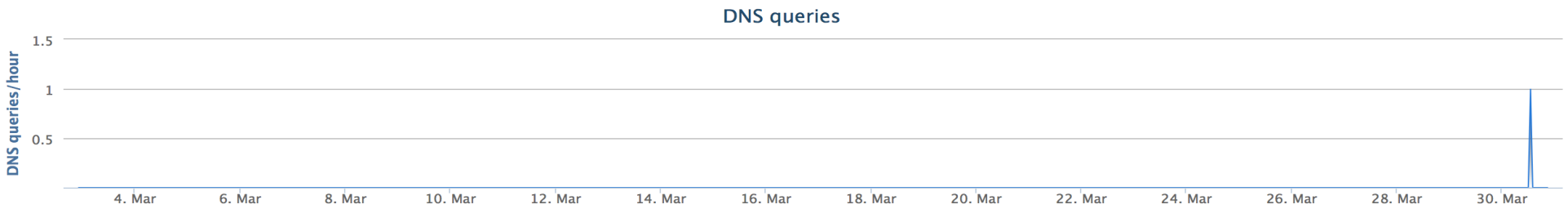
Search in Google

Search in VirusTotal

This domain is currently in the OpenDNS Security Labs block list

This domain is associated with the following type of threat: Phishing

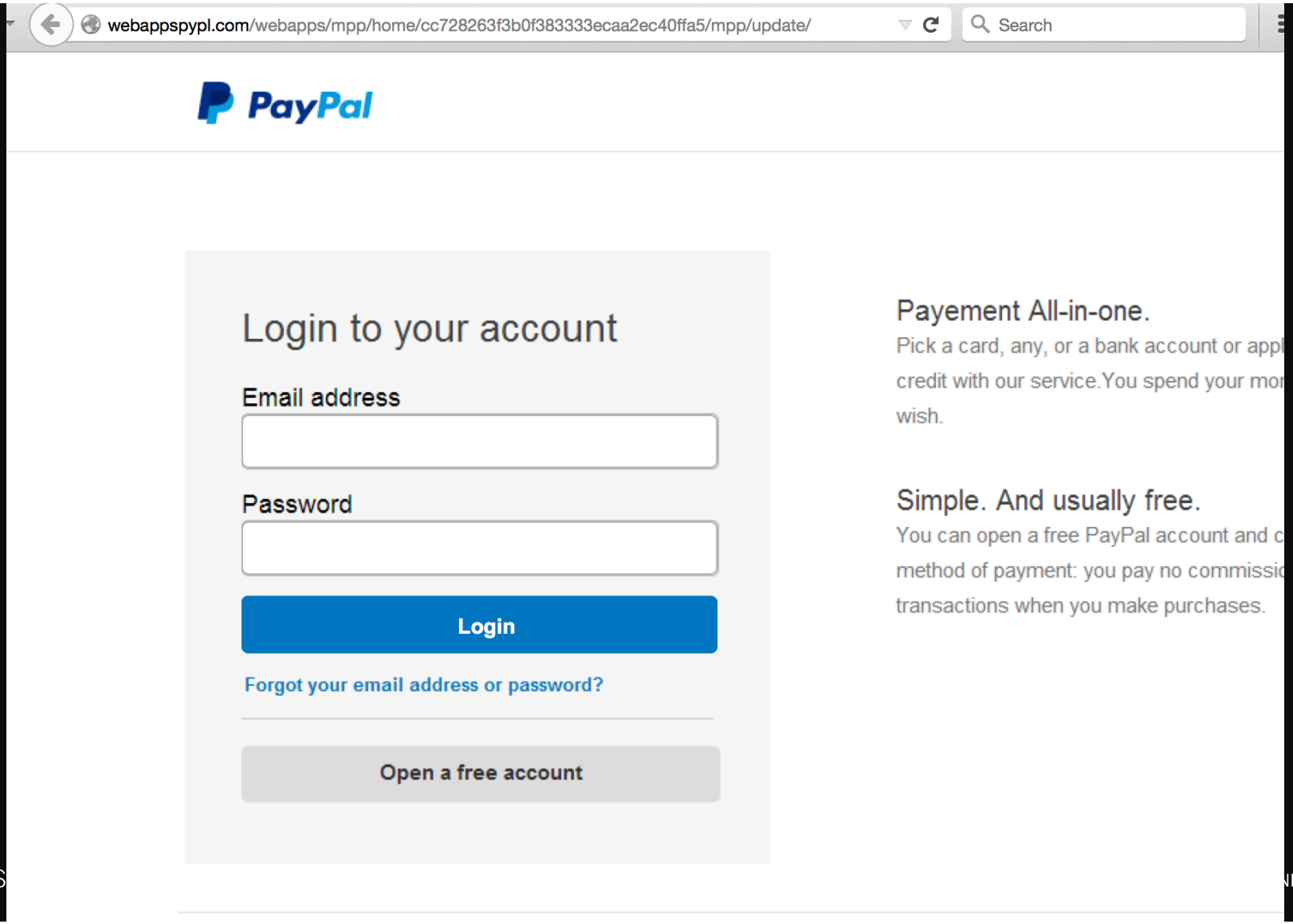
This domain may have been created using a domain generation algorithm (DGA)



WHOIS RECORD DATA

Registrar Name: Namecheap **IANAID:** 1068 Last retrieved March 30, 2016 Get latest

Created: March 30, 2016 NEW **Updated:** March 30, 2016 **Expires:** March 30, 2017 Raw data


webappspypl.com/webapps/mpp/home/cc728263f3b0f383333ecaa2ec40ffa5/mpp/update/
Search
PayPal
Login to your account
Email address
Password
Login
Forgot your email address or password?
Open a free account
Payement All-in-one.
Pick a card, any, or a bank account or appl
credit with our service.You spend your mor
wish.
Simple. And usually free.
You can open a free PayPal account and c
method of payment: you pay no commissio
transactions when you make purchases.

2016-02-11T01:02:00.974Z, webappspypl.com, 0.9960060715675354, Paypal/bluespruceus.com.html

webappspypl.com

INVESTIGATE

Visualize

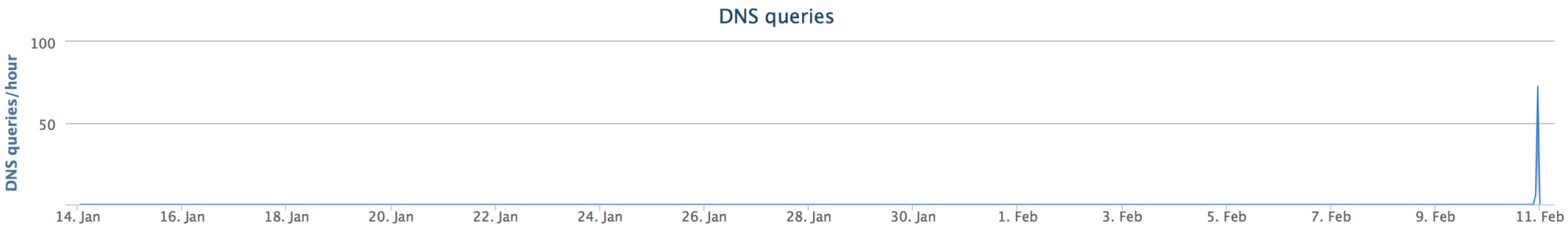
DETAILS FOR WEBAPPSPYPL.COM

Search in Google

Download as CSV

Classifier prediction: benign

OpenDNS Security Graph Score: +62



WHOIS RECORD DATA

Registrar Name: NAMEBAY **IANAID:** 88

Last retrieved February 11, 2016

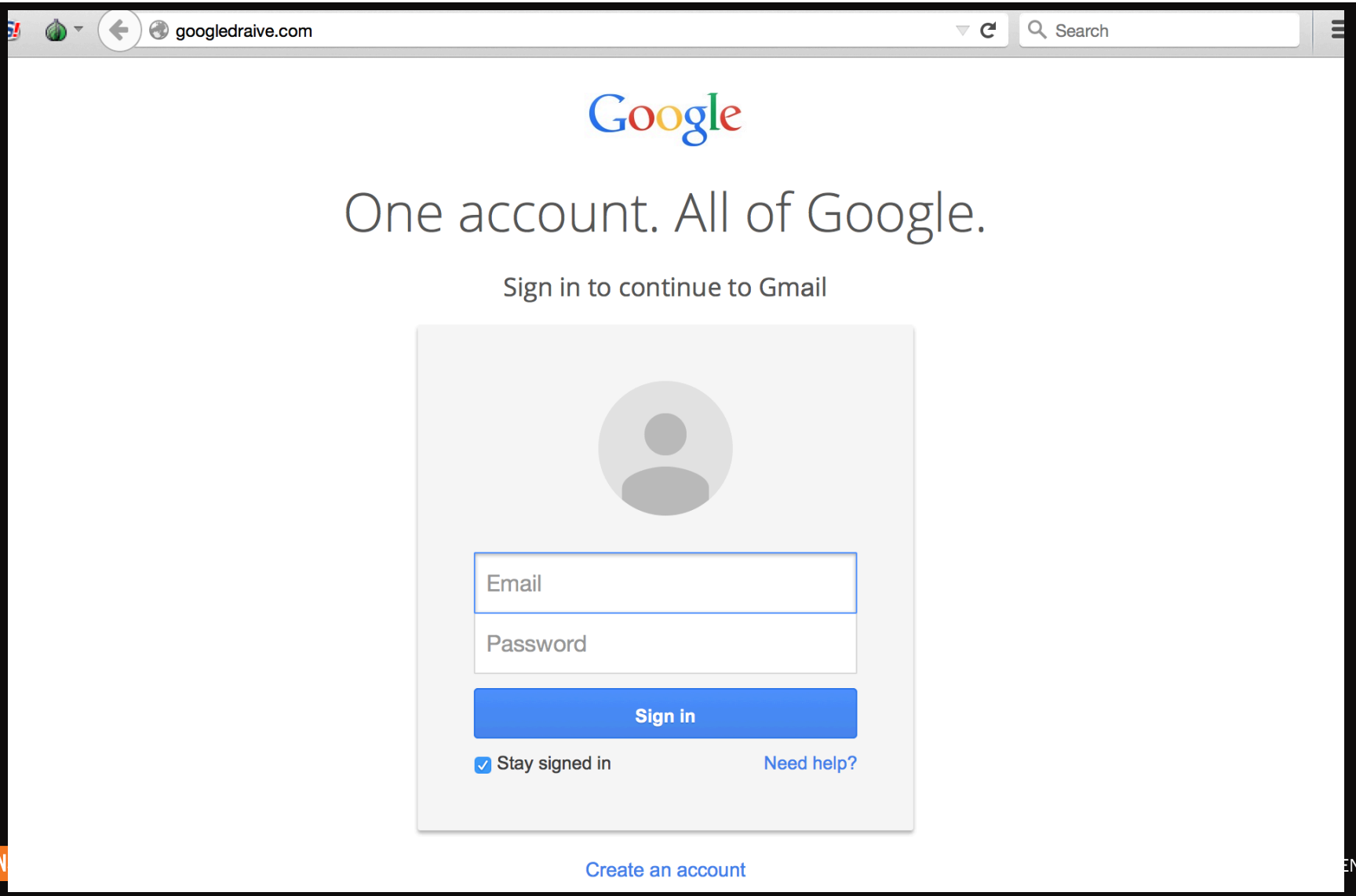
Get latest

Created: February 10, 2016 NEW

Updated: February 10, 2016

Expires: February 10, 2017

Raw data


googledraive.com
Search
Google
One account. All of Google.
Sign in to continue to Gmail
Email
Password
Sign in
Stay signed in
Need help?
Create an account

googledraive.com 0.9555937051773071, Google/www.rotisseriebuongusto.com.br

DETAILS FOR GOOGLEDRAIVE.COM

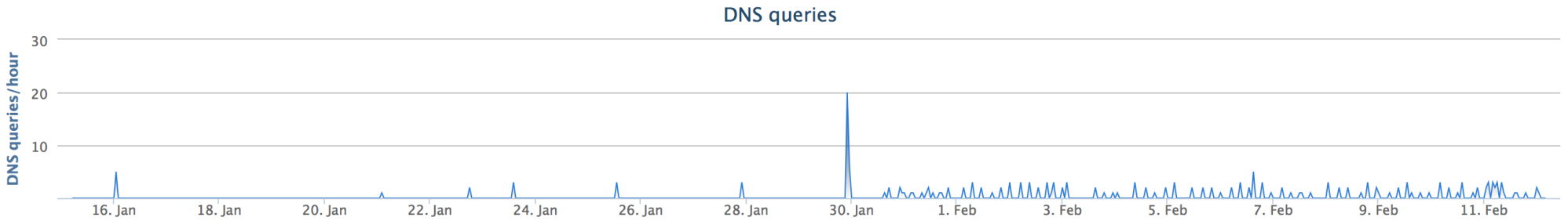
Search in Google

Download as CSV

This domain is currently in the OpenDNS Security Labs block list

Classifier prediction: suspicious

OpenDNS Security Graph Score: -94



WHOIS RECORD DATA

Registrar Name: PDR Ltd. d/b/a PublicDomainRegistry.com **IANAID:** 303

Last retrieved February 6, 2016 Get latest

Created: September 4, 2015 **Updated:** September 4, 2015 **Expires:** September 4, 2016 Raw data

Email Address	Associated Domains	Email Type	Last Observed
s.miloshevich@yandex.ru	8 Total - 3 malicious	Administrative, Registrant, Technical	Current

s.miloshevich@yandex.ru

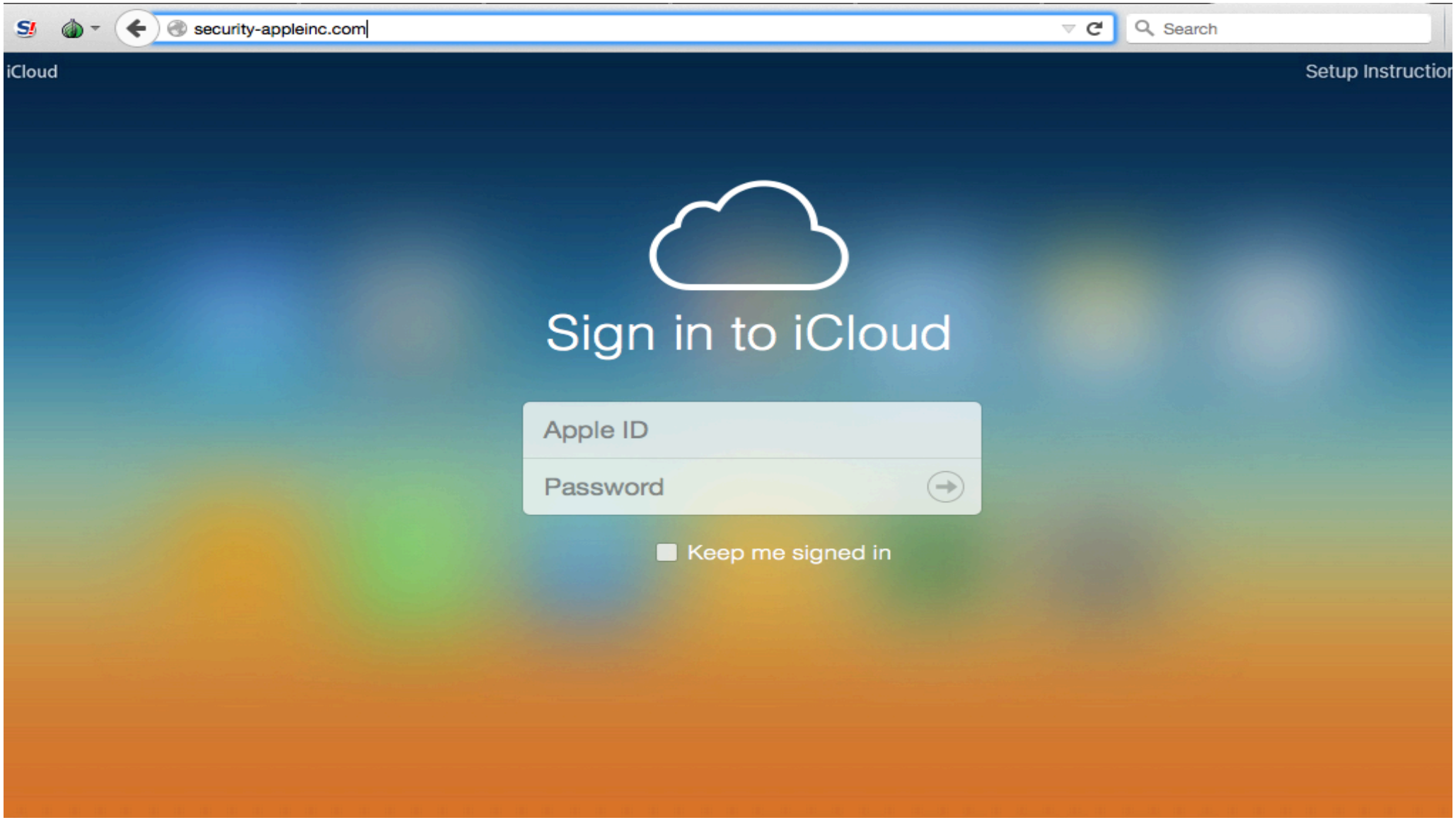
INVESTIGATE

Visualize

DOMAINS ASSOCIATED WITH S.MILOSHEVICH@YANDEX.RU

Domain Name	Security Categories	Content Categories	Last Observed
alert-login-gmail.com	Malware, Phishing		Current
suporteng.com	Malware		Current
docsautentification.com			Current
g000glemail.com			Current
googledraive.com			Current
googlsupport.com			Current
membrana52.com			Current
pwdrecover.com			Current

Showing 8 of 8 results


security-appleinc.com
Search
iCloud
Setup Instructio
Sign in to iCloud
Apple ID
Password
Keep me signed in

security-appleinc.com 0.9436904191970825, Apple/www.lcloudid-ds.top.html

DETAILS FOR SECURITY-APPLEINC.COM

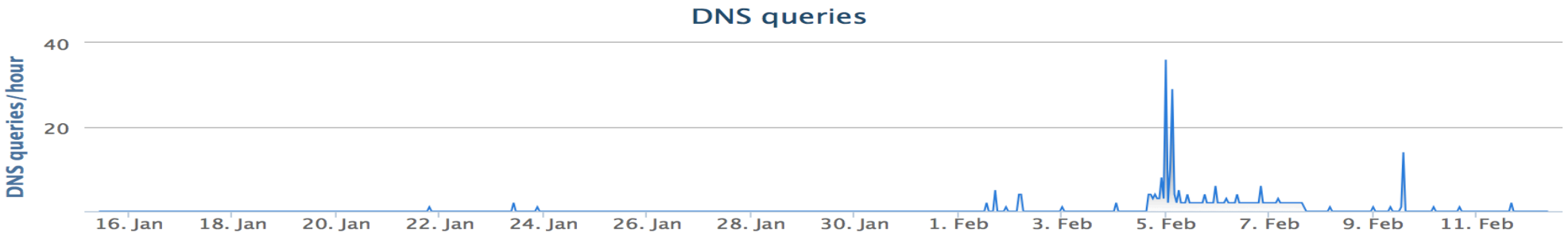
Search in Google

Download as CSV

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 Get latest

Created: January 8, 2016 **Updated:** January 16, 2016 **Expires:** January 8, 2017 Raw data

Email Address	Associated Domains	Email Type	Last Observed
trustedmon@gmail.com	11 Total - 8 malicious	Administrative, Registrant,	Current

trustedmon@gmail.com

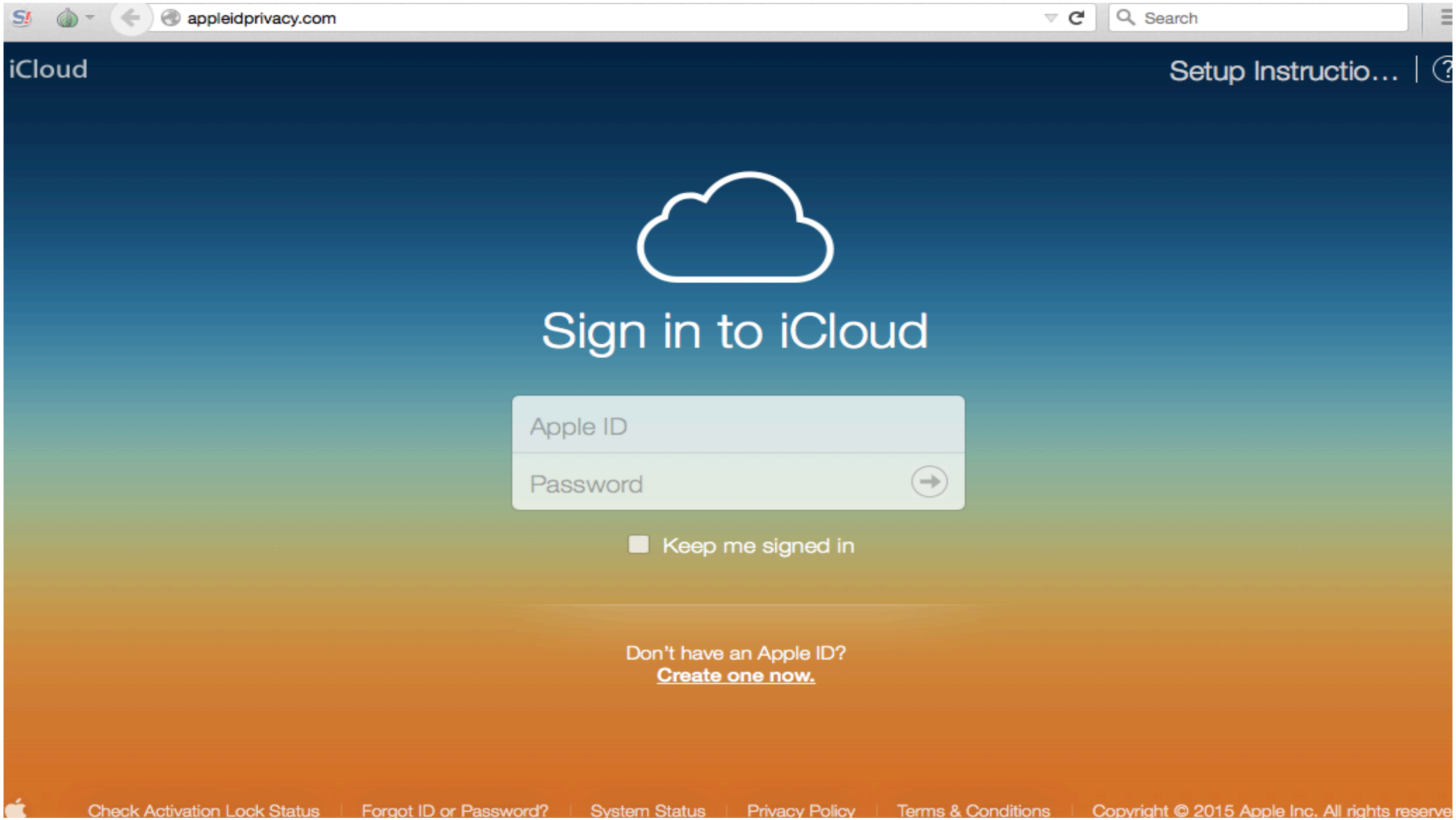
INVESTIGATE

Visualize

DOMAINS ASSOCIATED WITH TRUSTEDMON@GMAIL.COM

Domain Name	Security Categories	Content Categories	Last Observed
appleidprivacy.com			Current
appleinc-security.com			Current
appleinc-support.com			Current
icloudprivacy.com			Current
inc-appleid.info			Current
secure-appleinc.com			Current
security-apple.com			Current
security-appleinc.com			Current

Showing 8 of 8 results


appleidprivacy.com
Search
iCloud
Setup Instructio…
Sign in to iCloud
Apple ID
Password
Keep me signed in
Don't have an Apple ID?
Create one now.
Check Activation Lock Status
Forgot ID or Password?
System Status
Privacy Policy
Terms & Conditions
Copyright © 2015 Apple Inc. All rights reserve

2016-02-26T16:15:01.418Z www[.]gmail-remind[.]tk 0.9633020758628845 Google/www.gmail-edit.pw.html

104.207.132.165 INVESTIGATE Visualize Back to to

MALICIOUS DOMAINS HOSTED BY 104.207.132.165

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

FEATURES

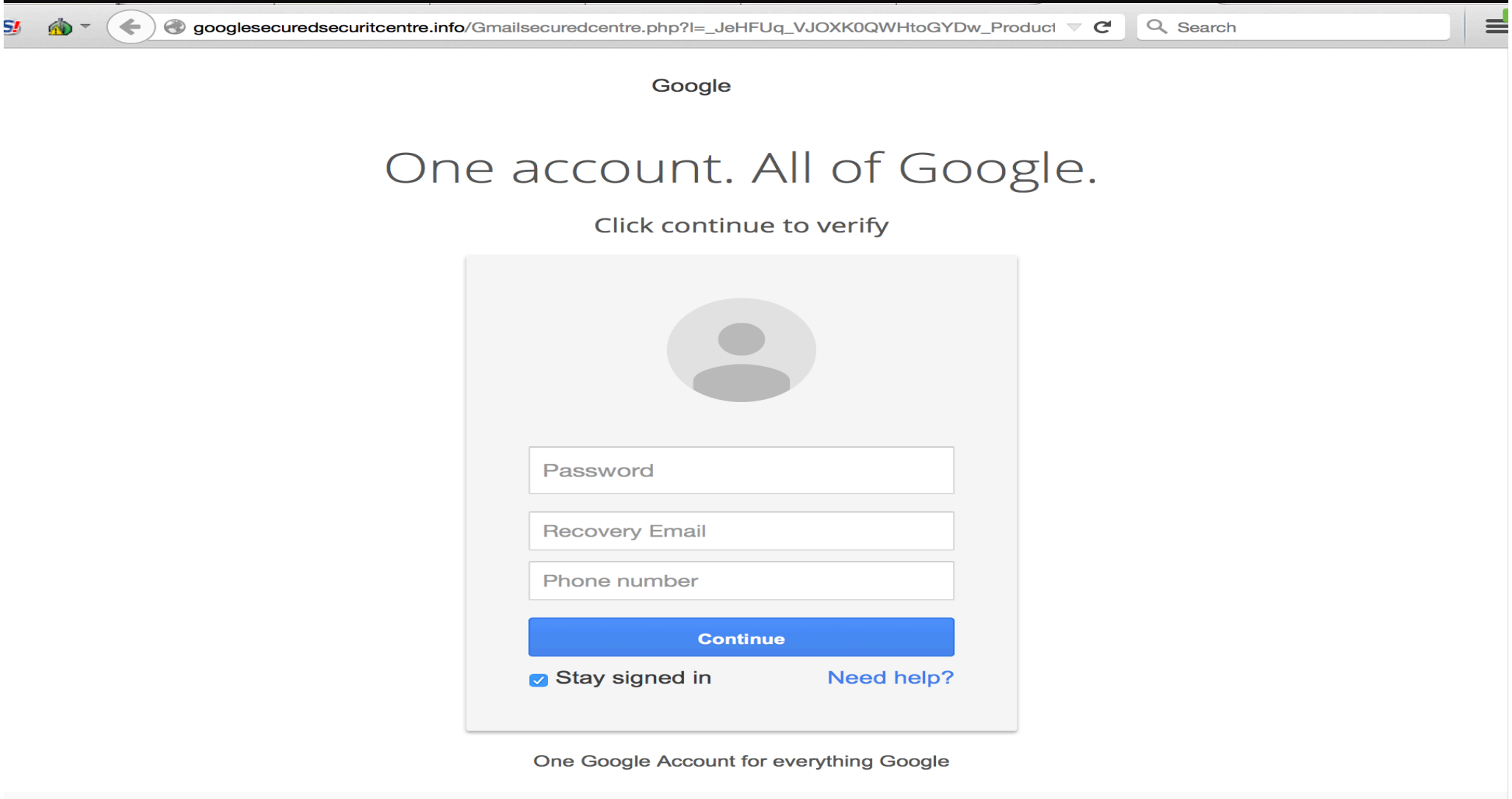
Known domains hosted at this IP	20
LD2 domains count	16
LD3 domains count	20
LD2-1 domains count	15
LD2-2 domains count	20
LD2 domains diversity	0.8
LD3 domains diversity	1
LD2-1 domains diversity	0.75
LD2-2 domains diversity	1

KNOWN DOMAINS HOSTED BY 104.207.132.165

sg-images.yahoo-images.com www.gmail-remind.tk www.gmail-secure.tk www.yahoo-noreply.tk yahoo-noreply.tk gmail-retry.tk www.gmail-retry.tk phpinfo.pw www.gmail-safety.pw www.yahoo-safety.com www.gmail-edit.pw www.yahoo-maintain.pw yahoo-maintain.pw yahoo-protect.com www.gmail-maintain.tk www.yahoo-protected.tk www.yahoo-operation.tk accounts-163.tk www.yahoo-protect.com yahoo-protect.tk


www.yahoo-noreply.tk
Search
YAHOO!
奇摩
服務說明
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。
YAHOO!
奇摩
登入您的帳號
電子信箱
密碼
保持我的登入狀態
登入
無法存取自己的帳號？
第一次使用 Yahoo奇摩？
註冊新帳號
服務條款 | 隱私權

2016-03-23T04:25:54.807Z googlesecuredsecuritcentre[.]info 0.9594852328300476 Google/192.185.217.156.html



OpenDNS Investigate

DOCUMENTATION INVESTIGATE SUPPORT jeremiah@opendns.com Sign Out

SEARCH PATTERN SEARCH

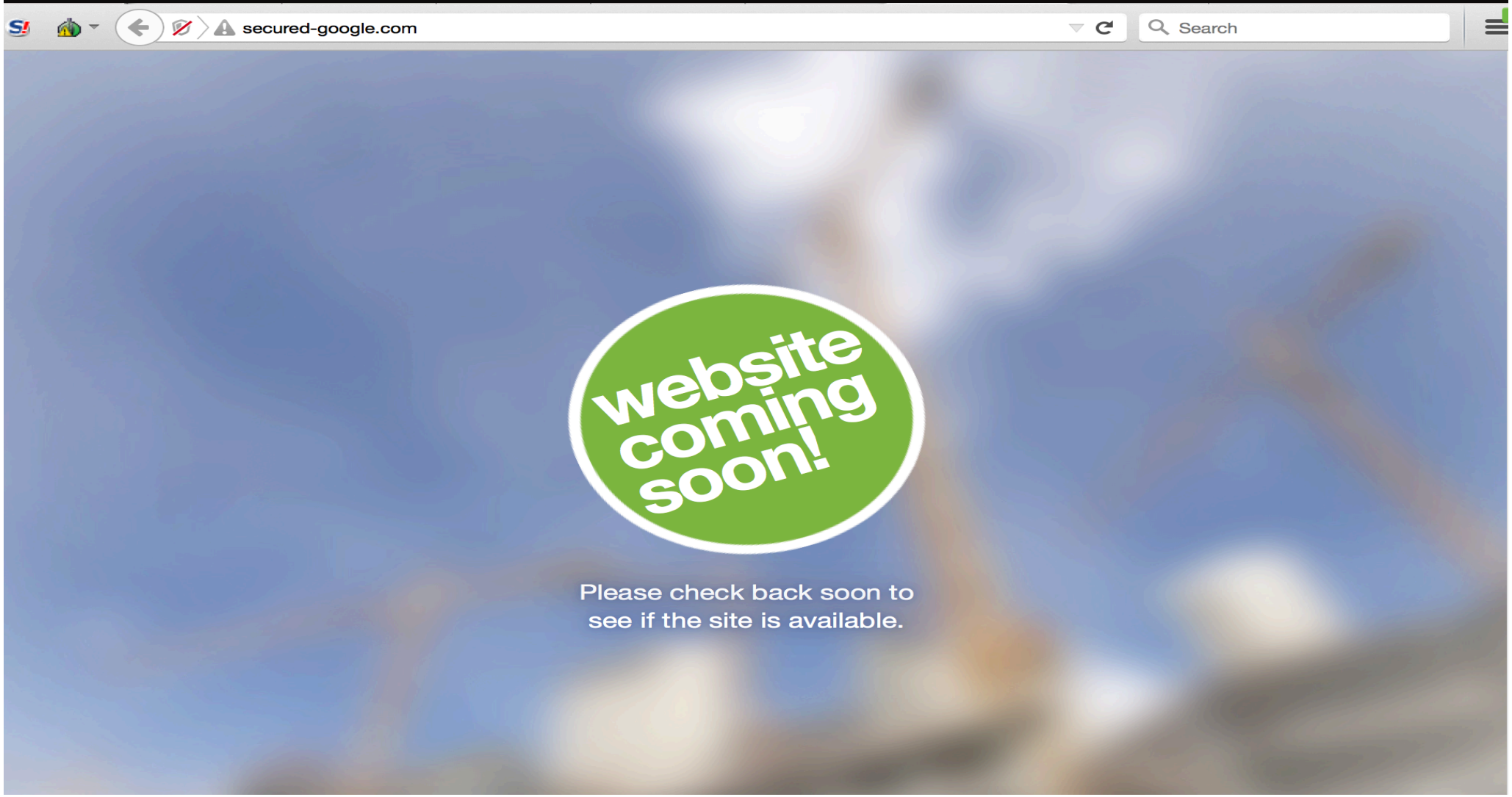
sjwdrum@gmail.com INVESTIGATE

Visualize

DOMAINS ASSOCIATED WITH SJWDRUM@GMAIL.COM

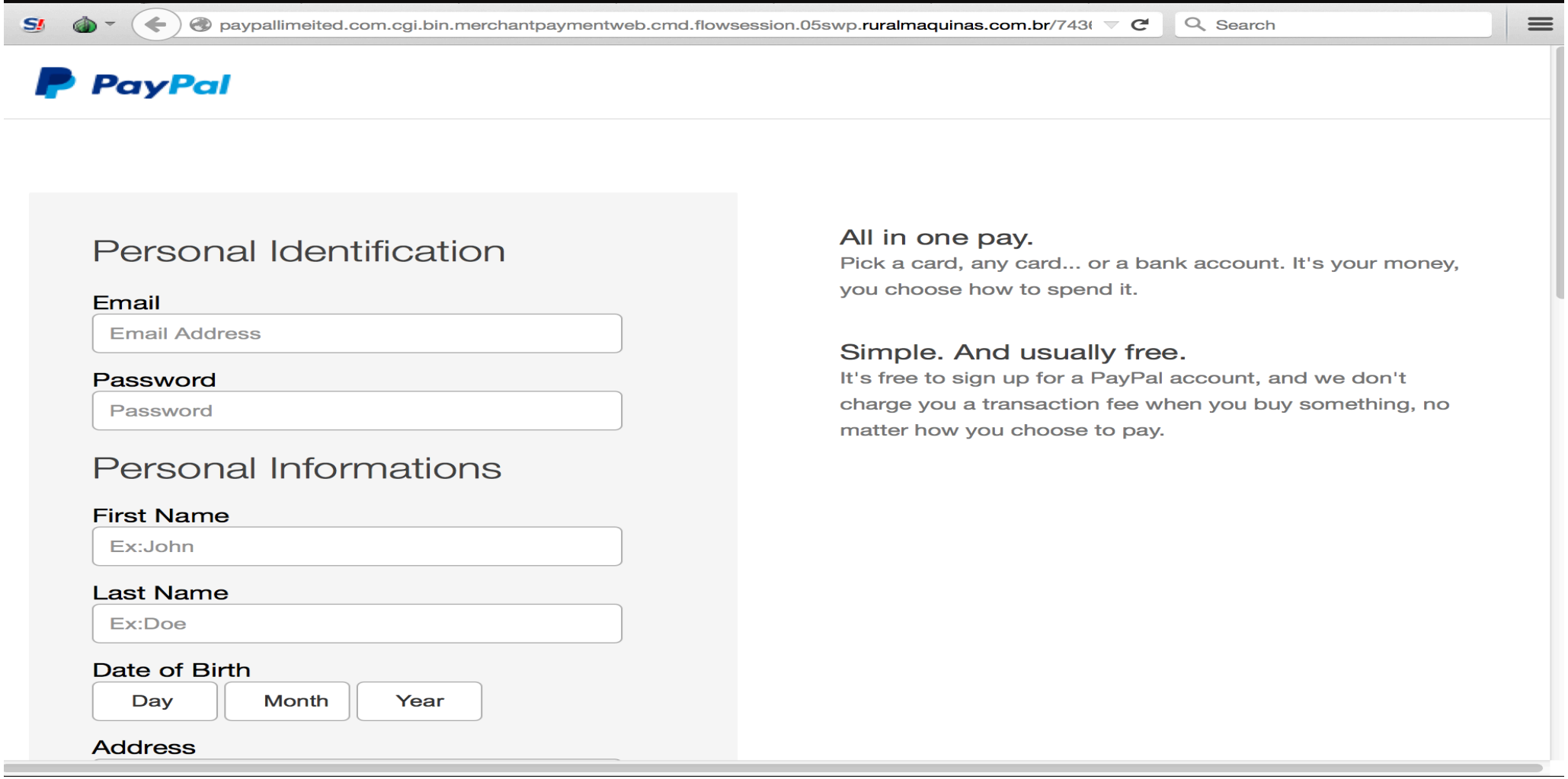
Domain Name	Security Categories	Content Categories	Last Observed
dhlpackagecentresecuredlink.info			Current
googlesecuredsecuritcentre.info			Current
mightyhighband.com			Current
realquickverifiedserv.info			Current
samueljwelch.com			Current
secured-google.com			Current
simplefmtservice.com			Current
welchtuningsystems.com			Current
welchtuningsytems.com			Current

Dare I say predictive??



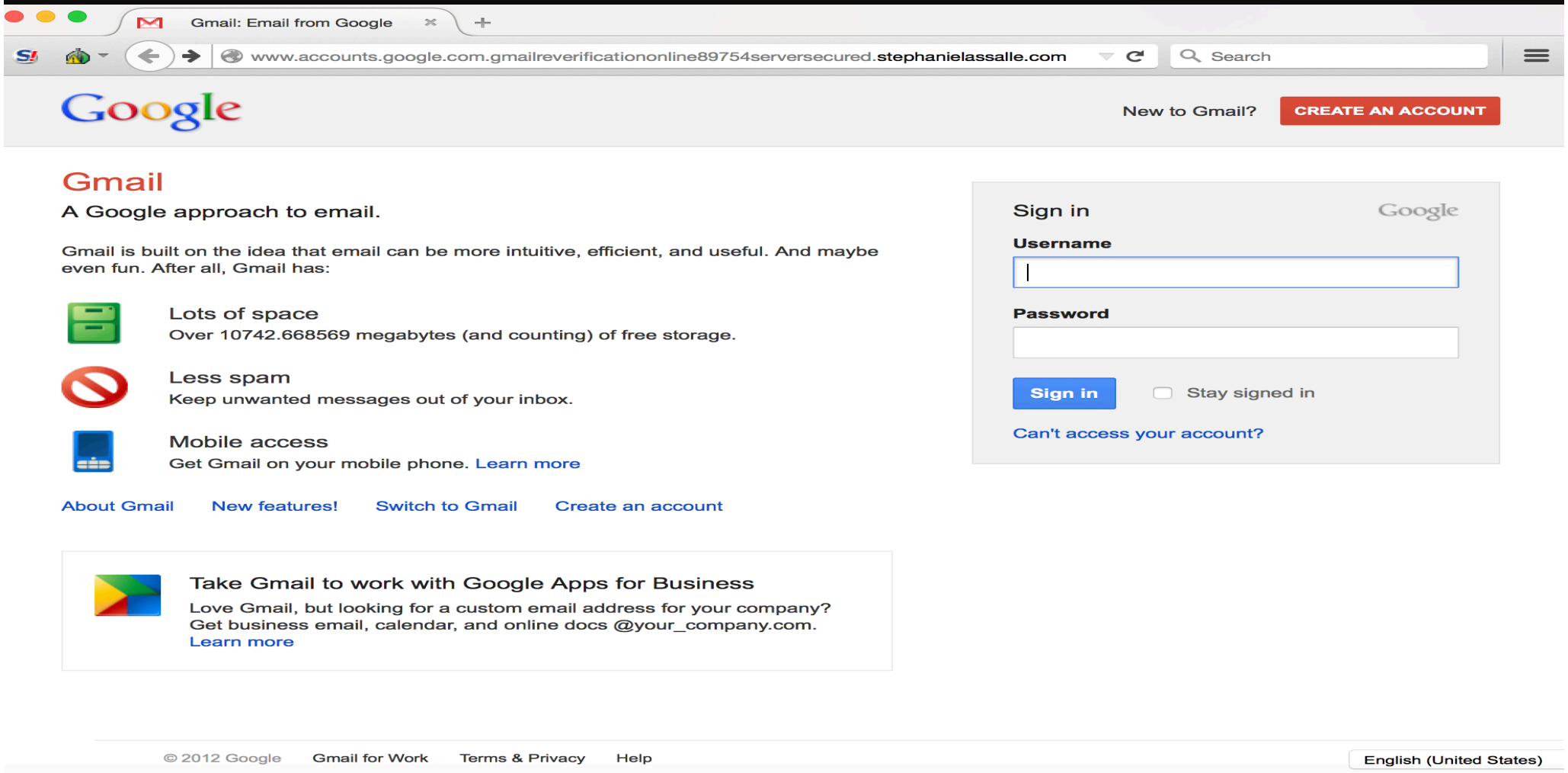
COMPROMISED SAMPLES

2016-03-04T04:30:09.099Z
paypallimeited[.]com[.]cgi[.]bin[.]merchantpaymentweb[.]cmd[.]flowsession[.]05swp[.]ruralmaquinas[.]com[.]br
0.9028750658035278 Paypal/scoalabranzei.ro.html




Selamat data...
Faceboo...
Gmail
iCloud
Yahoo - login
Facebook - L...
Rural ...
Update A...
ruralmaquinas.com.br
Search
www.ruralmaquinas.com.br
Rural Máquinas
Fazendo mais por você!
Siga-nos:
Início
A Empresa
Produtos
Assistência Técnica
Máquinas Usadas
Notícias
Orçamento
Contato
NOGUEIRA
Essencial no campo
1
2
3
4
inroda
MFW
DeLaval
IKEDA
Mais verde na terra !
NOGUEIRA
weg
Informativo - News
Limpeza do Equipamento de
Gerador de Energia
Conforto Animal
ruralmaquinas.com.br/index.php

2016-03-26T10:47:26.379Z
www[.]accounts[.]google[.]com[.]gmailreverificationonline89754serversecured[.]stephanielassalle[.]com
0.9815937280654907 Google/www.whiskey-memoirs.com.html



stephanielassalle.com/Site/Bienvenue.html

Search

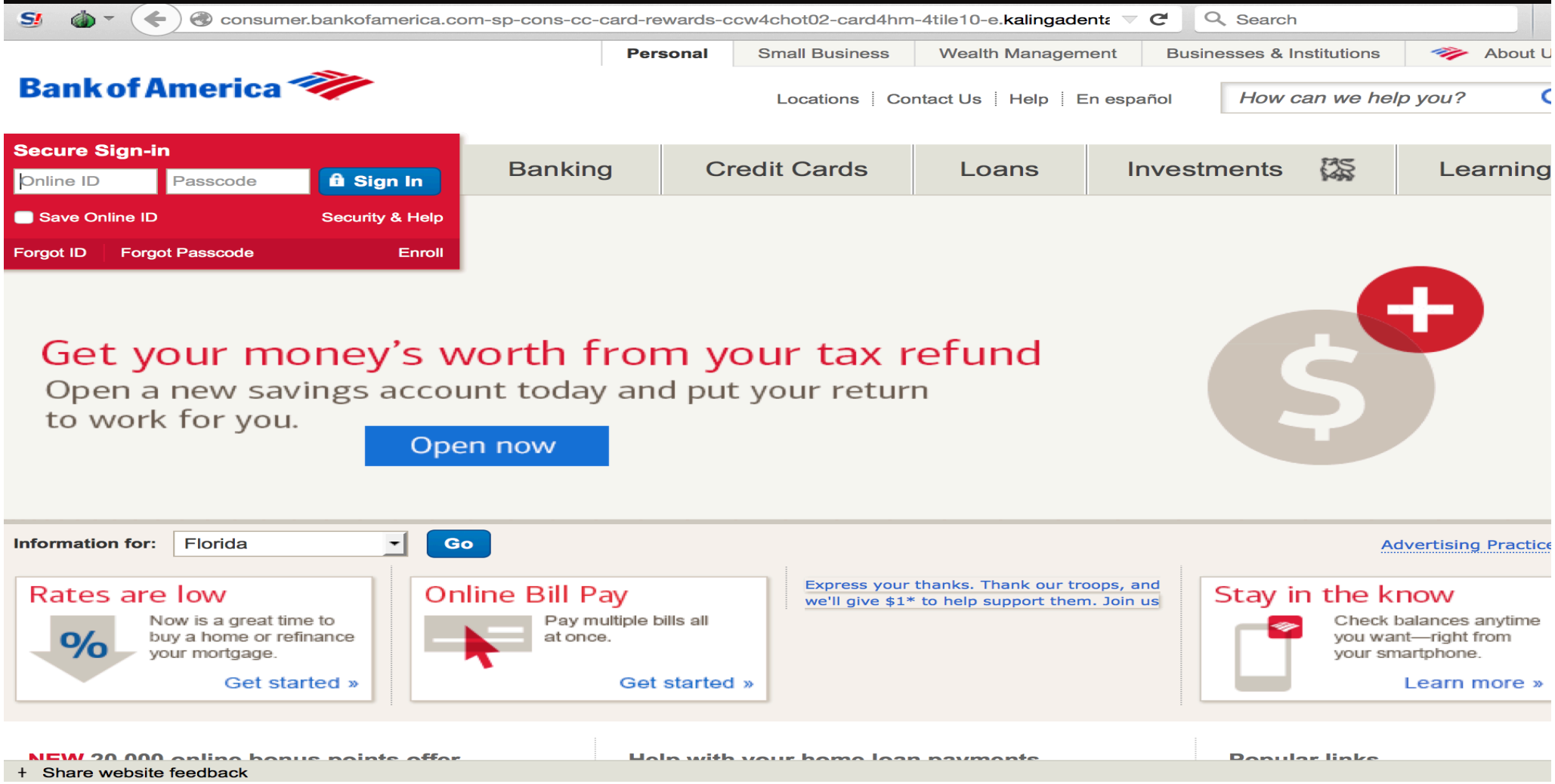
STÉPHANIE
LASSALLE

Photographie

ENTRER

Français

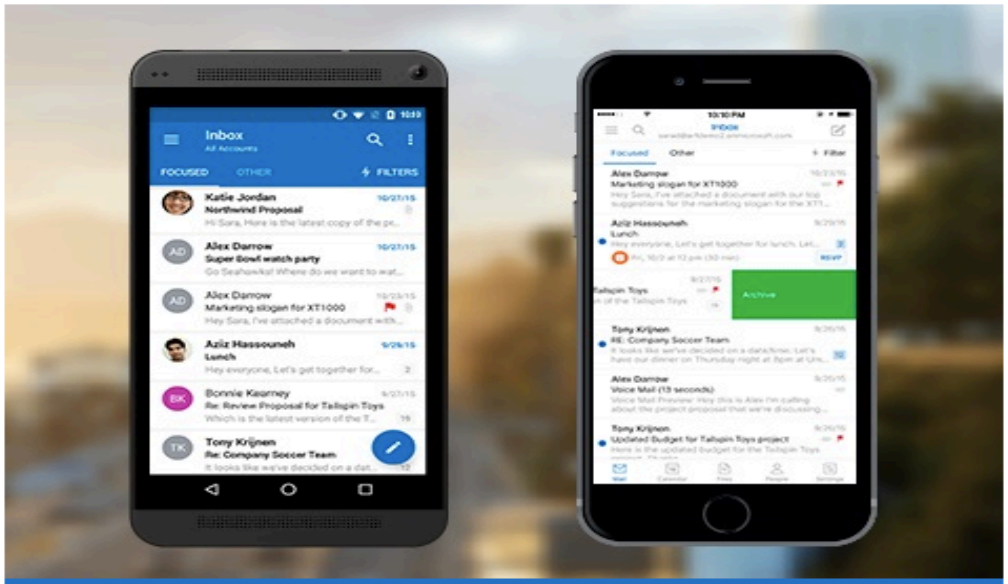
2016-03-24T04:10:53.709Z consumer[.]bankofamerica[.]com-sp-cons-cc-card-rewards-ccw4chot02-card4hm-4tile10-e[.]kalingadentalcare[.]com 0.9887999892234802 BankOfAmerica/goodhealthyeating.com.html



2016-03-11T10:21:30.644Z microsoft-outlook[.]ogspy[.]net 0.9925749897956848 Outlook/amazonripoff.com.html

outlook-microsoft.ogspy.net/1ced7d30a4d717b3ccc4f7eb36dfe0ff/

Search



Emportez Outlook avec vous

Vivez une expérience de messagerie optimale sur appareils iOS, Android et Windows.

En savoir plus

Outlook

Compte Microsoft Qu'est-ce que c'est?

Adresse e-mail ou Telephone

Mot de Passe

Maintenir la connexion

Se connecter

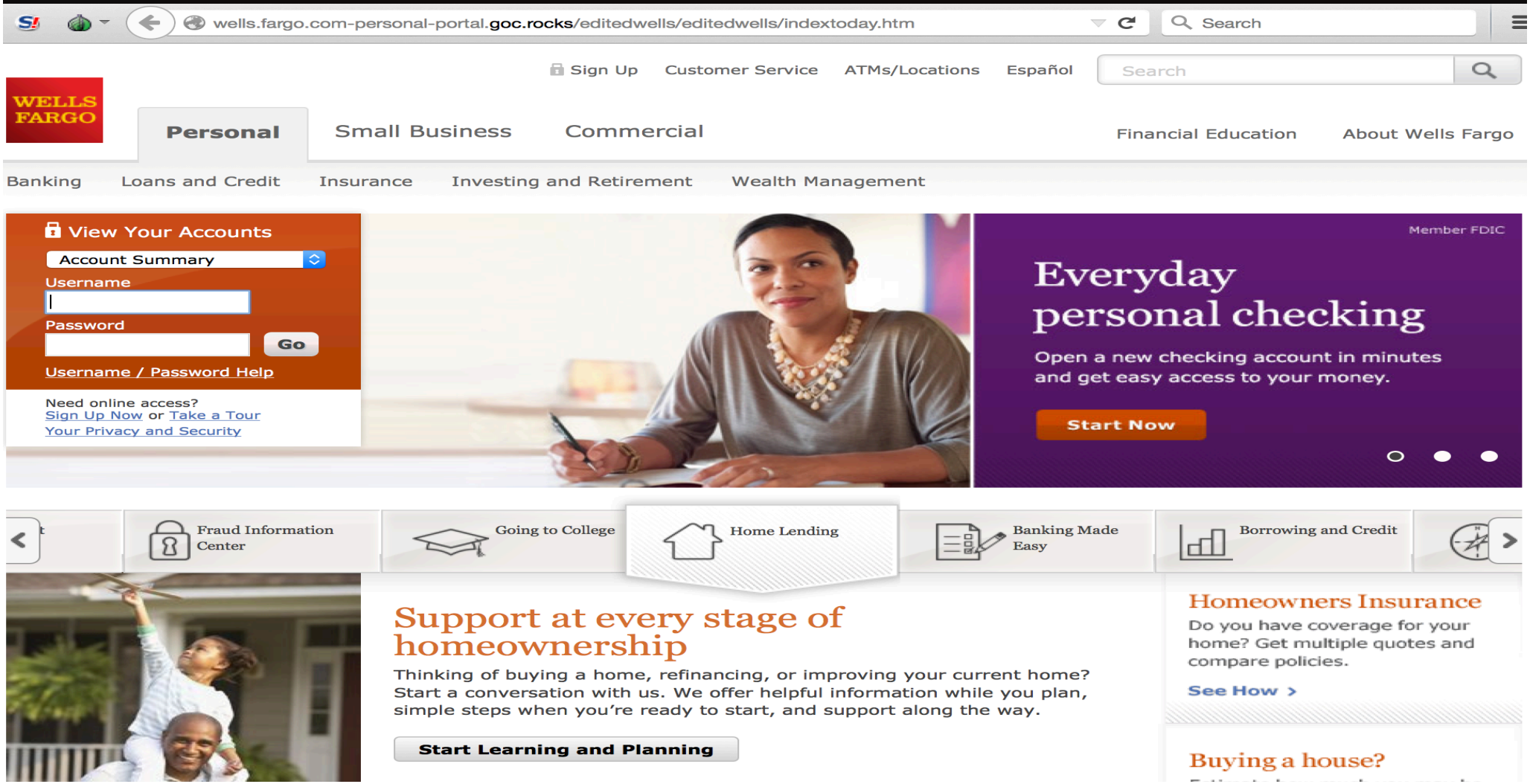
Votre compte n'est pas accessible?

Se connecter avec un code a usage unique

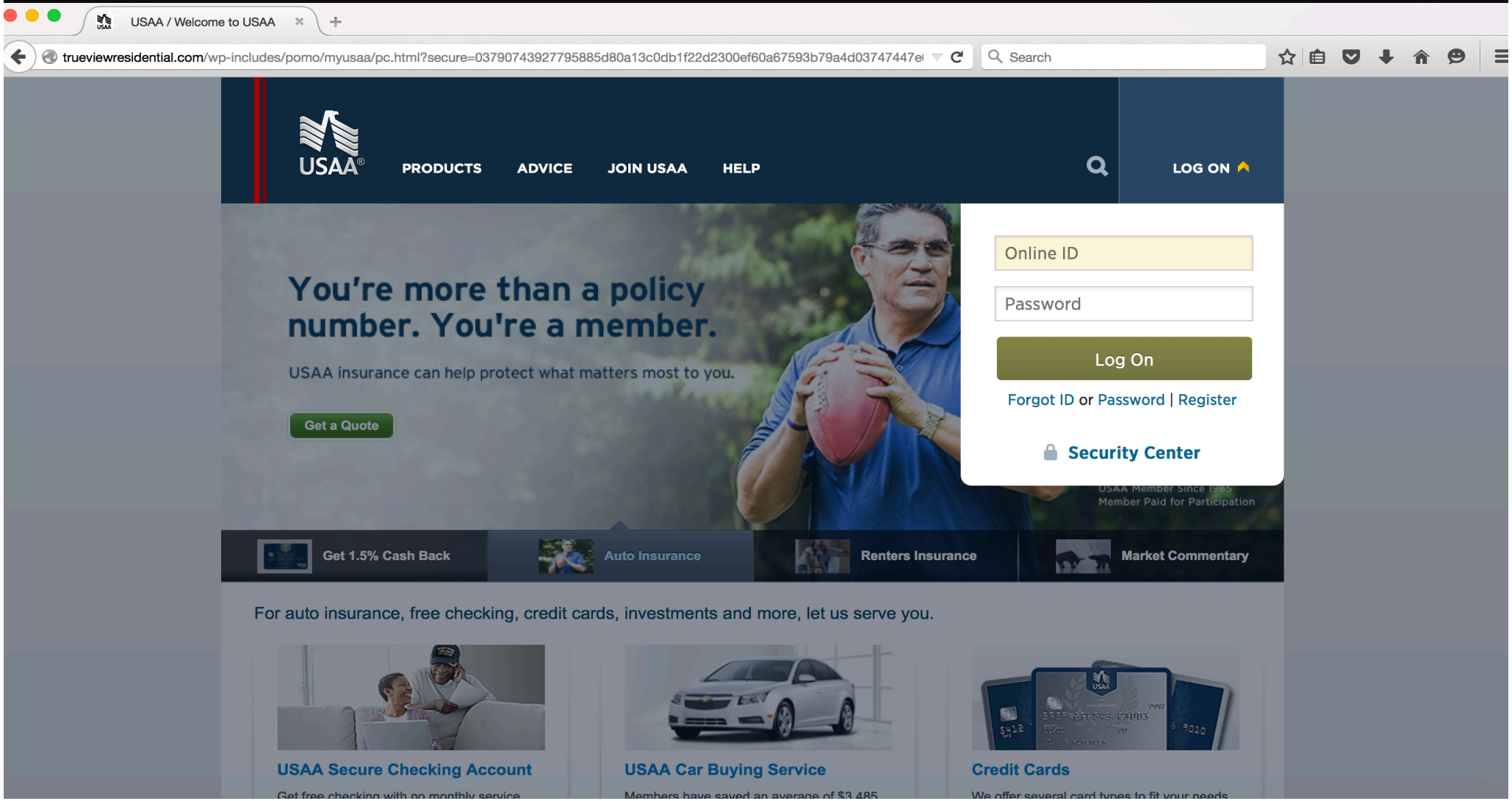
Vous n'avez pas encore de compte Microsoft?

Cree un compte maintenant

2016-02-28T16:11:08.036Z wells[.]fargo[.]com-personal-portal[.]goc[.]rocks 0.93414705991745
WellsFargo/www.sharprocket.com.ph.html



2016-03-30T10:52:18.824Z usaa[.]paramountweb[.]com[.]au 0.9032782316207886 USAA/gselitefitness-warrington.co.uk.htm



www.paramountweb.com.au

Search

PARAMOUNT
WEBSITE DESIGN

Home | About | Design | Portfolio | Marketing | Contact

2016-03-16T10:18:39.161Z dropbox[.]oktube[.]biz 0.8946635723114014 Dropbox/aquamarinedesalinators.com.html

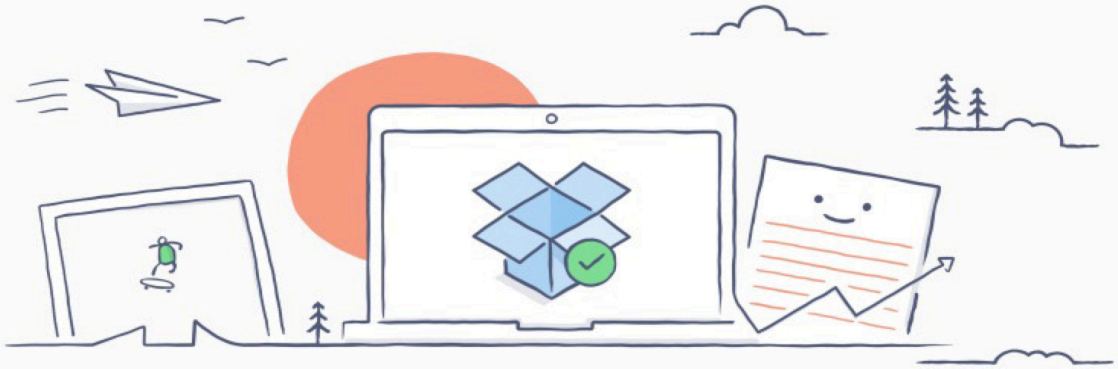
dropbox.oktube.biz

Search

Try Dropbox Business

Dropbox

Download the app • Sign in



Dropbox works the way you do

Get to all your files from anywhere, on any device, and share them with anyone.

Full name

Email

Password

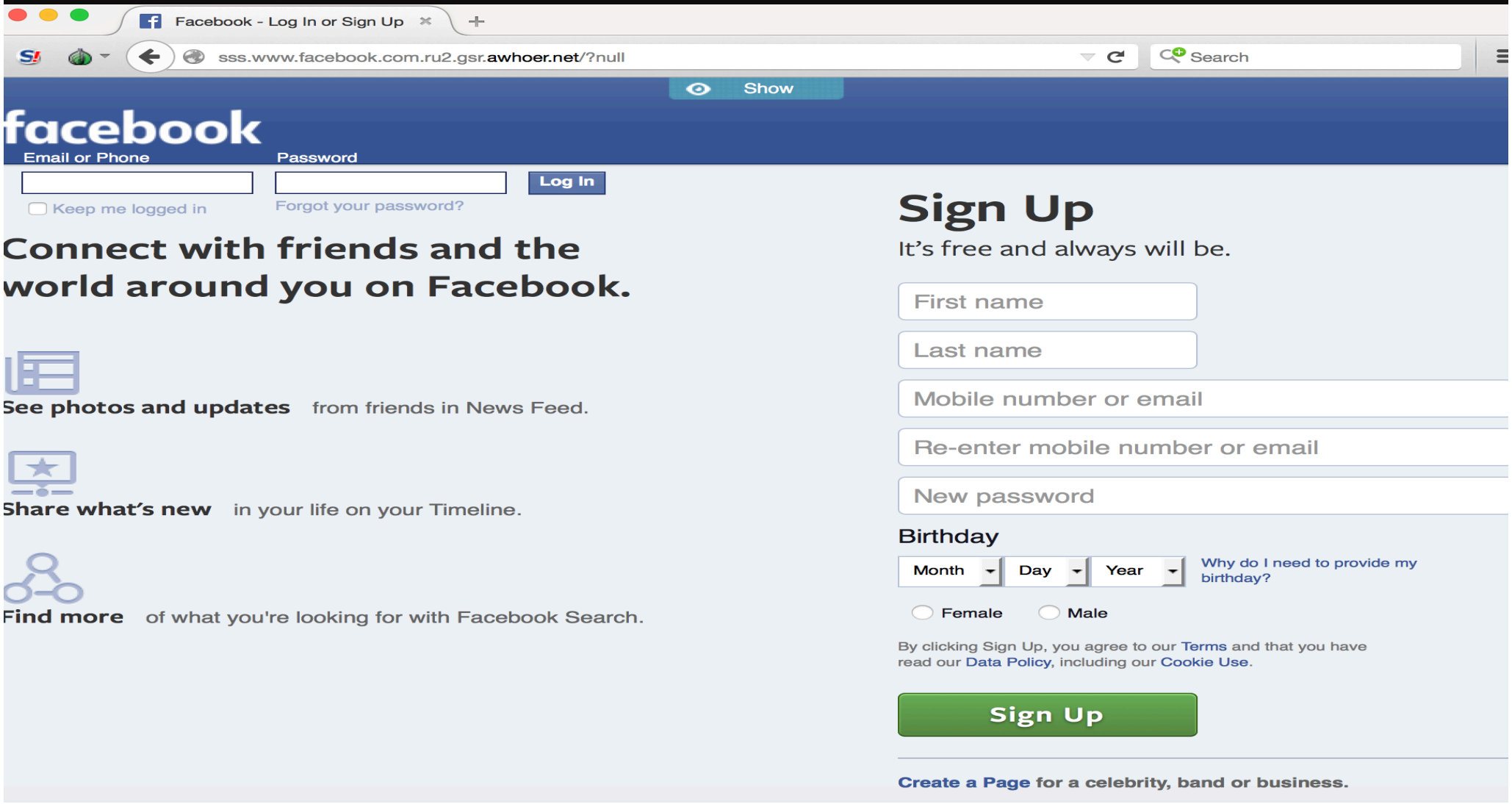
I agree to Dropbox terms

Sign up for free

or

Sign up free with Google

2016-03-30T05:03:42.960Z sss[.]www[.]facebook[.]com[.]ru2[.]gsr[.]awhoer[.]net 0.9346507787704468 Facebook/facebook.com.html



FALSE POSITIVES

```
Wed Mar 02 2016 10:28:19 GMT+0000 (UTC)

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/appleidverifysuport.com.html
2016-03-02T04:14:22.609Z pacebook[.]online 0.9935976266860962 Facebook/khansports.net.html
2016-03-02T04:01:19.971Z bankofameica[.]com 0.9831417202949524 BankOfAmerica/janwill.co.ke.html
2016-03-02T04:07:35.164Z bankofameric[.]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 bankofameria[.]com 0.9831417202949524 BankOfAmerica/janwill.co.ke.html
2016-03-02T04:24:50.095Z wellsafargo[.]com 0.9378761053085327 WellsFargo/wellsfargo.com.html
2016-03-02T04:26:46.840Z welllsfargo[.]com 0.9378761053085327 WellsFargo/wellsfargo.com.html
2016-03-02T04:22:28.324Z wellsfaargo[.]com 0.9378566145896912 WellsFargo/wellsfargo.com.html
2016-03-02T04:27:51.324Z wellsfarago[.]com 0.9378566145896912 WellsFargo/wellsfargo.com.html
2016-03-02T04:27:10.762Z wellsgfargo[.]com 0.9376039505004883 WellsFargo/wellsfargo.com.html
2016-03-02T04:27:28.924Z wellssfargo[.]com 0.9375162720680237 WellsFargo/wellsfargo.com.html
2016-03-02T04:22:43.132Z wellsfarog[.]com 0.9375162720680237 WellsFargo/wellsfargo.com.html
2016-03-02T04:27:52.661Z wellsfsrgo[.]com 0.9375162720680237 WellsFargo/wellsfargo.com.html
2016-03-02T04:22:16.901Z wallsfargo[.]com 0.9375162720680237 WellsFargo/wellsfargo.com.html
2016-03-02T04:27:50.260Z wellfargobank[.]com 0.9375162720680237 WellsFargo/wellsfargo.com.html
2016-03-02T04:29:40.807Z welsfargo[.]com 0.9375162720680237 WellsFargo/wellsfargo.com.html
2016-03-02T04:28:28.860Z wwwwellsfargo[.]com 0.9373353123664856 WellsFargo/wellsfargo.com.html
2016-03-02T04:29:44.988Z welssfargo[.]com 0.9371532797813416 WellsFargo/wellsfargo.com.html
2016-03-02T04:27:52.220Z wellsfargow[.]com 0.9369886517524719 WellsFargo/wellsfargo.com.html
2016-03-02T04:29:31.825Z wellsfergo[.]com 0.9369886517524719 WellsFargo/wellsfargo.com.html
2016-03-02T04:15:28.254Z ellsfargo[.]com 0.936801016330719 WellsFargo/wellsfargo.com.html
2016-03-02T04:22:28.324Z wellsfago[.]com 0.936801016330719 WellsFargo/wellsfargo.com.html
2016-03-02T04:26:44.585Z weellsfargo[.]com 0.9366189241409302 WellsFargo/wellsfargo.com.html
2016-03-02T04:28:16.260Z wellsfaro[.]com 0.9361376166343689 WellsFargo/wellsfargo.com.html
2016-03-02T04:22:22.710Z seguridad[.]dsal-apple[.]com 0.8931735754013062 Apple/randoserc.com.html
2016-03-02T04:15:16.043Z mcms[.]innovicloud[.]com 0.8765116930007935 Paypal/www.paypal-united.com.html

Wed Mar 02 2016 10:28:22 GMT+0000 (UTC)
```

https://investigate.opendns.com/domain-view/name/wellsafargo.com/view

Search

wellsafargo.com INVESTIGATE

Visualize

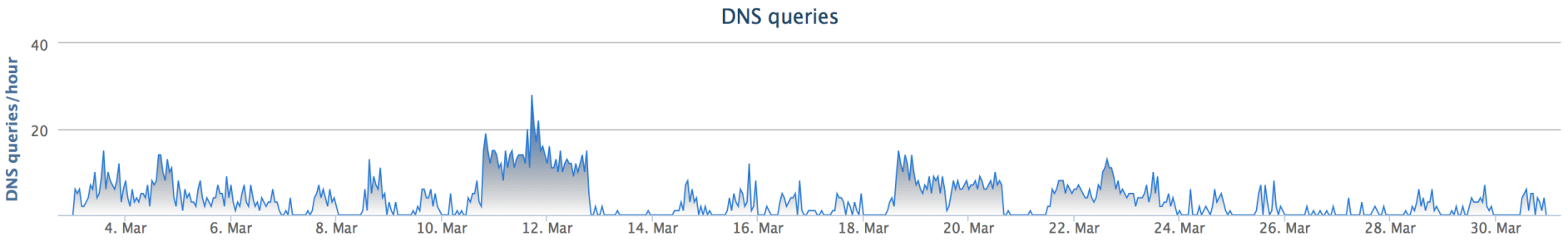
DETAILS FOR WELLSAFARGO.COM

Search in Google

Search in VirusTotal

Classifier prediction: benign

OpenDNS Security Graph Score: +100



WHOIS RECORD DATA

Registrar Name: MarkMonitor, Inc. **IANAID:** 292

Last retrieved March 31, 2016 Get latest

Created: October 11, 2002 **Updated:** August 28, 2014 **Expires:** October 11, 2016 Raw data

Email Address	Associated Domains	Email Type	Last Observed
hostmaster@wellsfargo.com	Greater than 500 Total	Administrative, Registrant, Technical	Current

Show past data

Showing 1 of 2 Results

play.googles.re 0.9210382103919983, Paypal/avantebus.com.br.html

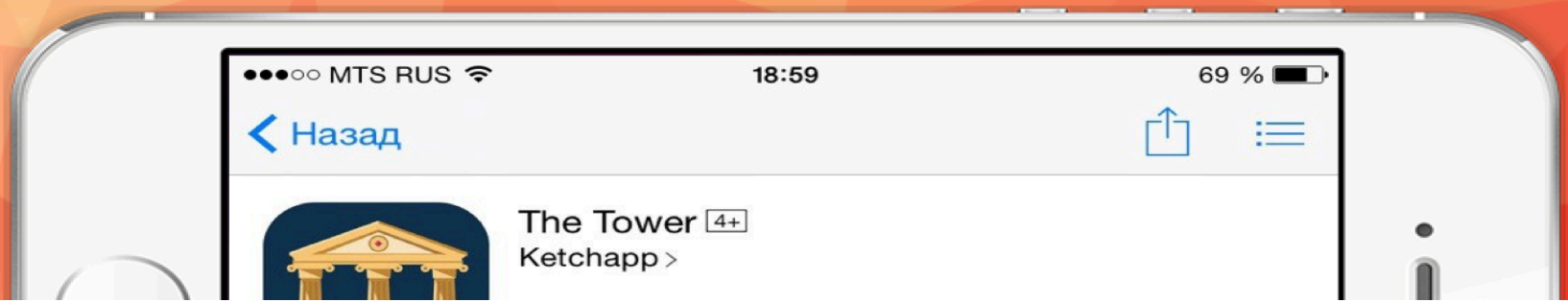
payforinstall.ru

Search

Pay For Install

РЕКЛАМОДАТЕЛЯМ КОНТАКТЫ ЛИЧНЫЙ КАБИНЕТ

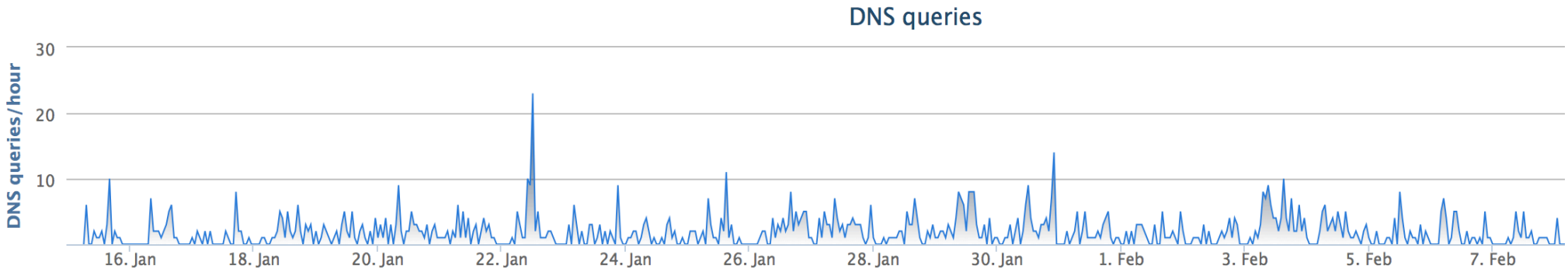
Мы представляем вам уникальный сервис с помощью которого вы сможете быстро и легко заработать



DETAILS FOR PLAY.GOOGLES.RE

Classifier prediction: benign

OpenDNS Security Graph Score: +100



WHOIS RECORD DATA

Registrar Name: TLD Registrar Solutions Ltd **IANAID:** 1564

Last retrieved Febru

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

Email Address	Associated Domains	Email Type	Last Observed
topxaker@gmail.com	7 Total - 1 malicious	Administrative, Registrant	Current
AfnicRoleObject@internet.bs	9 Total - 3 malicious	Technical	Current

SEARCH PATTERN SEARCH

topxaker@gmail.com

INVESTIGATE

Visualize

DOMAINS ASSOCIATED WITH TOPXAKER@GMAIL.COM

Domain Name	Security Categories	Content Categories	Last Observed
rr3.us	Drive-by Downloads/Exploits		Current
app1e.us			Current
apples.re			Current
googles.re			Current
miniopera.us			Current
rovio.us			Current
skyperu.us			Current
swizard.us			Current

pyplusers.com 0.9960060715675354, Paypal/bluespruceus.com.html
googledraive.com 0.9555937051773071, Google/www.rotisseriebuongusto.com.br
idmappleinc.info 0.9283350110054016, Apple/www.lcloudid-ds.top.html
supportapples.com 0.9700624942779541, Apple/santuariodocorpo.com.html
wellllsfargoservice.goldenhomes.co.za 0.9379003047943115, WellsFargo/wellsfargo.com.html
facebooksytfa.tr.gg 0.9258965253829956, Facebook/facebook.com.html
facebookcomments.netne.net 0.9551351070404053, Facebook/facebook.com.html
google-gems.pe.hu 0.9421336650848389, Google/gmailstealer.tk.html
isdmaapple-step2.com 0.9403073787689209, Apple/carsdrive.com.ar.html
lindungifacebookmu.tk 0.9336977601051331, Facebook/facebook.com.html
atfaceebooktt45128765670988900872345766798598346530989O.notiexpreesht.com, 0.9466191530227661, facebook.com.html
www.ifacebok.pl 0.9442495107650757, Facebook/facebook.com.html
apple.idevicetrackers.com 0.9769512414932251, Apple/www.lcloudid-ds.top.html
mobile-bankofamerica.com.ssnverify.user5cnkktq4at3.ssnid6530.us 0.9848669171333313, BankOfAmerica/janwill.co.ke.html
mobile-bankofamerica.com.ssnverify.userkbvjzqes28m.ssnid6530.us 0.9848620295524597, BankOfAmerica/janwill.co.ke.html
mobile-bankofamerica.com.ssnverify.userewdjrdhekze.ssnid6530.us 0.9848669171333313, BankOfAmerica/janwill.co.ke.html
free-facebook.gq 0.9977947473526001, Facebook/khansports.net.html
transaction.paypalcom.verification.global.kimora-fashion.com 0.9959402084350586, Paypal/dinastiaacessorios.com.br.html
bankofamerica.com.ssnid6530.us 0.9848620295524597, BankOfAmerica/janwill.co.ke.html
fac3book2.hostingsiteforfree.com 0.9887204170227051, Facebook/faceb000k.com.html
ymfacebook.ml 0.9958401918411255, Facebook/khansports.net.html
www.facebook-page-security.somee.com 0.9825869798660278, Facebook/www.secure.logs-facebook.com.html

STACKING MODELS

2016-03-04T10:39:38.372Z www[.]icloud-safri[.]com 0.9479695558547974 Apple/applelock.ru.html

https://investigate.opendns.com/domain-view/name/www.icloud-safri.com/view

Search

www.icloud-safri.com INVESTIGATE Visualize Back to top

DETAILS FOR WWW.ICLOUD-SAFRI.COM

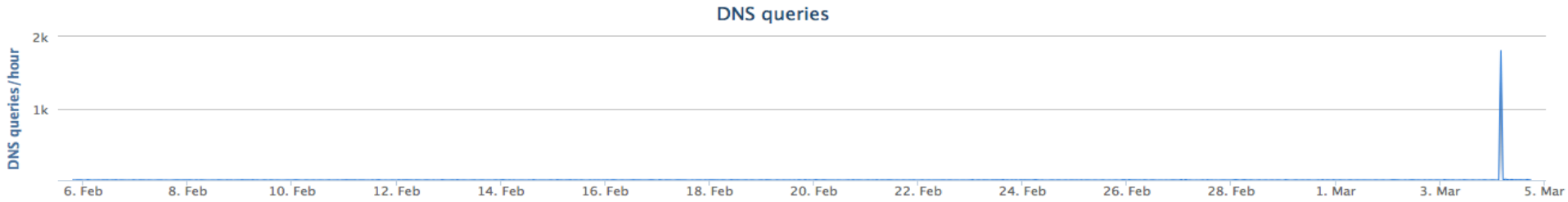
Search in Google

Search in VirusTotal

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

Last retrieved March 4, 2016 Get latest

Created: October 15, 2015 **Updated:** October 15, 2015 **Expires:** October 15, 2016 Raw data

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

http://facebook[.]com.accounts[.]login[.]userid[.]280964[.]2nso3s[.]com/we
next=http%3A%2F%2Fwww.facebook.com%2videos%2F%3A%4A%4ID%1

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

```
<!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

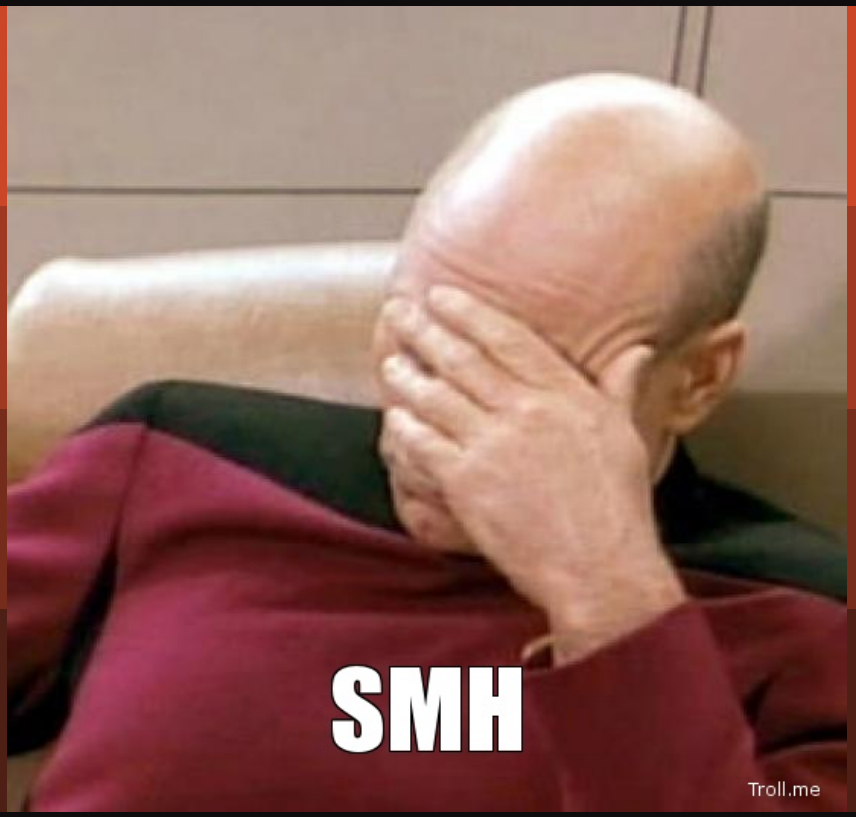
Whois information

Registration date	2013-09-03 00:00:00 +0000
Registrar name	GODADDY.COM, LLC
Registrant	li ning < li2384826402@yahoo.com >
Registrant contact address	guangdongsheng guangzhoushi Alabama UNITED STATES

21,533 DOMAINS???

crowcasinovip.biz mybestbrand.biz mybestbrands.biz huarenceluewangzhi.com icbczay.com boyinbocai5.com
haoyunc3.com bocaiwangzhenqianpingtai.com zuqiubocaiwangzhan7.com weinisirenyulecheng94.com
xinquanxunwang244.com dfjdh.com yaojiyulecheng9898.com wanbaoluyulecheng94.com xinpujingyule15.com
toabaao.com jinbaiyiyulecheng26.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 zhucecaipiaosongcaijin.com ceoylcdf.com
zhucesongcaijindewangzhan62.com aomenduchangyouxiyounaxie30.com mengtekaluoylcb.com
guojihuangguanyule40.com huangguantiyupingtai93.com huangguanxianjinwangxinyu37.com
aomenduchangpaixing27.com 500wanylcyu.com dajihuiylc686.com ruifengguojiyy.com makeboluoylcb.com
jincaigjylc.com xindongfangylc869.com aomenduchangzainali50.com wangshangyulekaihusongcaijin.com
huangguanxjwkh.com jinbangylc77.com baijialeqo.com yataigjylc.com baishenggjylcwe.com bocaigongsiqe.com
wufagjylc.com moerbenylckk.com bogouylc1663.com huangguandailiwangzhi23.com bojueylcpo.com
bocaiwangzhanqe.com taoataao.com bbhunas.com sjzd36.com sjpt63.com bjlkh33.com
baijialebishengtouzhujiqiao20.com xijialiansaijifenbang57.com baijialeyule86.com xijiapaiming46.com
aomenbaijialechangying76.com baijialeyulepingtai34.com wangshangbaijialekaihusongcaijin76.com
ouzhouwudaliansaipaiming53.com wudaliansaitedian39.com baijialekaihusong50caijin17.com baijialeguize52.com
zhibobazuqiuzhibo2.com zuqiubifenqiutan88.com dejiasaichengbiao88.com zuqiuba85.com mahuitqzzjw83.com
sjzd01.com weixingjianting29.com cwanpp.com xingboyulezaixian86.com mwqpah.com
jiankongpingtairuanjian43.com zhenqianyulechengguanwang63.com njdyyytj.com fanheer.com 999coin.com
shenganna74.com jackwolfskinsalejp.com zaozhuangcq.com bjl7788.com ruhejiankongshouji2.com
aomenduchangyingqianliao75.com shoujidingweichaxunruanjian12.com shoujijiantingshebei46.com aomen916.com
shoujikajiantingqi77.com zhenqianyouxipaixing2.com rysevw.com wanzhenqianwangzhan36.com vrcgw.com
feilvbinshengannayulecheng20.com duchangyingqianmijue81.com zzvqo.com
wangshangzhenqianzhajinhuayouxi62.com zql2014.com bocwtr.com baojiyulezaixian87.com pacyfjg.com

OpenDNS


SMH
Troll.me

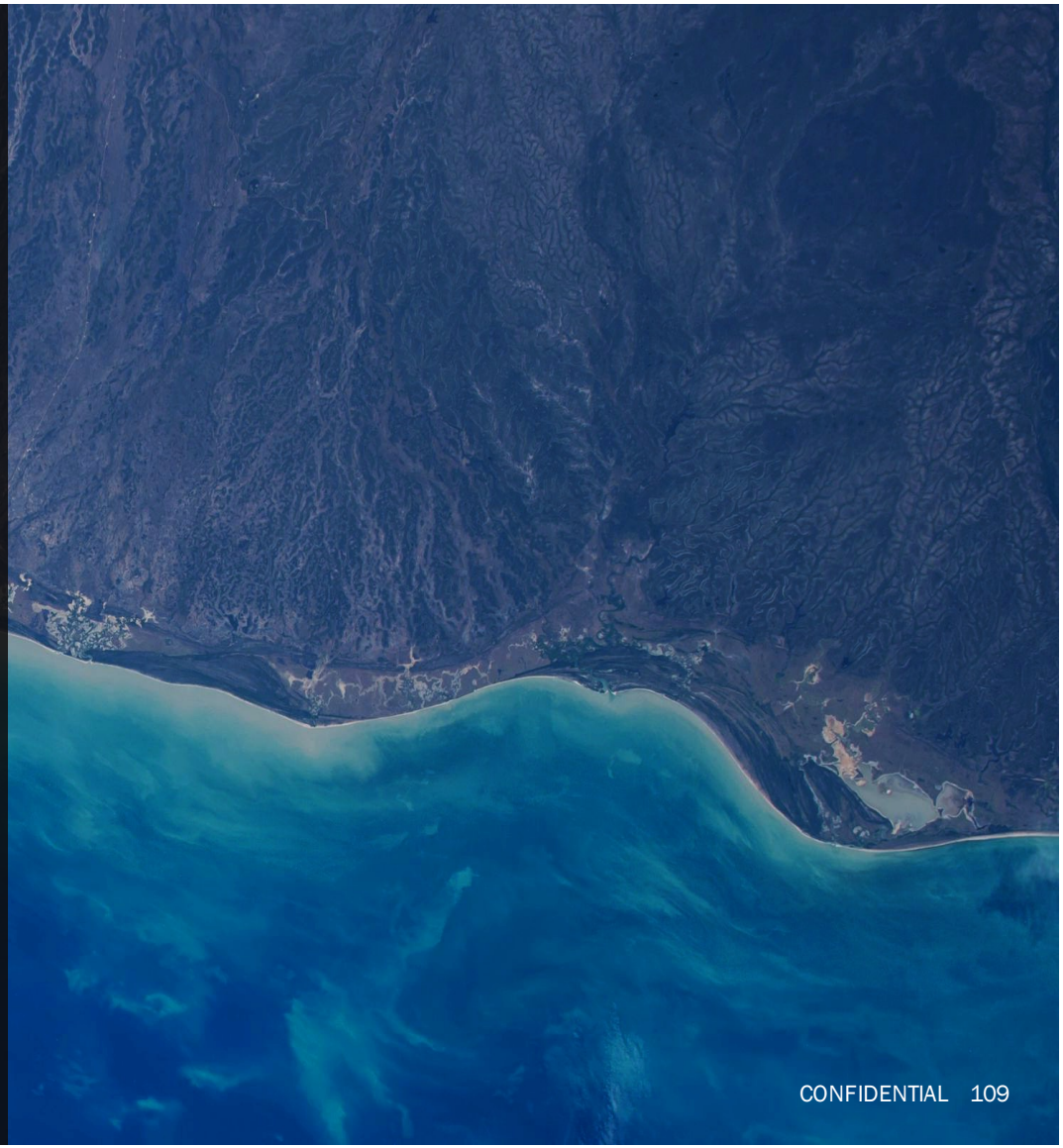
SAKULA / THREATCONNECT REPORT

1	Domain Name: TOPSEC2014.COM	1	Domain Name: TOPSEC2014.COM
2	Registry Domain ID: 1857525015_DOMAIN_COM-VRSN	2	Registry Domain ID: 1857525015_DOMAIN_COM-VRSN
3	Registrar WHOIS Server: whois.godaddy.com	3	Registrar WHOIS Server: whois.godaddy.com
4	Registrar URL: http://www.godaddy.com	4	Registrar URL: http://www.godaddy.com
5	Update Date:	5	Update Date: 2014-05-06 04:52:21
6	Creation Date: 2014-05-06 04:48:49	6	Creation Date: 2014-05-06 04:48:49
7	Registrar Registration Expiration Date: 2015-05-06 04:48:49	7	Registrar Registration Expiration Date: 2015-05-06 04:48:49
8	Registrar: GoDaddy.com, LLC	8	Registrar: GoDaddy.com, LLC
9	Registrar IANA ID: 146	9	Registrar IANA ID: 146
10	Registrar Abuse Contact Email: abuse@godaddy.com	10	Registrar Abuse Contact Email: abuse@godaddy.com
11	Registrar Abuse Contact Phone: +1.480-624-2505	11	Registrar Abuse Contact Phone: +1.480-624-2505
12	Domain Status: ok	12	Domain Status: clientTransferProhibited
		13	Domain Status: clientUpdateProhibited
		14	Domain Status: clientRenewProhibited
		15	Domain Status: clientDeleteProhibited
13	Registry Registrant ID:	16	Registry Registrant ID:
14	Registrant Name: li ning	17	Registrant Name: Top Sec
15	Registrant Organization:	18	Registrant Organization: TopSec
16	Registrant Street: guangdongsheng	19	Registrant Street: china
17	Registrant City: guangzhoushi	20	Registrant City: china
18	Registrant State/Province: Alabama	21	Registrant State/Province: china
19	Registrant Postal Code: 54152	22	Registrant Postal Code: 100000
20	Registrant Country: United States	23	Registrant Country: China
21	Registrant Phone: +1.4805428751	24	Registrant Phone: +1.82776666
22	Registrant Phone Ext:	25	Registrant Phone Ext:
23	Registrant Fax:	26	Registrant Fax:
24	Registrant Fax Ext:	27	Registrant Fax Ext:
25	Registrant Email: li2384826402@yahoo.com	28	Registrant Email: TopSec_2014@163.com

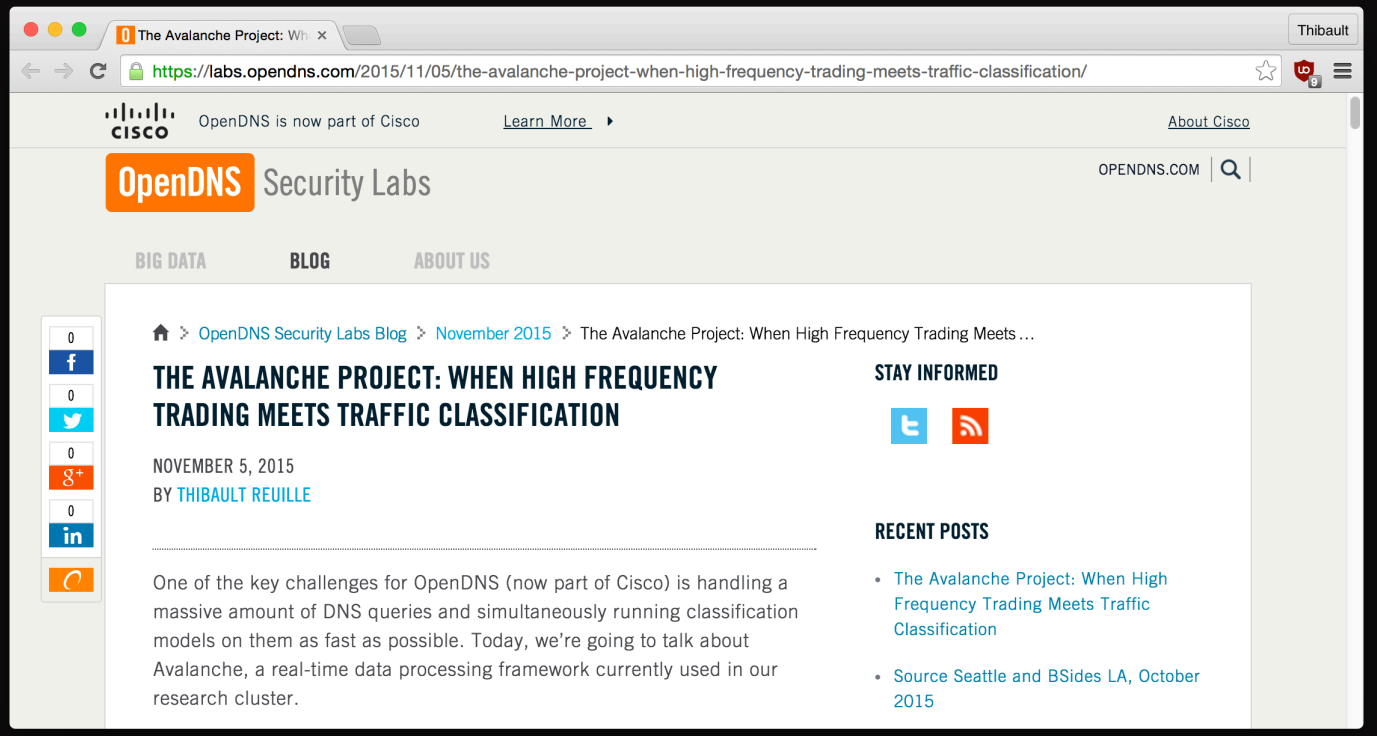
Future Work

- 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

CONCLUSION

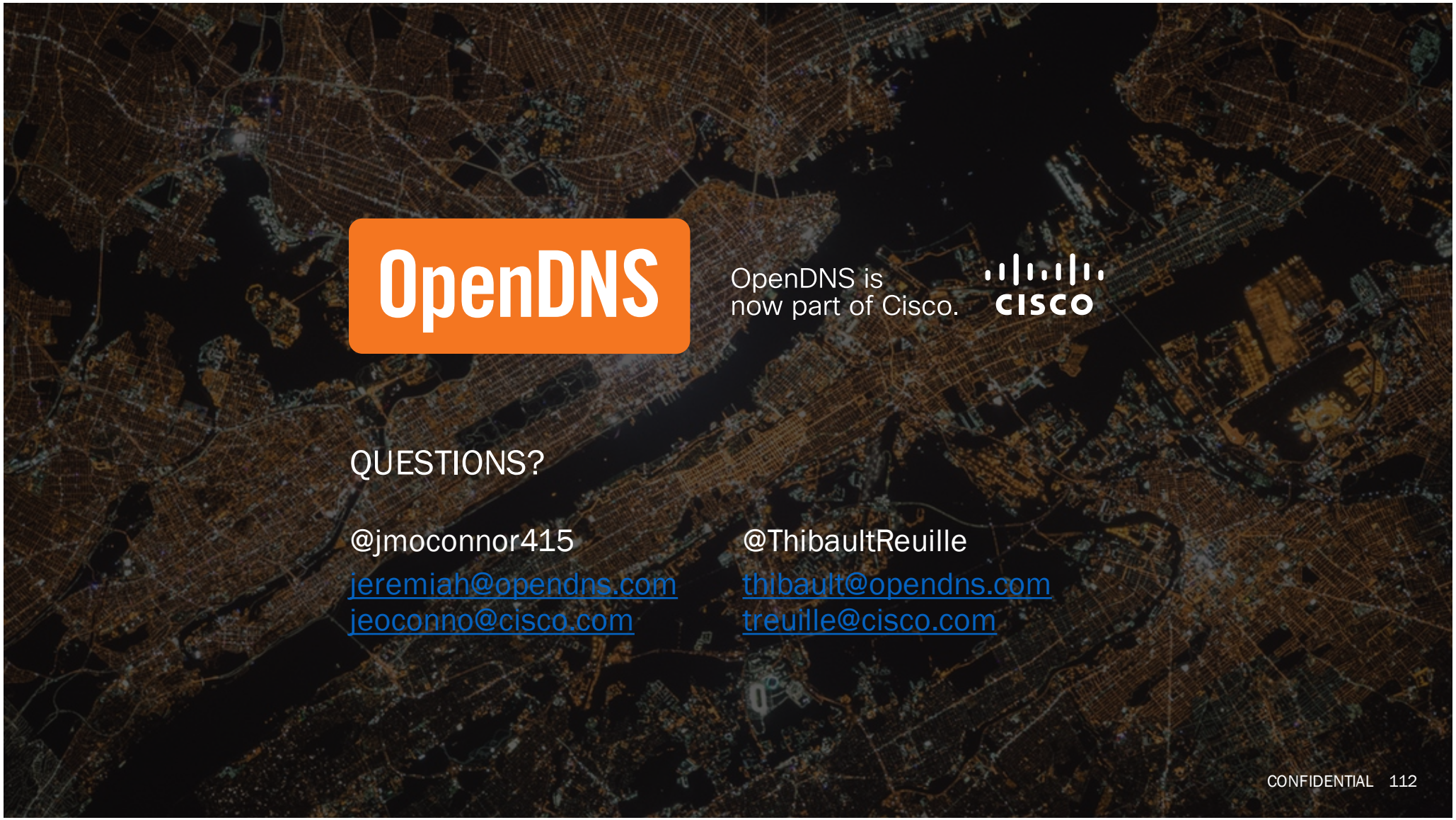


BLOG POST LIVE : labs.opendns.com



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


OpenDNS
OpenDNS is
now part of Cisco.
cisco
QUESTIONS?
@jmoconnor415
jeremiah@opendns.com
jeoconno@cisco.com
@ThibaultReuille
thibault@opendns.com
treuille@cisco.com