Intrusion Detection

Eka Renardi

Cyber Attacks News

Kremlin Hit by **Cyber Attacks**
eSecurity Planet - Sep 18, 2015
"Yesterday someone attempted to **hack** our website and alter the data ... "The discussion, though brief, of China's **cyber-attacks** on the U.S. in ...

China-based **Cyber Attacks** On US Military Are 'Advance…
Forbes - Sep 18, 2015
Today's followup post is here: Report: Chinese **Hackers** Used OPM Data ... Tiger: Exploring Chinese **Cyber** Espionage **Attacks** on U.S. Defense ...

Cyber attacks from China against the US may be slowing…
Business Insider - Sep 19, 2015
Government-supported **hackers** in China may have backed off recently as Chinese and U.S. officials began negotiating in earnest over **cyber** security ahead of ...

Hack Attack: Pentagon Kills Top ISIS **Cyber** Warrior
Foreign Policy (blog) - Aug 28, 2015
Dramatically shifting the types of militants it targets, the Pentagon confirmed that an American drone killed a prominent Islamic State **hacker** and ...
Rapid News Netw...

US, China appear close on **cyber** economic espionage deal
PCWorld - 4 hours ago
Hours later, The Wall Street Journal published an interview with President Xi, who denied any part in **hacking attacks** on U.S. businesses and ...
Is The U.S. And China's Cyberwar Reaching A Detente Or A **...**
In-Depth - TechCrunch - Sep 21, 2015
Explore in depth (2,258 more articles)

US University to spend $3 million on **cyber** security after …
ITProPortal - Aug 26, 2015
Rutgers University in New Brunswick is to spend up to $3 million (£1.9 million) on **cyber** security to prevent **hackers** crippling the university's ...

Types of Attacks

Active

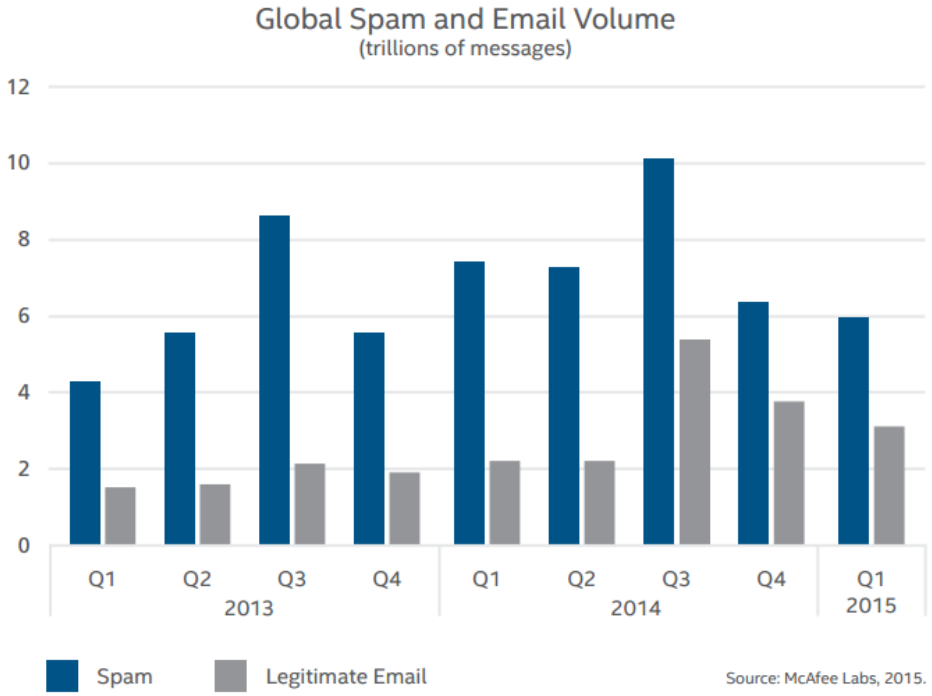
- Port scanner
- Wiretapping

Passive

- DDOS
- Man in the middle
- DNS spoofing
- Buffer overflow
- Format string attack
- SQL injection
- Spam

Source: https://en.wikipedia.org/wiki/Network_security

Messaging and network threats



Source: http://www.mcafee.com/us/resources/reports/rp-quarterly-threat-q1-2015.pdf

The Question

How do I accurately detect that an intrusion is actually occurring, and not just an anomalous activity?

OR

How can I predict that this anomalous activity is an intrusion?

Example

Typical Users

http://server.com?query=book

Server

Malicious User

http://server.com?query=book;DROP TABLE USERS;

Dataset

Given sensitive nature of this, most companies are reluctant to provide the dataset, mine included :(

Public dataset for experimentation, http://users.aber.ac.uk/pds7/csic_dataset/csic2010http.html

Dataset < 500K rows, and labelled.

Sample Data

Columns

"index","method","url","protocol","userAgent","pragma","cacheControl","accept","acceptEncoding","acceptCharset","acceptLanguage","host","connection","contentLength","contentType","cookie","payload","label"

Method : PUT, GET, POST
URL: unique, only 1

Label

- “norm” - normal
- “anom” - anomalous

Sample Data

Sample Good Data

"7","GET","http://localhost:8080/tienda1/publico/pagar.jsp","HTTP/1.1","Mozilla/5.0 (compatible; Konqueror/3.5; Linux) KHTML/3.5.8 (like Gecko)","no-cache","no-cache","text/xml,application/xml,application/xhtml+xml,text/html;q=0.9,text/plain;q=0.8,image/png,*/*;q=0.5","x-gzip, x-deflate, gzip, deflate","utf-8, utf-8;q=0.5, *;q=0.5","en","localhost:8080","close","null","null","JSESSIONID=535529056CAF606E044AC8C76F6E6C40","modo=insertar","norm"

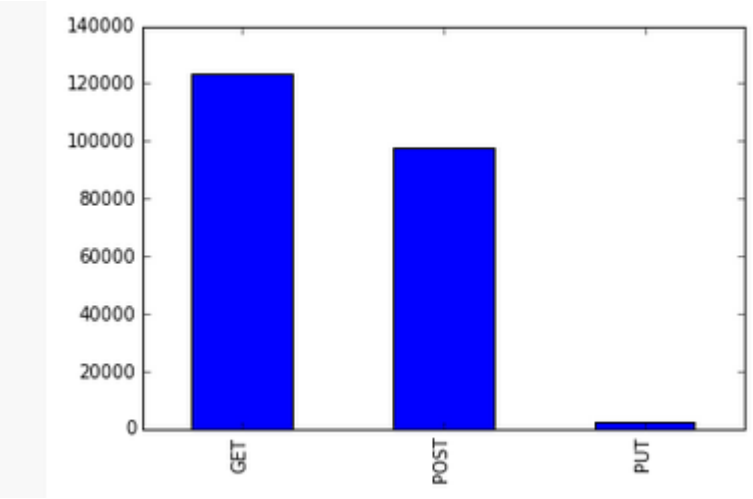
Sample Bad Data

"0","GET","http://localhost:8080/tienda1/publico/anadir.jsp","HTTP/1.1","Mozilla/5.0 (compatible; Konqueror/3.5; Linux) KHTML/3.5.8 (like Gecko)","no-cache","no-cache","text/xml,application/xml,application/xhtml+xml,text/html;q=0.9,text/plain;q=0.8,image/png,*/*;q=0.5","x-gzip, x-deflate, gzip, deflate","utf-8, utf-8;q=0.5, *;q=0.5","en","localhost:8080","close","null","null","JSESSIONID=B92A8B48B9008CD29F622A994E0F650D","**cantidad='; DROP TABLE usuarios; SELECT * FROM datos WHERE nombre LIKE '%"**,"anom"

"3144","GET","http://localhost:8080/tienda1/publico/registro.jsp","HTTP/1.1"....
*;q=0.5","en","localhost:8080","close","null","null","JSESSIONID=0474EC2F9E229A3DDDC7F7FCFB15B054","**ntc=9812245040414546','0','0','0','0');waitfor delay '0:0:15';--"**,"anom"

Exploratory Charts

GET, POST, PUT methods




```
In [134]: pd.crosstab(df.method, df.is_anom)
```

Out[134]:

is_anom	0.0	1.0
method		
GET	61998	61447
POST	41998	55939
PUT	0	2193

1st Model

TFIDF on the “payload”, and “is_anom” as response.

LogisticRegression classifier

Accuracy

```
In [157]: from sklearn.metrics import accuracy_score
          print("accuracy ", accuracy_score(y_test, predictions))

          accuracy  0.776505528321
```

Null Accuracy is around 0.50

Next Steps - Iterate

Better model (incorporate NLTK, additional classifiers, cross validation, etc)

Better charts