Malicious CDNs: Identifying Zbot Domains en Masse via SSL Certificates and Bipartite Graphs

Dhia Mahjoub and Thomas Mathew

Overview

- The goal of this talk is to provide a series of simple statistical methods that allow a researcher to identify ZBot domains using SSL data

- All of the data discussed is open source and can be obtained at scans.io/study/sonar.ssl (thanks rapid7!)

SSL

- **S**ecure **S**ocket **L**ayer
- Method of encrypting traffic over HTTP
- Increase in websites employing SSL

SSL (cont)

- Interested in the x509 certificate
    - Contains information regarding the issuer, subject, creation date, etc
    - Each certificate associated with IP(s)
    - x509 contain a CommonName field which can be either blank or contain an alphanumeric string
        - Our interest is in CommonNames that are valid domain names

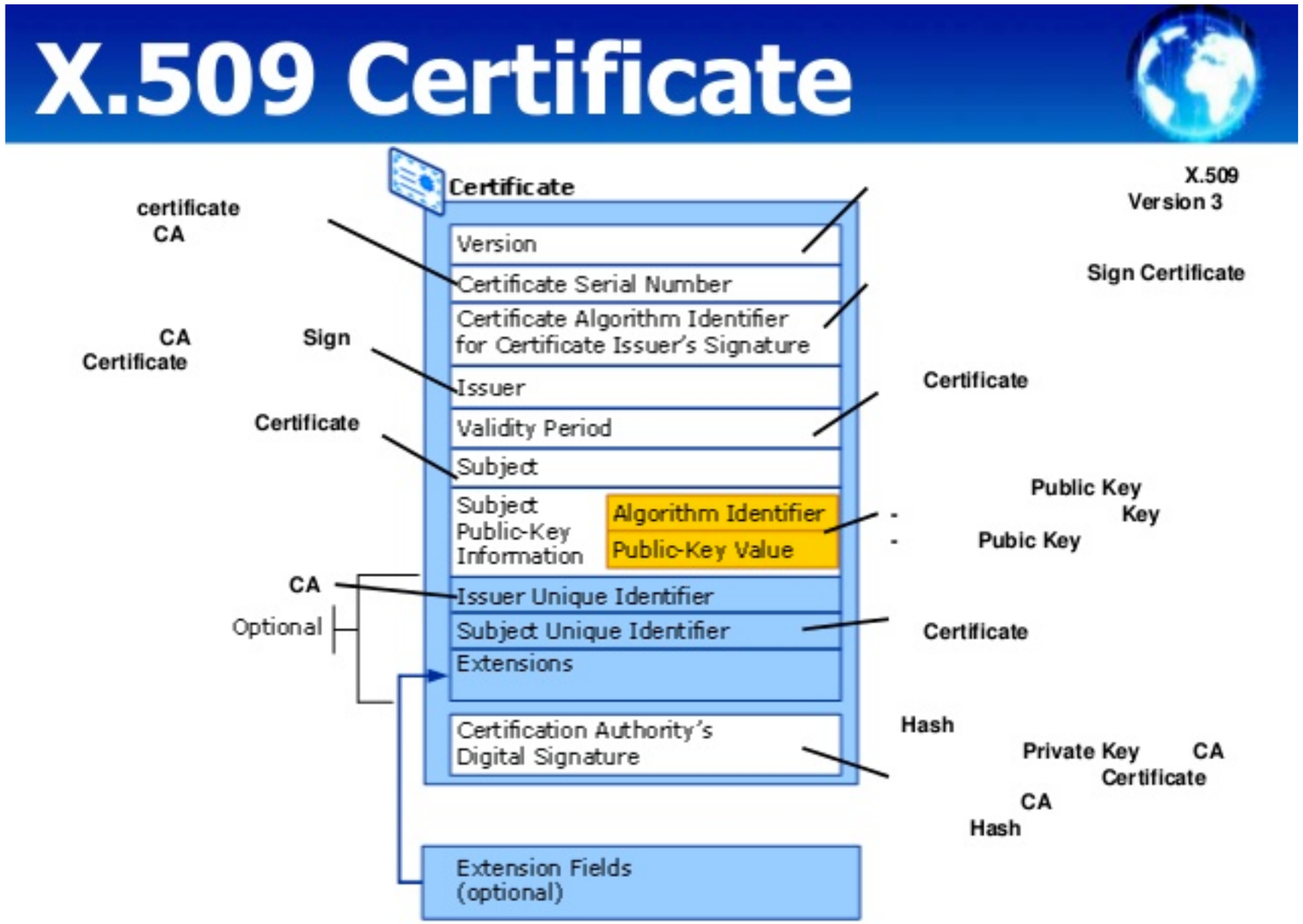
SSL (Cont)

- x509 certificates reveal valuable network information

- Maps relationship between domestic vs commercial IP space

- Gives information regarding ownership of a particular IP

SSL (Cont)


X.509 Certificate
Certificate
Version
Certificate Serial Number
Certificate Algorithm Identifier for Certificate Issuer's Signature
Issuer
Validity Period
Subject
Subject Public-Key Information
Algorithm Identifier
Public-Key Value
Issuer Unique Identifier
Subject Unique Identifier
Extensions
Certification Authority's Digital Signature
Extension Fields (optional)
Optional
certificate CA
CA Certificate
Sign
Certificate
CA
X.509 Version 3
Sign Certificate
Certificate
Public Key
Key
Pubic Key
Certificate
Hash
Private Key
CA Certificate
CA
Hash

CDNs, SSL, Zbot?

- CDNs serve as content delivery for popular domains
- CDNs help establish similar connections to a domain by hosting identical x509 certificates or x509 certificates with the same common name
- In recent times Zbot has attempted to use its network of hacked machines as CDN for malicious actors (hosts carding forums, trojans, etc)
- Can we use SSL data to help us identify Zbot?

Sonar Data

- 2x or 4x monthly scan of the entire IPv4 IP space
- x509 —> IP pairing
- Sampled data because of restrictions

Sonar Data

- Longitudinal study of Sonar SSL data. Examine hosting patterns over a 5 month time period
  - Hope to identify anomalous behavior

Sonar Data

- Table documents the number of SHAs and CommonNames per month

- Manually infeasible

	Unique SHAs	Unique Common Names
JAN	1,068,402	850,236
FEB	692,542	589,609
MARCH	977,484	813,773
APRIL	249,252	233,834
MAY	1,098,914	958,321

Investigation

- Data easier to work with if it has a natural structure

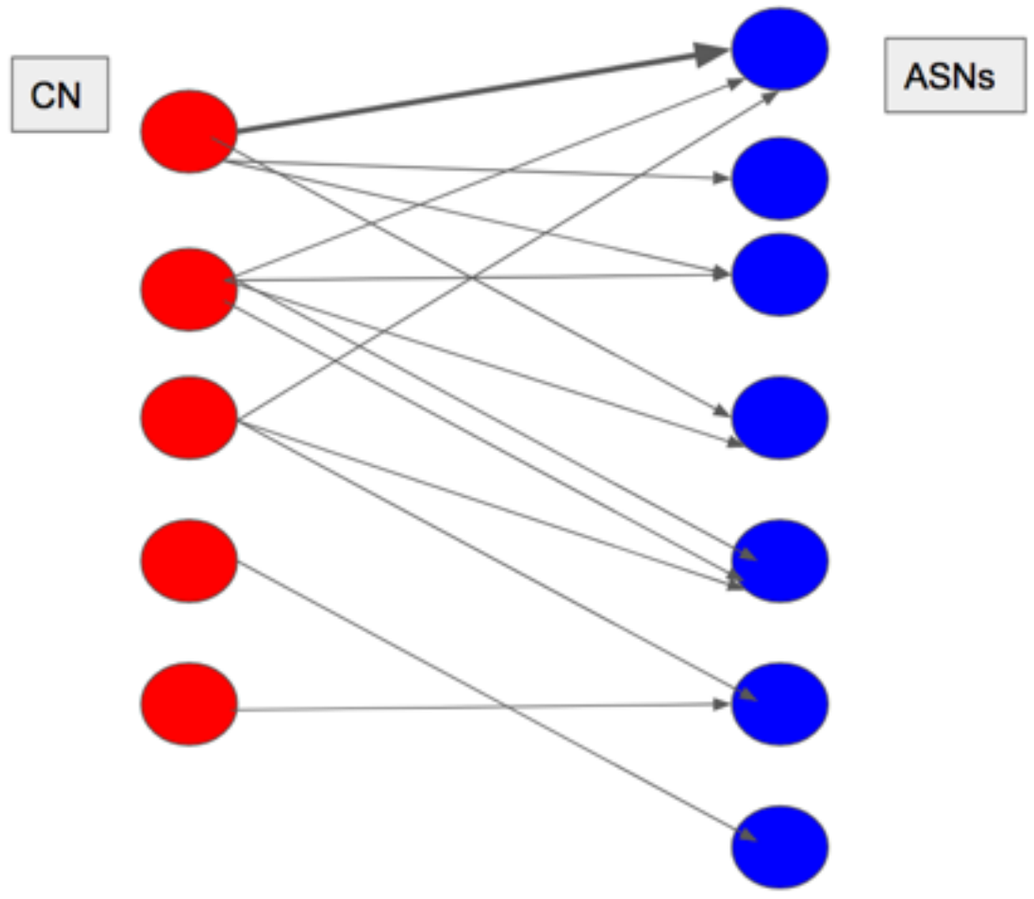
- What structure naturally captures this type of data?

Investigation

- Graphs can be used to model network data

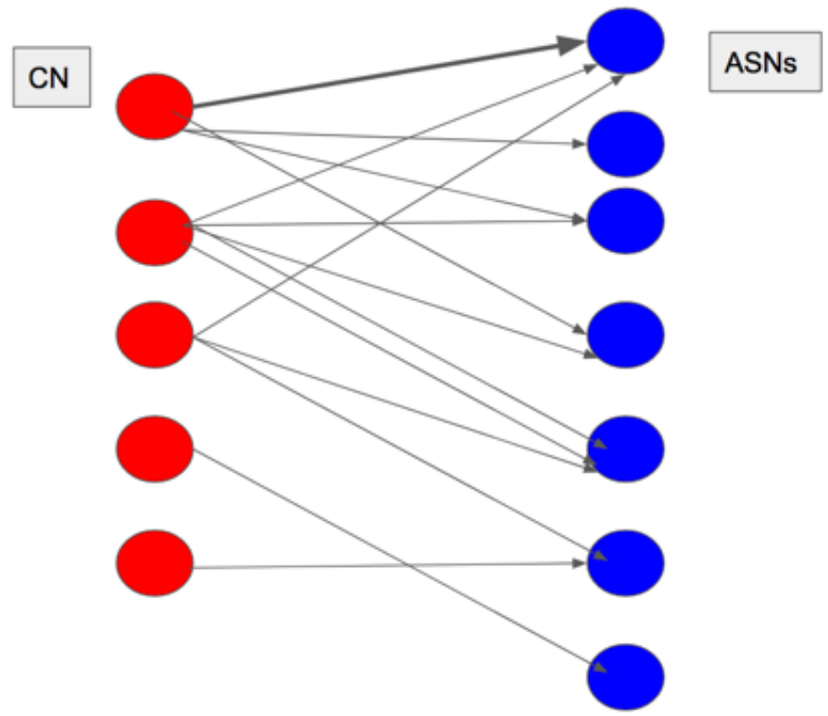
- We will work with bipartite graphs

- A bipartite graph is a graph whose vertices can be split into two disjoint sets



Bipartite Graphs

- Two sets:
  - Common Names
  - ASNs
- Could create a bipartite graph between CommonName —> IP, etc
- Choose ASN because provides best resolution to analyze the data

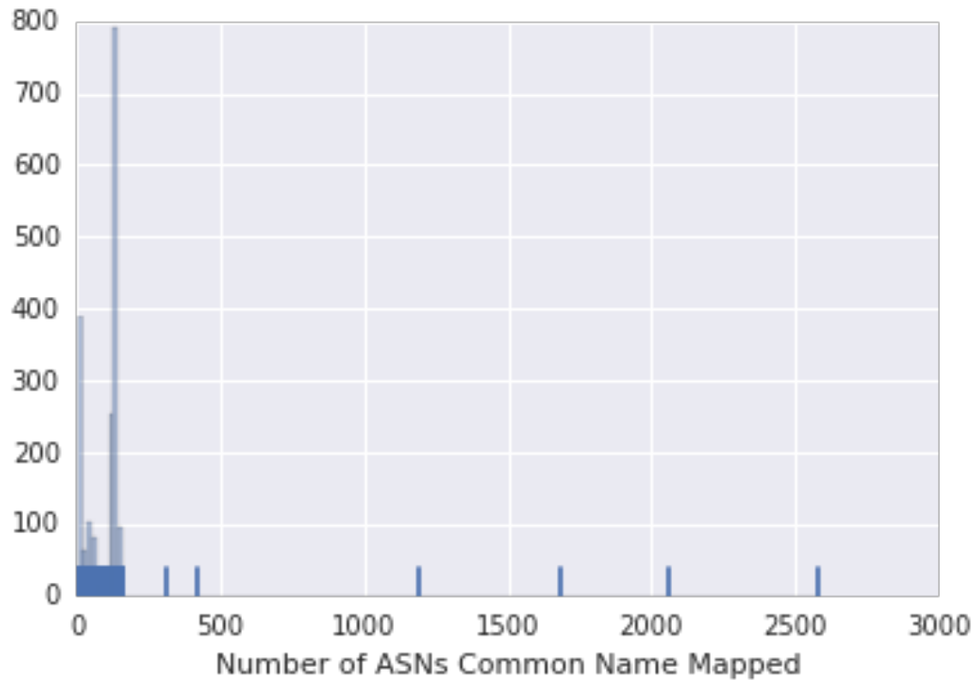


Bipartite Graph

- Multiple methods of analyzing the graph:
  - Graph factorization
  - Identify minimum connected components
  - Extract minimum spanning tree

Graph Analytics

- Our interest is in anomalous substructures within the graph



- Examining out degree of CommonNames to discover anomalies

- Plot histogram of graph degree counts

Graph Analytics

- Histogram shows that there are a few outliers in the tail
- Majority of the domains bounded between 1-200
- These outliers are expected. google and dlink would be widely distributed domain names across IP space

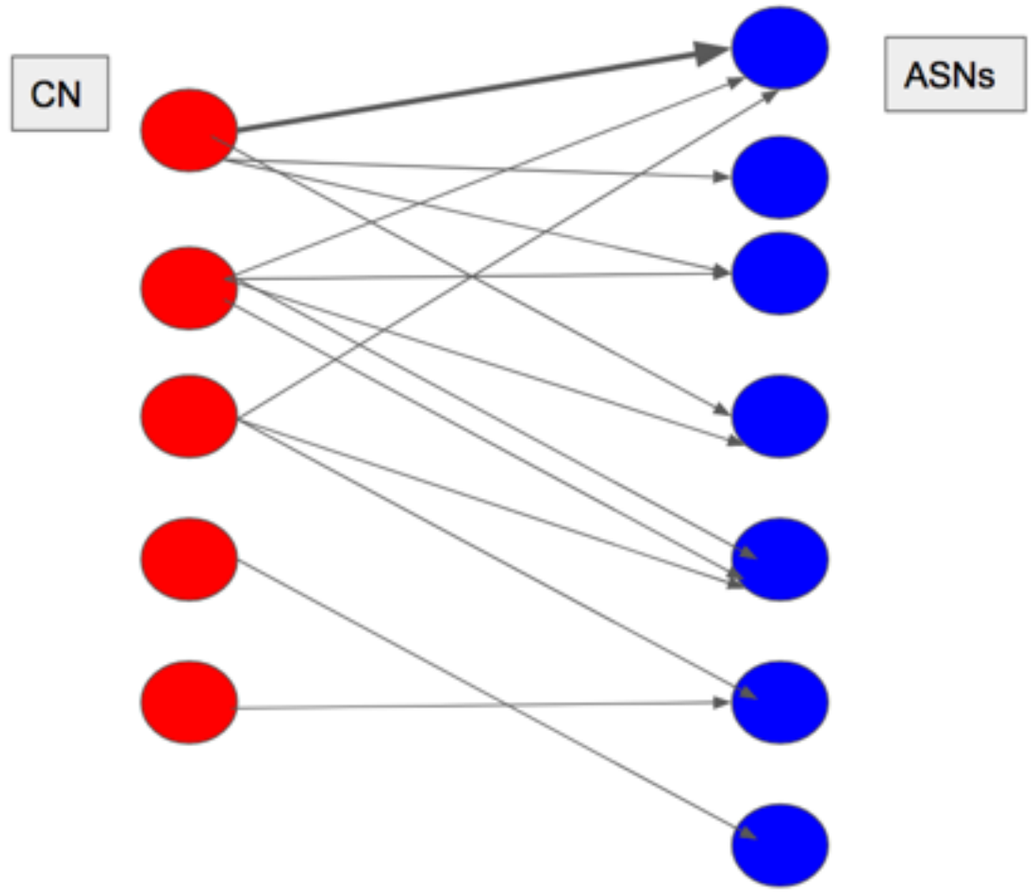
	domain	asn_count
0	www.dlink.com	2577
1	googlevideo.com	2060
2	google.com	1681
3	synology.com	1195
4	www.example.com	416
5	www.dlink.com.tw	311
6	itunes.apple.com	160
7	image-glb.qpyou.cn	159
8	asos-media.com	158
9	download.mcafee.com	157
10	www.koreanair.com	150

Popular Graphs

- Goal: Create a measure of popularity for each CommonName

- Have it based on topological features of the bipartite graph

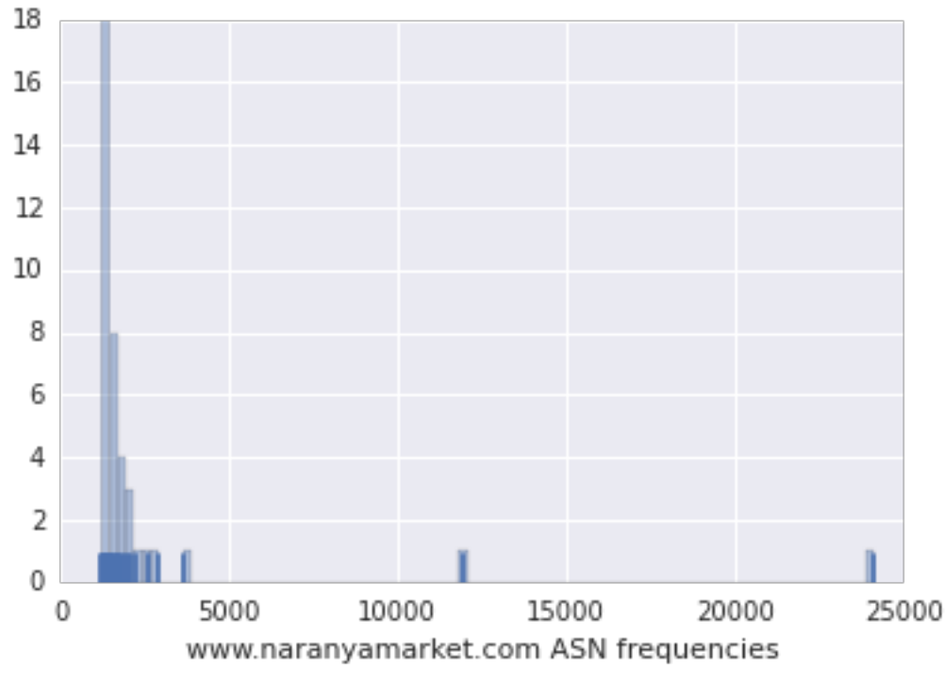
- Calculate frequencies of what type of ASNs host each common name



Popular Graphs

- Two different types of histogram distribution

- naranyamarket.com is found on some extremely popular ASNs



Filters for Anomaly Detection

- We want a filter that is more sensitive to ‘low frequencies’

- Fits our hypothesis regarding hosting patterns of popular domains

- To create such a filter we increase the number of bands in the lower frequency spectrum (ie 1-5 domains map to an ASN).

Filters for Anomaly Detection

- Bucket the frequencies into 9 different bands.
  - ie 1-5, 5-10, 10-20, 20-50, 50-100, 1000-2000, etc
  - notice the higher resolution for lower frequencies
- Each domain is then associated with a 9-d vector

Filters for Anomaly Detection

- Use the following algorithm:
    - Split CommonNames by increments of 10.
    - Within each of these groups create the distance matrix (n x n) between each pairwise CommonName using the Euclidean norm
    - Calculate the norm for each column vector

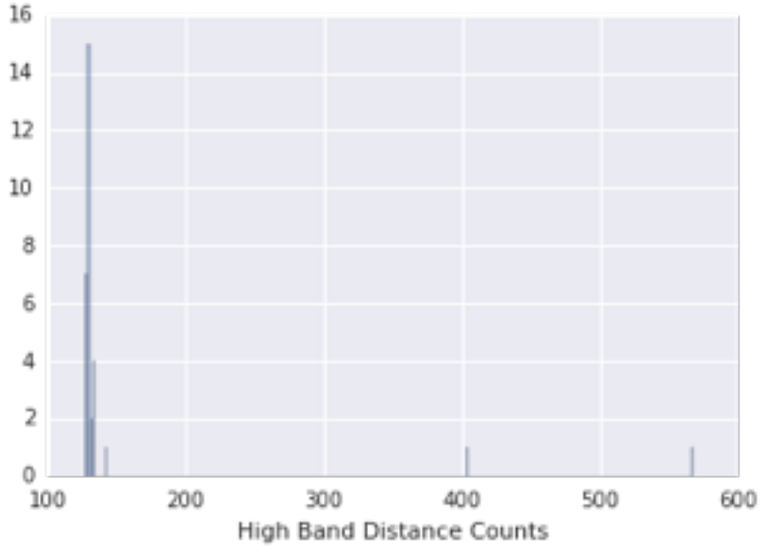
Algorithm Results

- This is the result from a particularly interesting interval (110-110)

- The histogram clearly shows a significant outlier (more than 2 std away)

- The outlier in the high band was ‘tangerine-secure.com’ which we ended up verifying as a ZBot domain

```
Out[94]: [('leonisavirtual.leonisa.com', 141.70744511139841),
          ('uis.uat.usajobs.gov', 132.38579984273238),
          ('frswebservices.org', 133.77966960640916),
          ('clarovideo.net', 133.31916591398252),
          ('tangerine-secure.com', 567.00352732588885),
          ('mobileonline.td.com', 131.44200241931802),
          ('www.getnet.com.br', 133.09019498069722),
          ('sanssl-014.bsdtools.com', 131.67004215082488),
          ('www.toysrus.de', 130.00384609695206),
          ('www.toysrus.fr', 130.00384609695206),
          ('cert2.coxmediagroup.com', 129.0310040261642),
          ('www.toysrus.at', 130.00384609695206),
          ('api.services.westjet.com', 130.00384609695206),
          ('stage.ritzcarlton.com', 130.00384609695206),
          ('www.brp.com', 128.7944098165755),
          ('edge-cdn.net', 129.73048986263791),
          ('midatlantic.aaa.com', 128.05467582247826),
          ('webssl.chinanetcenter.com', 403.43277011170028),
          ('tickets.cirquedusoleil.com', 128.7944098165755),
          ('www.santander.cl', 128.7944098165755),
          ('www.borbonese.com', 128.7944098165755),
          ('secure.boulanger.fr', 128.7944098165755),
          ('media1.1800flowers.com', 128.7944098165755),
          ('stage.brighttalk.net', 128.7944098165755),
          ('www.atgstores.com', 127.81627439414747),
          ('blueapron.com', 129.77287852244012),
```



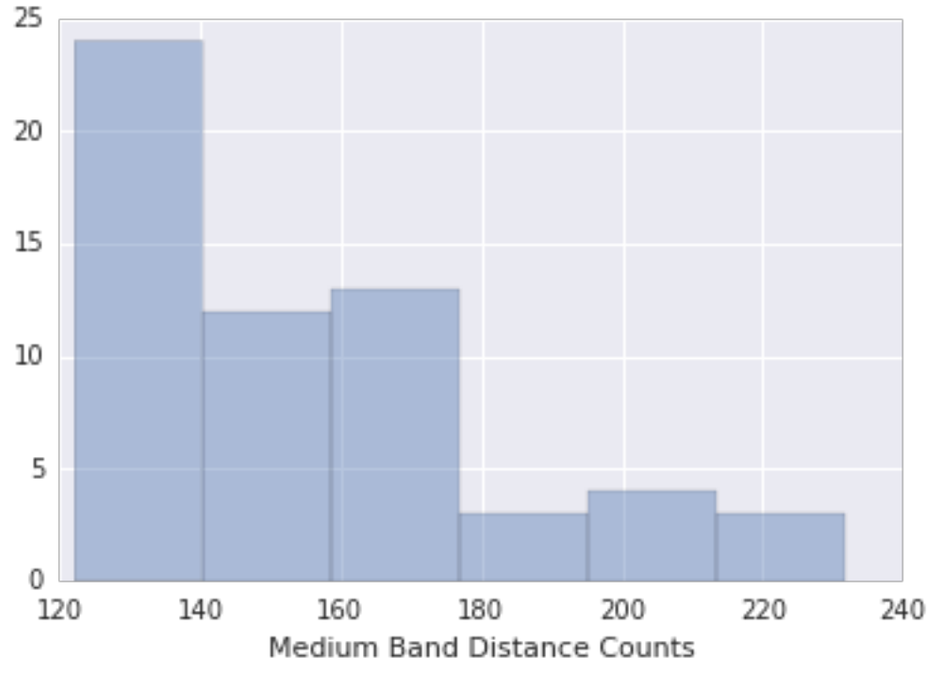
Algorithm Results

- Another interesting band is the 30-40 ASN range

- Much tighter spectrum of results

- The tail end of the histogram contains interesting domains

```
Out[104]: [('casino.netbet.com', 154.65445354078881),
           ('r-ak.bstatic.com', 160.09996876951601),
           ('example.com', 219.68613975396809),
           ('meenyousecu.com', 231.53401477968632),
           ('ssl2.cdngc.net', 161.44968256394932),
           ('www.naranyamarket.com', 142.09855734665288),
           ('swf.dougaku.tv', 160.09996876951601),
           ('www.wondershare.com', 154.65445354078881),
           ('realvu.net', 154.83216720048841),
           ('scdn1.ssl.altcdn.com', 154.83216720048841),
           ('secure.dot-st.com', 154.83216720048841),
           ('ustat.petpic.jp', 154.83216720048841),
           ('test-p-drv-ds1.support.ricoh.com', 154.83216720048841),
           ('www.exceda.com.ar', 141.04254677224174),
           ('www.dom.daimler.com', 141.04254677224174),
           ('js.rtoaster.jp', 154.83216720048841),
           ('cdn.zamimg.com', 154.83216720048841),
           ('bingowithkerry.com', 133.92535234226565),
           ('support3.cdnetworks.net', 166.99401186868948),
           ('support.cdnetworks.net', 173.88789492083686),
           ('fabulousbingo.co.uk', 128.7206277175496),
           ('ssl.cdngc.net', 165.14539048971363),
           ('sunbingo.co.uk', 127.09051892253804),
           ('support2.cdnetworks.net', 165.99397579430405),
           ('galacasino.com', 127.09051892253804),
           ('www.infoblox.com', 189.01058171435798),
           ('bdydns.com', 210.09521650908667),
```



Algorithm Results

- Out of these 5 domains - three are ZBot (meenyousecu.com, securedatassl.net, secure.tangerineaccess.com')
- Identified via further probing
- This anomaly detection method was able to reduce ~500k domains down to a manageable list of 8 domains
- Gave us actionable intelligence regarding which ASNs to monitor more closely

```
meenyousecu.com:231.53401478
bdydns.com:210.095216509
securedatassl.net:206.300751332
secure.tangerineaccess.com:214.35717856
flxdns.com:204.514058196
```