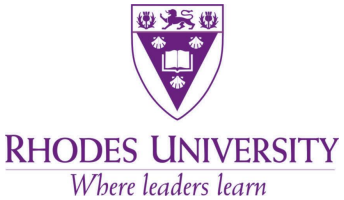
DNS Based Botnet C2 Server Detection

Spatial Statistics as a detection metric

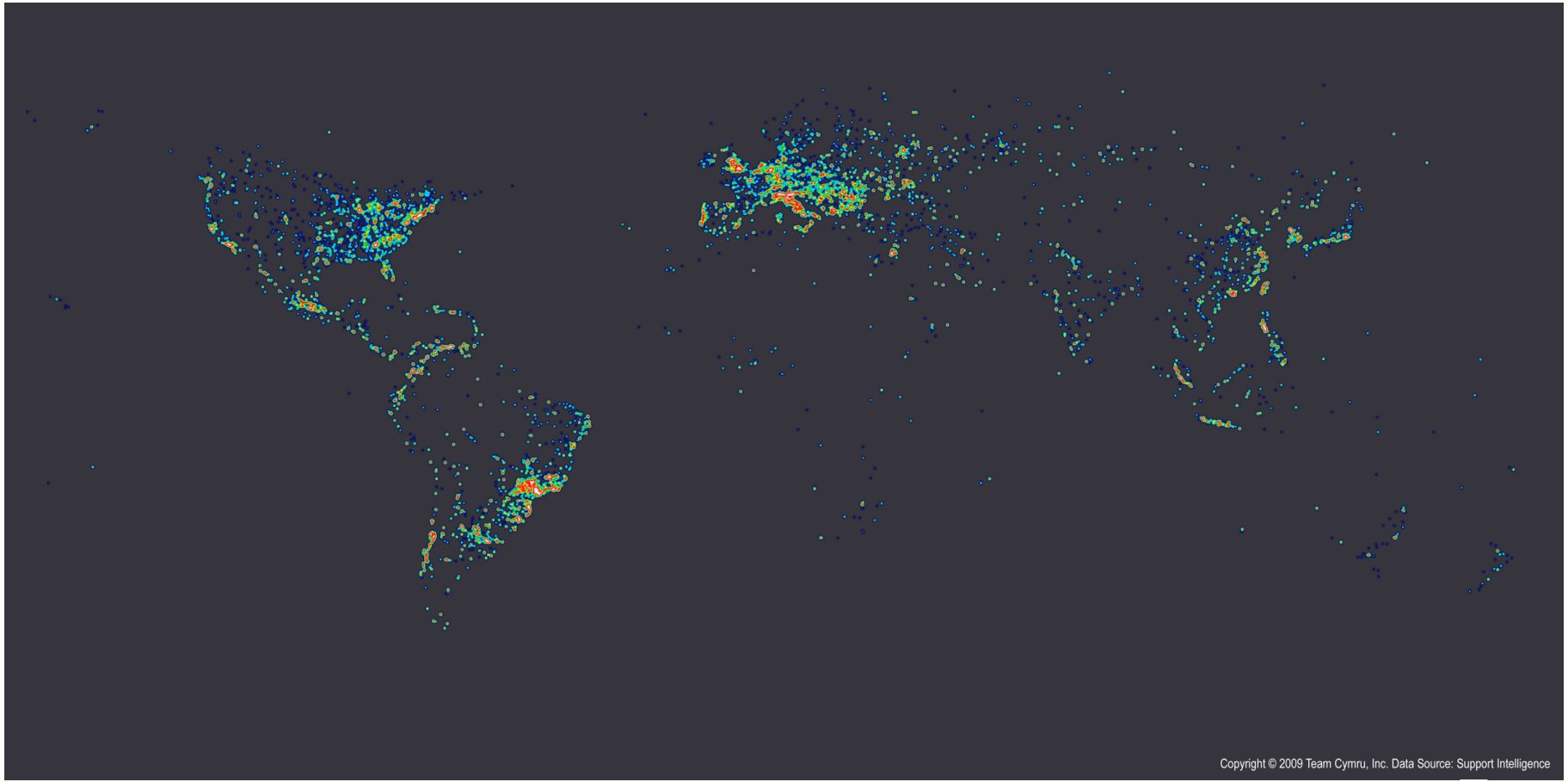


@kamp_staaldraad

etienne@sensepost.com

botconf 2013
First botnet fighting conference

sensepost



https://www.team-cymru.org/images/conficker-2009-01-29-dark-full.jpg

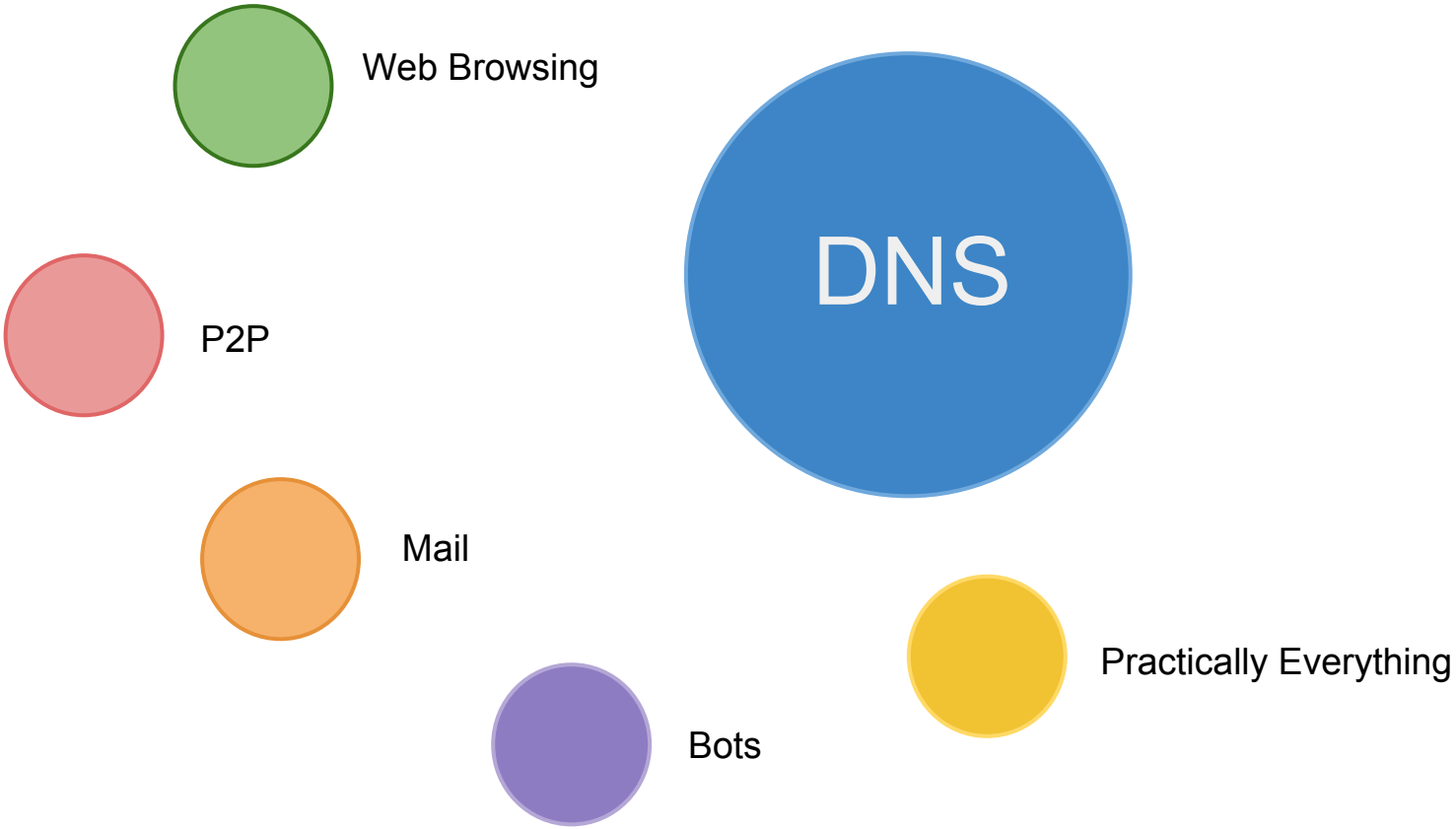
Geographic Analysis

sensepost

Research Goals

- Accurately detect botnet traffic
  - Assume no prior knowledge
  - Lightweight
  - Fast
  - Adaptable
- Early detection

sensepost



Examining DNS

sensepost

DNS Fast-Flux

```
; <<>> DiG 9.8.1-P1 <<>> lovenewgirl.com
;; global options: +cmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 12744
;; flags: qr rd ra; QUERY: 1, ANSWER: 5, AUTHORITY: 5, ADDITIONAL: 0

;; QUESTION SECTION:
;lovenewgirl.com.               IN      A

;; ANSWER SECTION:
lovenewgirl.com.        600     IN      A       67.190.124.84
lovenewgirl.com.        600     IN      A       78.60.45.112
lovenewgirl.com.        600     IN      A       118.14.224.139
lovenewgirl.com.        600     IN      A       194.90.36.187
lovenewgirl.com.        600     IN      A       222.106.31.112

;; AUTHORITY SECTION:
lovenewgirl.com.        86400   IN      NS      ns5.oldcitiesmaps.com.
lovenewgirl.com.        86400   IN      NS      ns2.oldcitiesmaps.com.
lovenewgirl.com.        86400   IN      NS      ns3.oldcitiesmaps.com.
lovenewgirl.com.        86400   IN      NS      ns4.oldcitiesmaps.com.
lovenewgirl.com.        86400   IN      NS      ns1.oldcitiesmaps.com.

;; Query time: 493 msec
;; SERVER: 127.0.0.1#53(127.0.0.1)
;; WHEN: Tue Aug  7 09:08:11 2012
;; MSG SIZE  rcvd: 217
```

- Short TTL
- Multiple A Records
- Different IP Ranges

sensepost

DNS Fast-Flux

```
;; ANSWER SECTION:
lovenewgirl.com.        600     IN      A       67.190.124.84
lovenewgirl.com.        600     IN      A       78.60.45.112
lovenewgirl.com.        600     IN      A       118.14.224.139
lovenewgirl.com.        600     IN      A       194.90.36.187
lovenewgirl.com.        600     IN      A       222.106.31.112
```

```
Lat: 39.7437 Lon: -104.9793
UTM: 37Z
MGRS: 13SED0177499311
```

```
Lat: 54.6833 Lon: 25.3167
UTM: 40R
MGRS: 35ULA9148060851
```

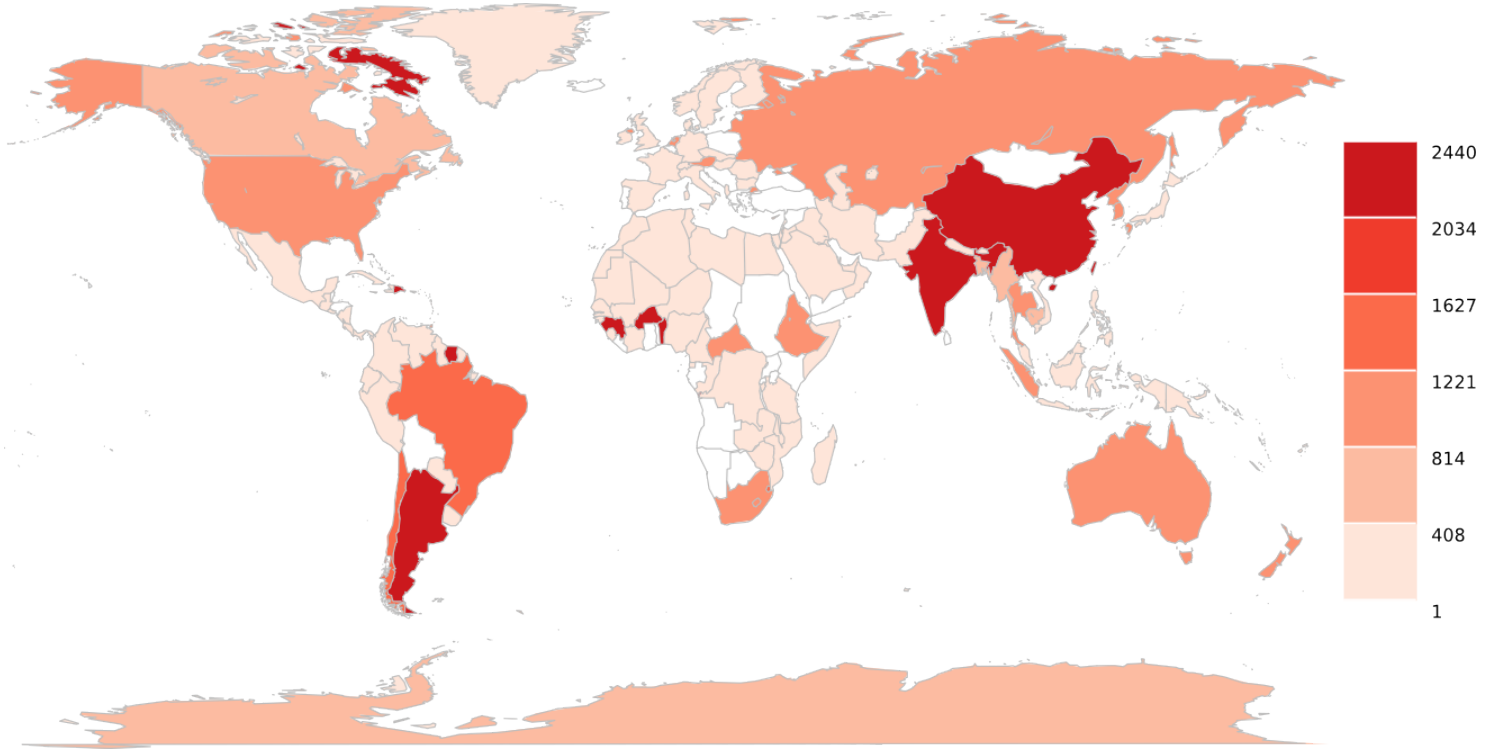
```
Lat: 34.6833 Lon: 135.8333
UTM: 36Z
MGRS: 53SNU7633338239
```

IP	ASN	Country	Timezone
67.190.124.84	AS7922	US	America/Denver
78.60.45.112	AS8764	LT	Europe/Vilnius
118.14.224.139	AS4713	JP	Asia/Tokyo
194.90.36.187	AS1680	IL	Asia/Jerusalem
222.106.31.112	AS4766	KR	Asia/Seoul

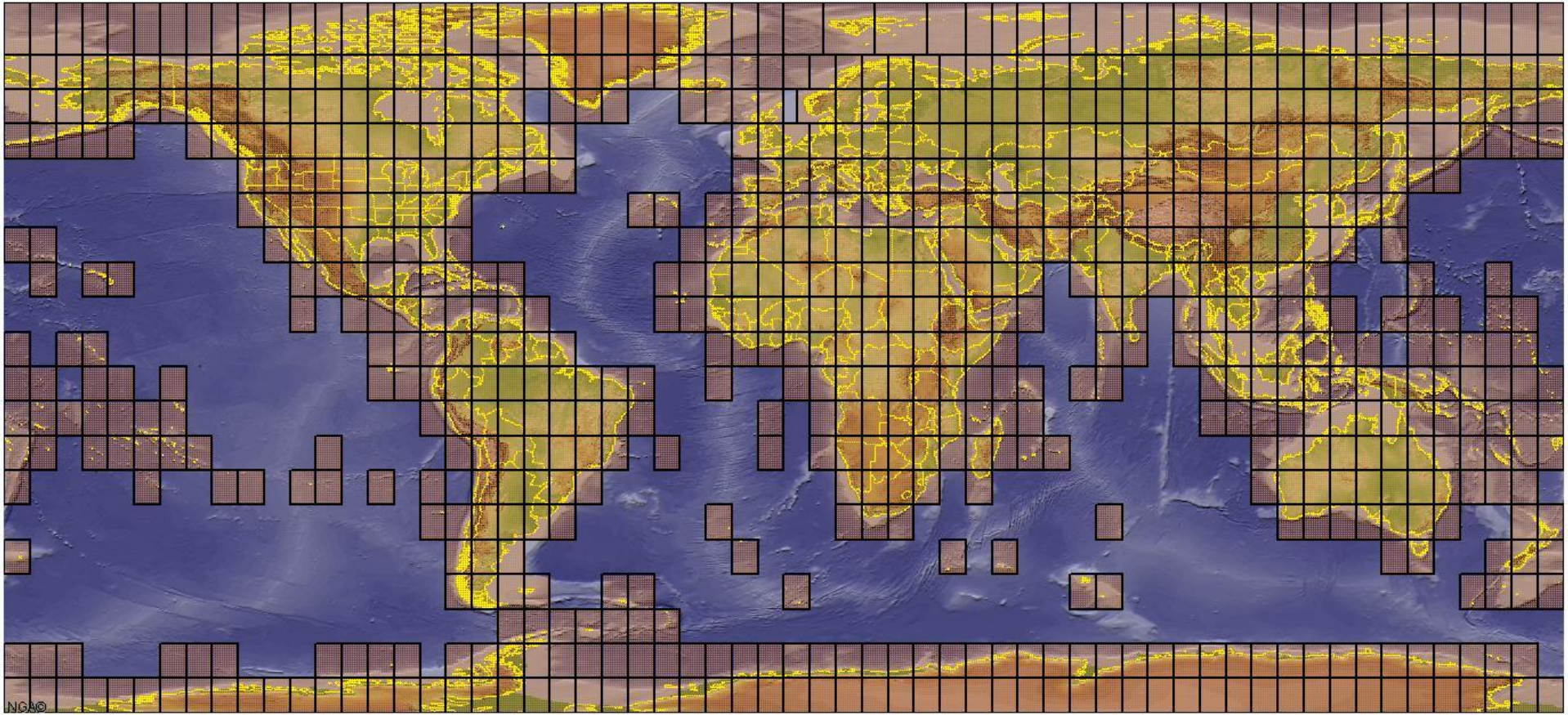
- Multiple ASNs
- Multiple Countries
- Multiple Timezones
- Multiple Unique Location Identifiers

sensepost

Widely Dispersed Networks



sensepost



Spatial Measures

sensepost

http://earth-info.nga.mil/GandG/coordsys/images/MGRS_1km_Polygon_Shapefiles_Coverage.jpg

Spatial Measures

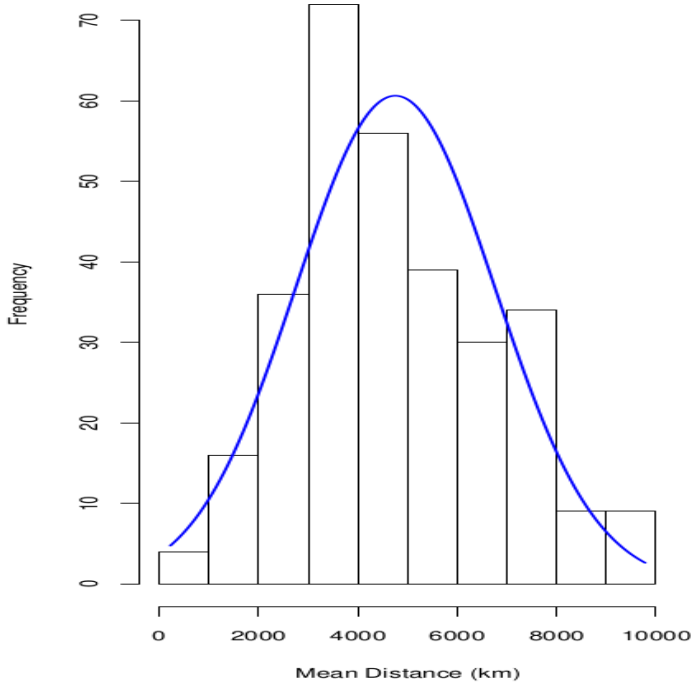
```
[08:32]etienne@null
↳ ~/Documents/Rhodes/Masters/Prototype > python presentationData.py
```

IP	ASN	Country	Timezone	UTM	MGRS
67.190.124.84	AS7922	US	America/Denver	37Z	13SED0177499311
78.60.45.112	AS8764	LT	Europe/Vilnius	40R	35ULA9148060851
118.14.224.139	AS4713	JP	Asia/Tokyo	36Z	53SNU7633338239
194.90.36.187	AS1680	IL	Asia/Jerusalem	36S	36SXB8622432597
222.106.31.112	AS4766	KR	Asia/Seoul	37Z	52SCG2334159589
79.108.149.71	AS6739	ES	Europe/Madrid	37M	30SYH0125936055
79.139.110.20	AS41740	PL	Europe/Warsaw	39Q	34UFA2837416063
79.139.110.20	AS41740	PL	Europe/Warsaw	39Q	34UFA2837416063
88.132.63.164	AS35311	HU	Europe/Budapest	38Q	34TDT0755809583
79.108.149.71	AS6739	ES	Europe/Madrid	37M	30SYH0125936055
124.6.3.225	AS24165	TW	Asia/Taipei	34Z	51QTF2762705352
89.229.214.126	AS21021	PL	Europe/Warsaw	39Q	34UCE6257855864
124.6.3.225	AS24165	TW	Asia/Taipei	34Z	51QTF2762705352
68.119.57.22	AS20115	US	America/New_York	36Z	17SKT4153099735

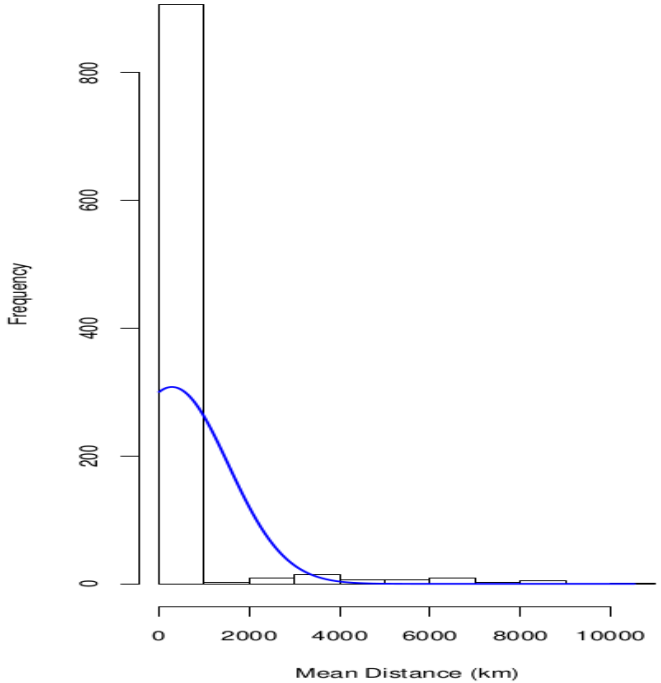
sensepost

Nearest Neighbours

Fast-Flux Domains



Benign Domains



sensepost

Spatial Statistics

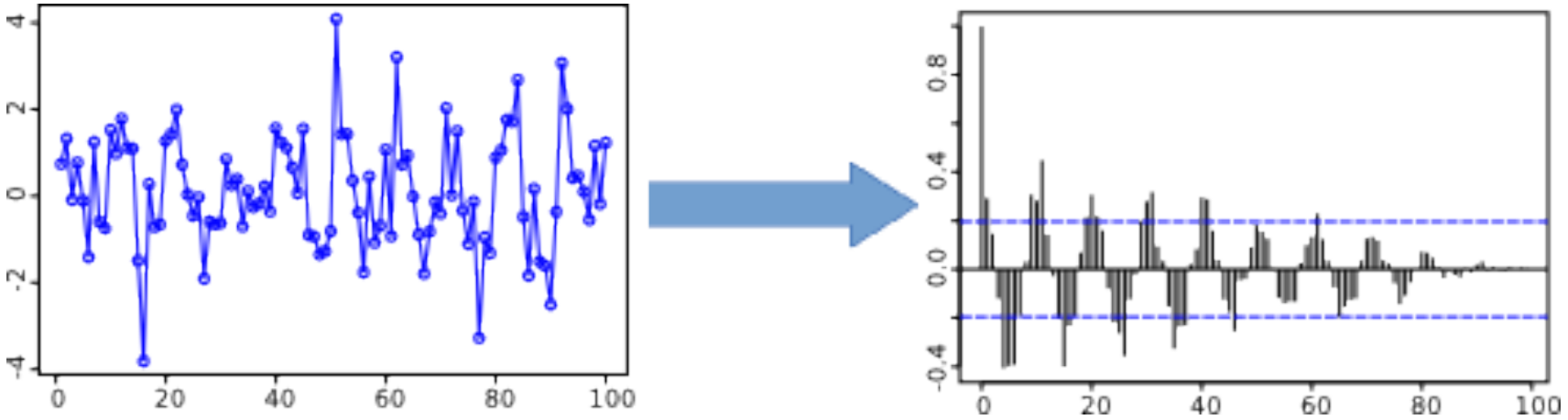
https://upload.wikimedia.org/wikipedia/commons/c/c7/Snow-cholera-map.jpg

First Law of Geography

"All things are related, but near things are more related than far things." - W. Tobler

sensepost

Autocorrelation



sensepost

Moran's Index

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2}$$

sensepost

Geary's Coefficient

$$C = \frac{(N-1)\sum_i \sum_j w_{ij}(X_i - X_j)^2}{2W\sum_i (X_i - \bar{X})^2}$$

sensepost

```
    def calcHaverDistance(self,lat1,lat2,lon1,lon2):
        '''
        Calculate the Distance between two locations using the Haver method
        @param latitude of location 1
        @param latitude of location 2
        @param longitude of location 1
        @param longitude of location 2
        @return the distance between the two locations
        '''
        dLat = math.radians(lat2-lat1)
        dLon = math.radians(lon2-lon1)

        a = math.sin(dLat/2) * math.sin(dLat/2) + math.sin(dLon/2) * math.sin(
        c = 2 * math.atan2(math.sqrt(a), math.sqrt(1-a))
        d = self.R * c
        return d

    def geary(self,matrix,values,meanv,N):
        vx1 =0
        vx2 =0
        vx3 =0
        W = 0
        for i,x1 in enumerate(values):
            for j,x2 in enumerate(values):
              #print matrix[i][j],x1,x2
              vx1 += matrix[i][j]*((x1-x2)*(x1-x2))

              W += matrix[i][j]
            vx3 += (x1-meanv)*(x1-meanv)
        vx2 = 2*W

        if vx3 ==0 or vx2==0:
            Ix = 0
        else:
            Ix = ((N-1)*vx1)/(vx2*vx3)

        return Ix

    def getMGRSVal(self,lat,lon):
        m = mgrs.MGRS()
        c = m.toMGRS(lat,lon)
        ind = 1
        for i in range(1,len(c)):
            if c[i].isalpha():
                ind = i
                break
        v1 = int(c[:ind])*(ord(c[ind:ind+1])+ord(c[ind+1:ind+2]))
        v2 = int(c[ind+3:])

        return (v1*v2,c)
```

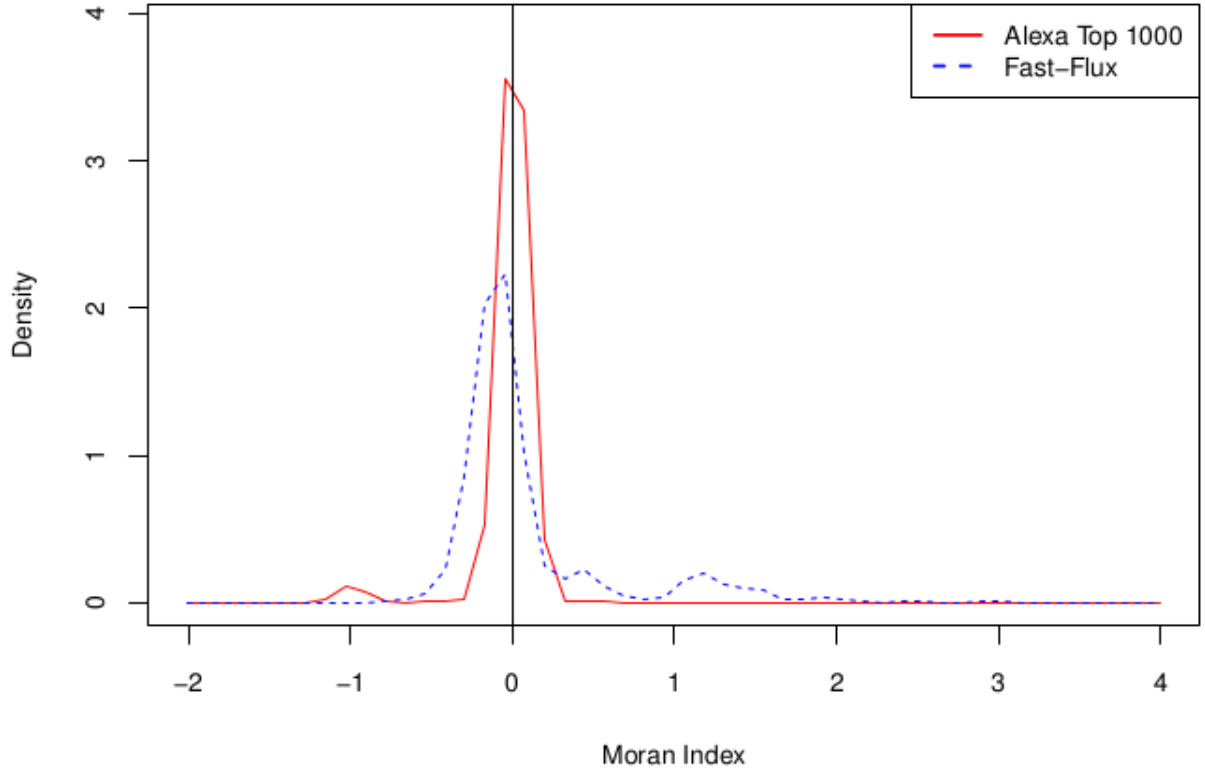
Building the Classifiers

sensepost

Classifier Training

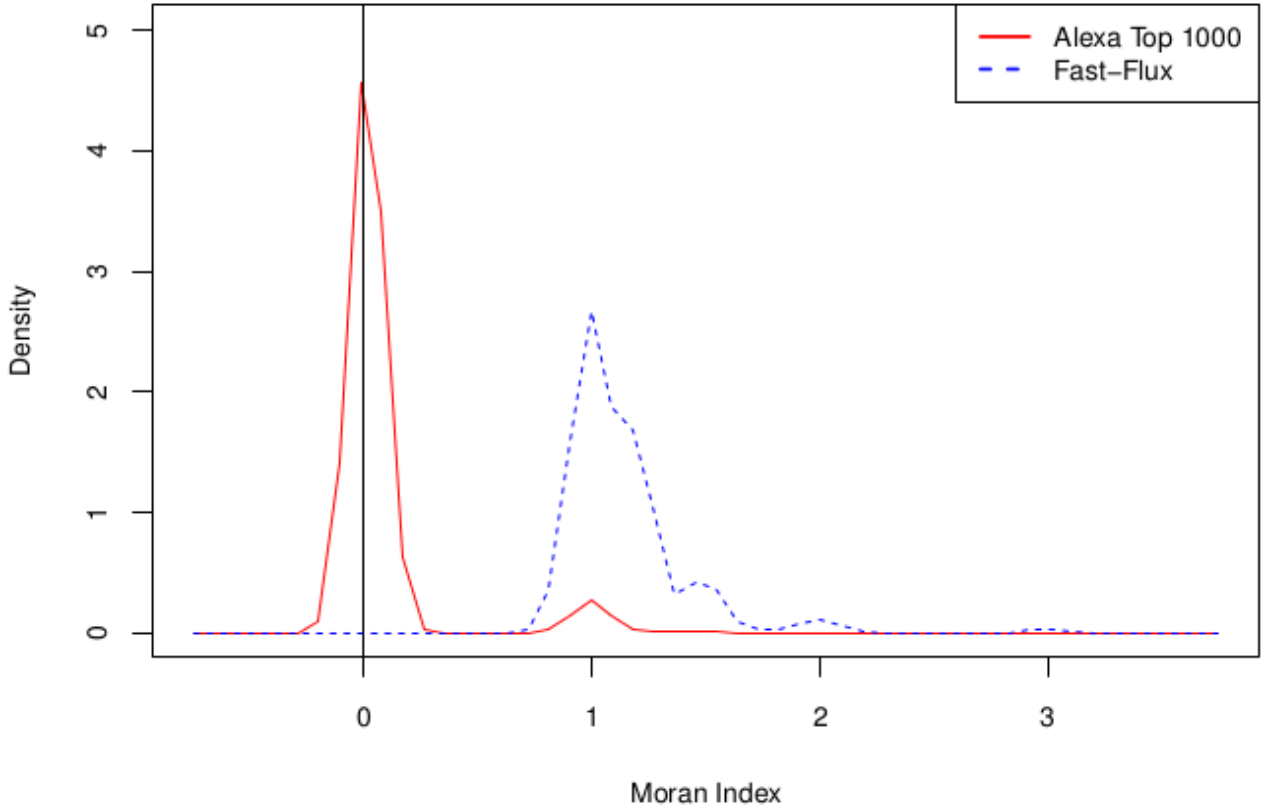
Benign Dataset

Fast-Flux Dataset



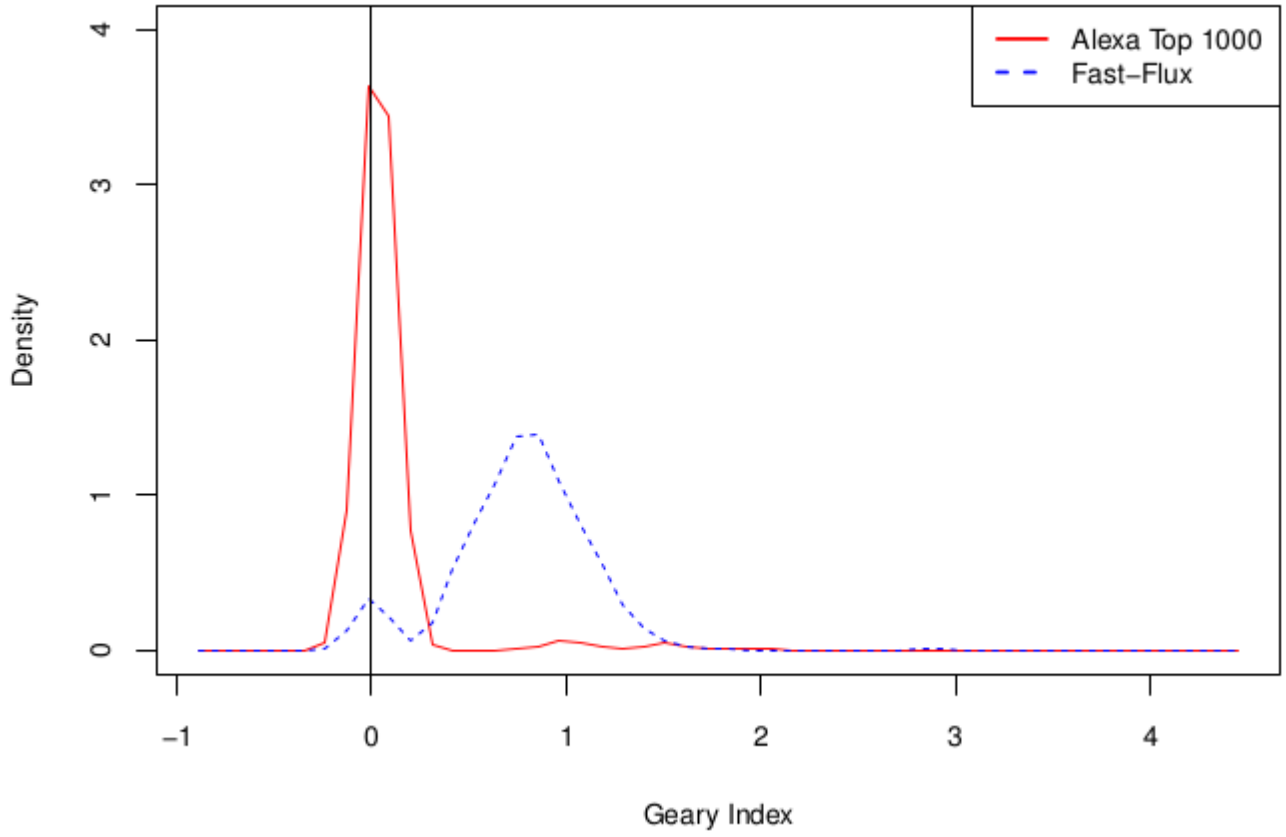
Moran's I: Timezones

sensepost



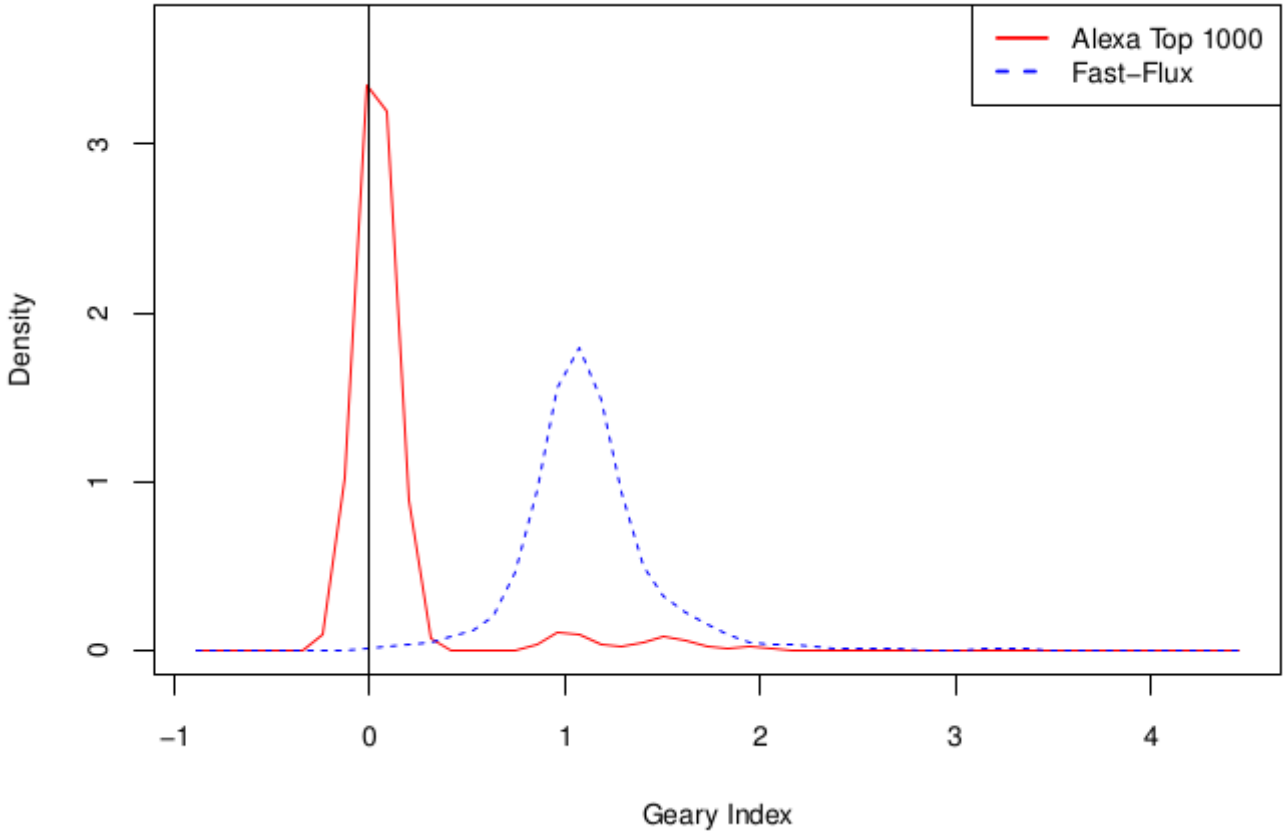
Moran's I: UTM

sensepost



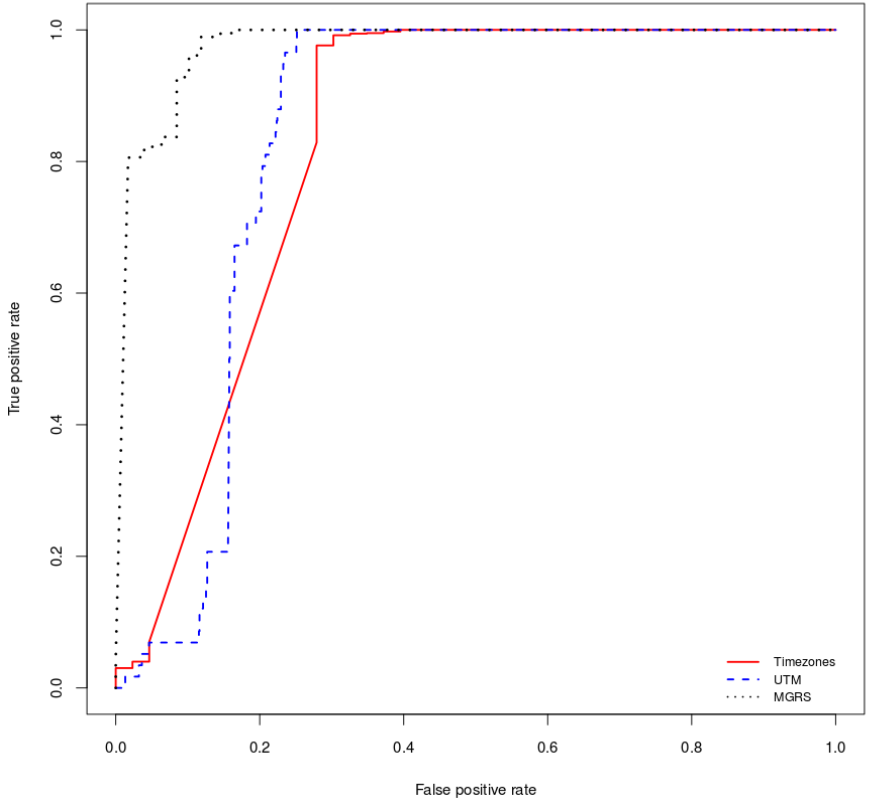
Geary's C: UTM

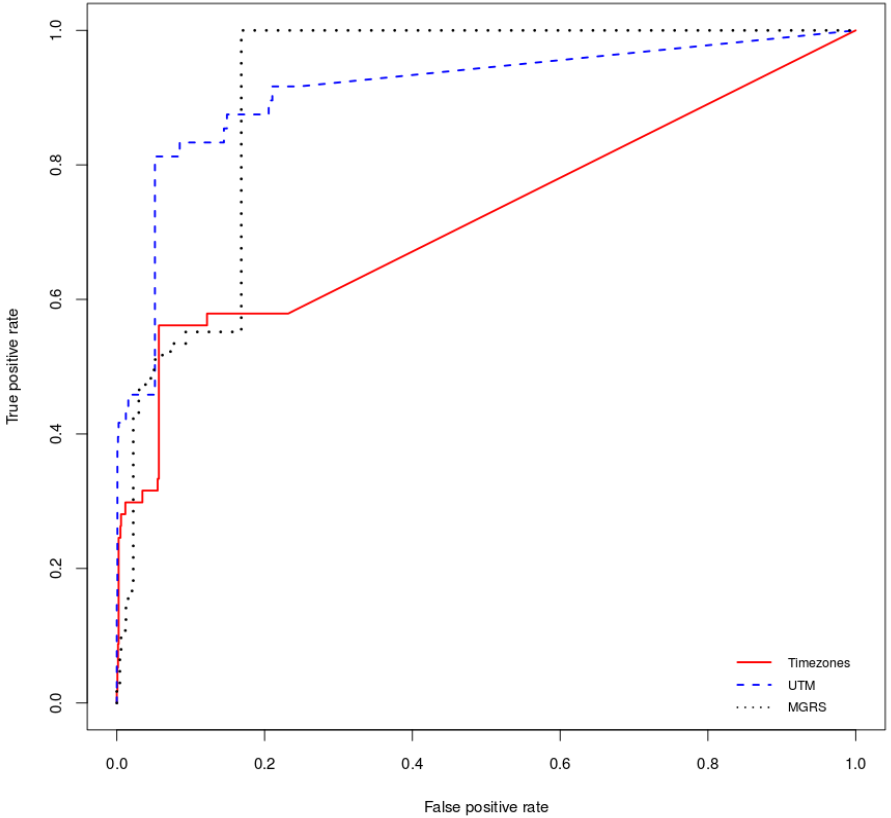
sensepost



Geary's C: MGRS

sensepost





Classifier Results

sensepost

Moran Classifier Results

97% Timezones

UTM 95%

95% MGRS

Accuracy

sensepost

Geary Classifier Results

95% Timezones

UTM 96%

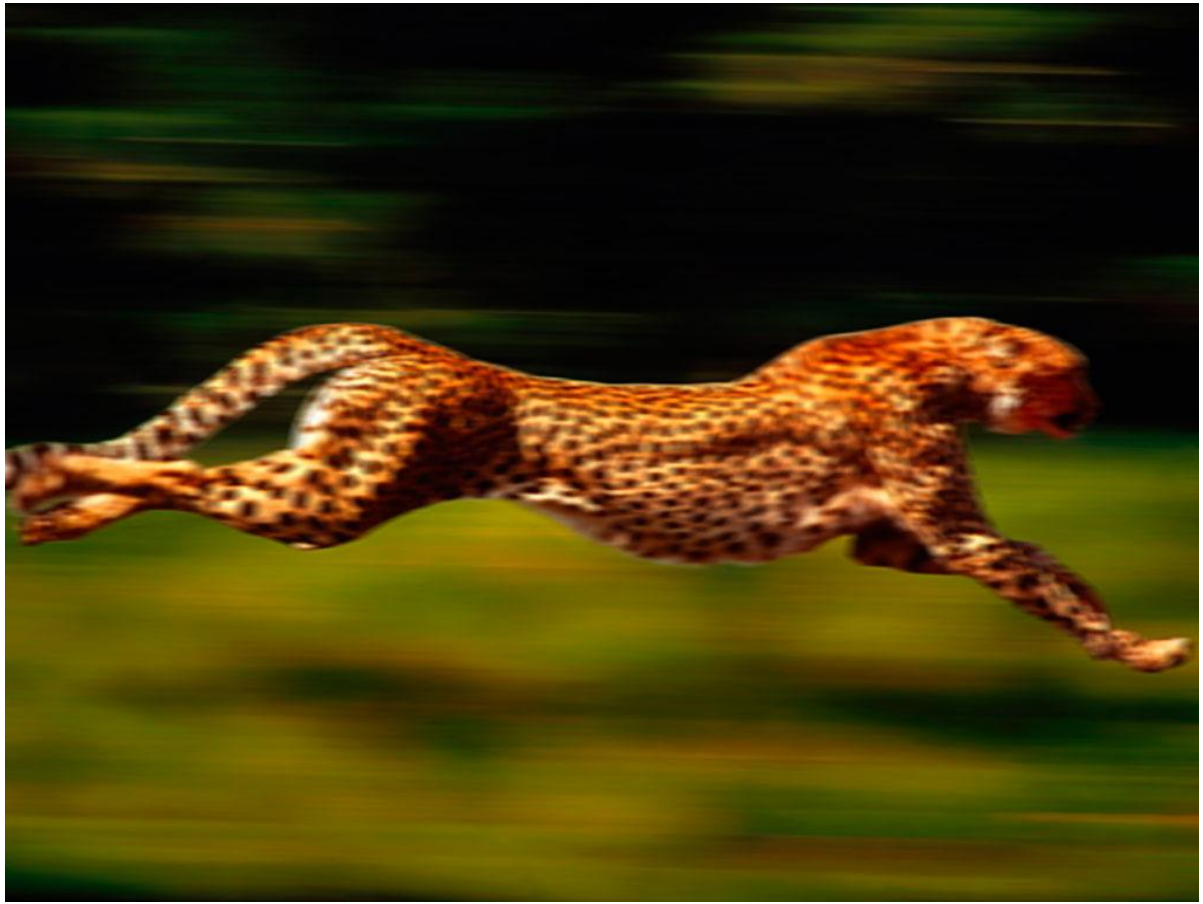
95% MGRS

Accuracy

sensepost

Evaluating Performance

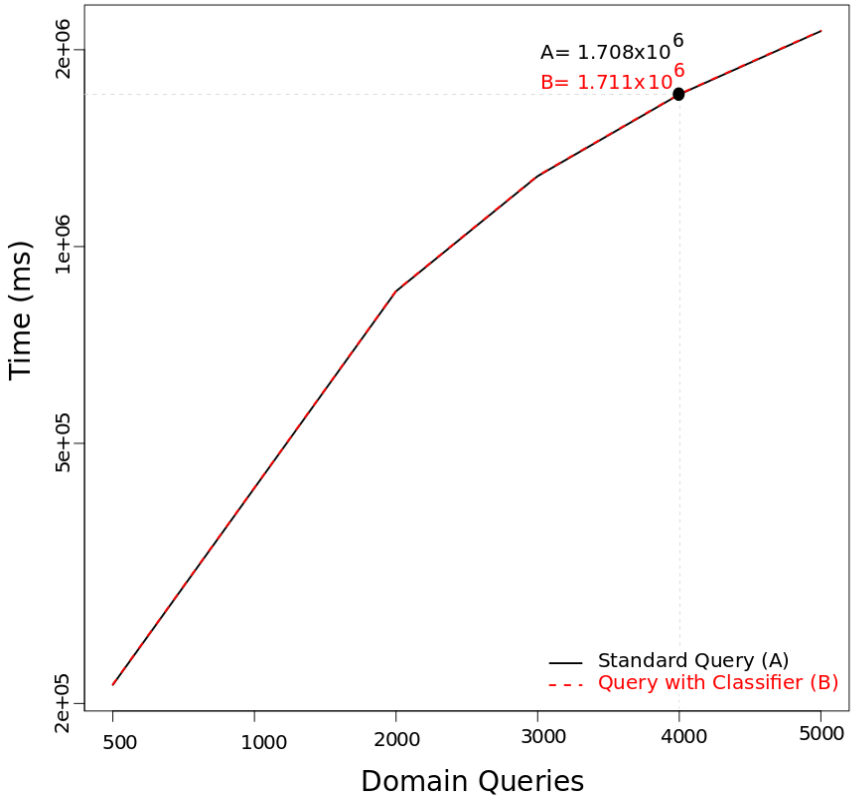
- Determine resource usage
- Impact on normal network performance
- Scalability



Classifier Performance Impact

http://beyond.customline.com/wp-content/uploads/2012/04/Cheetah-performance.jpg

sensepost



Measured Performance

sensepost

Measured Performance

20,000 domain lookups

Processed in 13 seconds

6.501×10^{-4} seconds per domain

sensepost

Benefits

Fast

Small

Low maintenance

Scalable

sensepost

Future Work

- Combine classifiers into stand-alone solution
- Combine detection and blocking
- Increase accuracy of geo-location

sensepost

Conclusion

@kamp_staaldraad

etienne@sensepost.com

sensepost