

**SNRG**

**Security and Networks  
Research Group**

# A FRAMEWORK FOR DNS BASED DETECTION AND MITIGATION OF MALWARE INFECTIONS ON A NETWORK

By: Etienne Stalmans

Supervisor: Dr Barry Irwin



**RHODES UNIVERSITY**  
*Where leaders learn*

# Bio

Who: Etienne Stalmans

Student at Rhodes University

Doing Computer Science (MSc)

Security and Networks Research Group

Web: <http://www.cs.ru.ac.za/research/g07S0924/>

Email: [g07s0924@campus.ru.ac.za](mailto:g07s0924@campus.ru.ac.za)

Twitter: [@kamp\\_staaldraad](https://twitter.com/kamp_staaldraad)

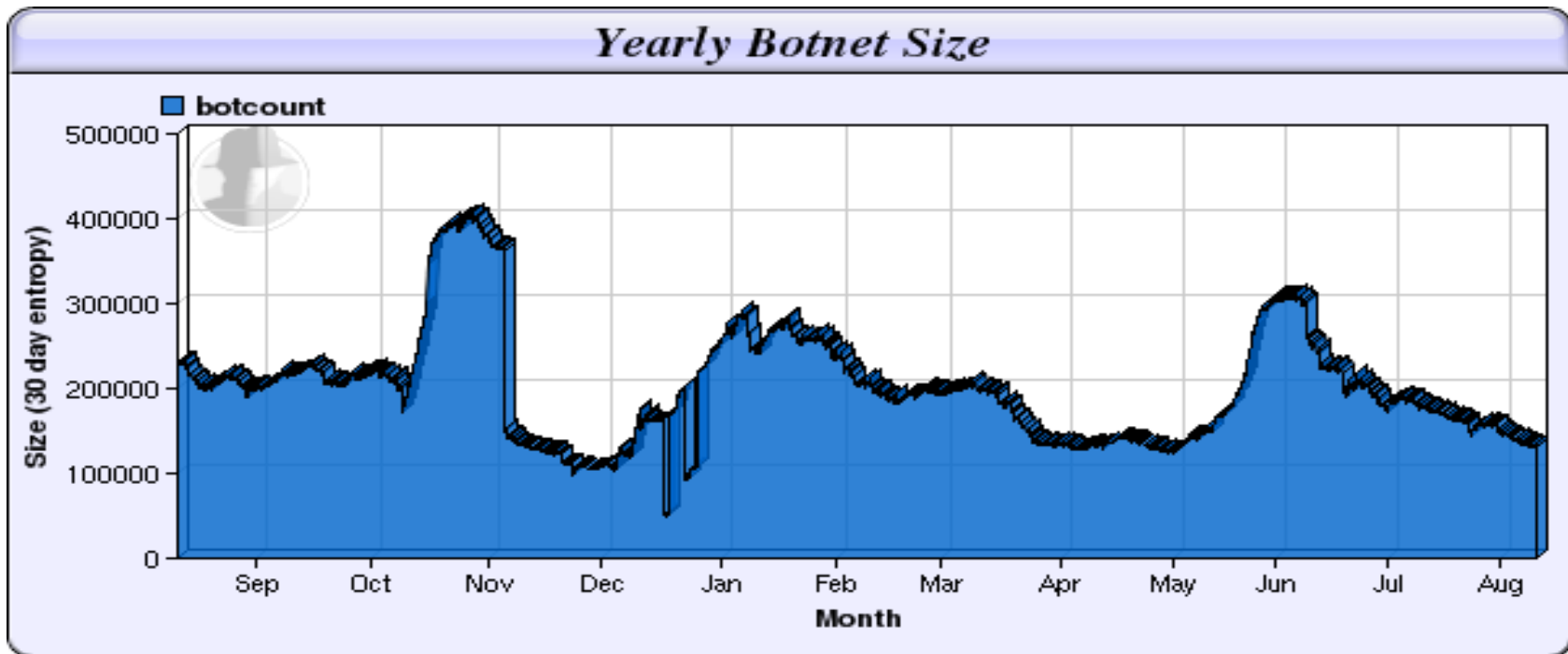
- System to detect Malware
  - Specifically hosts belonging to botnet
- Lightweight
  - Minimal extra processing
  - Little or no extra network load
- Scalable
  - Can be applied on network regardless of size
- Robust
- Adaptable
  - Can learn to recognise new infections

- Rely on blacklists/whitelists
  - Need to be kept up to date
  - Blind to 0-day attacks
- Manual
  - Register domains before Botnet owners
  - Blacklist domains on a daily basis
- Reverse Engineering of malware
  - Takes time
  - Even when successful, malware update invalidates results



- Use existing infrastructure
- Add sensor as a network tap
  - Or on DNS server
- Read DNS query responses
  
- Benefit:
  - Attempt to contact malicious host detected early
  - Detected before actual connection is made

# The Botnet Threat



<http://www.shadowserver.org/wiki/uploads/Stats/botcount-year.png>

- Peak of 400,000 infected hosts
- Average of 200,000 infected hosts
- Consistent over last 3 years

# What Are Botnets?

## A Botnet is:

- Private computers infected with malware
  - Known as a *Bot* or *Zombie*
  - Networked to form distributed botnet
  - Remotely controllable
- Used by criminals to send Spam Email, attack computer networks and to spread viruses
- Infected hosts are not always aware that they are infected
  - Rootkits
  - Detection avoidance techniques



[http://aipanic.com/wp/articleimages/computer\\_zombies.jpg](http://aipanic.com/wp/articleimages/computer_zombies.jpg)

# Detection Avoidance

## DNS Fast-flux

- Distributed Command and Control servers
- Numerous IP addresses associated with a single fully qualified domain name
- IP addresses swapped in and out rapidly
- Constantly changing DNS records



# DNS Fast-Flux

```
$ dig @NS2.WESTNS.COM wildcard.malaga-53.com a

; <<>> DiG 9.2.4 <<>> @NS2.WESTNS.COM wildcard.malaga-53.com a
; (1 server found)
;; global options: printcmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 13728
;; flags: qr aa rd; QUERY: 1, ANSWER: 5, AUTHORITY: 0, ADDITIONAL: 0

;; QUESTION SECTION:
;wildcard.malaga-53.com.          IN      A

;; ANSWER SECTION:
wildcard.malaga-53.com. 180     IN      A       68.126.240.182
wildcard.malaga-53.com. 180     IN      A       70.228.170.6
wildcard.malaga-53.com. 180     IN      A       68.74.207.77
wildcard.malaga-53.com. 180     IN      A       70.253.85.166
wildcard.malaga-53.com. 180     IN      A       67.37.184.67

;; Query time: 145 msec
;; SERVER: 67.190.128.40#53(67.190.128.40)
;; WHEN: Sun Sep  3 17:xx:xx 2006
;; MSG SIZE rcvd: 120
```

# DNS Fast-Flux

## Distribution of ZeuS Command and Control Servers



Imagery ©2011 TerraMetrics, NASA, Map data ©2011 Geocentre Consulting, MapLink, Tele Atlas - Terms of Use  
<https://zeustracker.abuse.ch/index.php>

# Domain Name Flux

## Another Technique to:

1. Prevent detection
2. Prevent shutdown
3. Increase lifetime

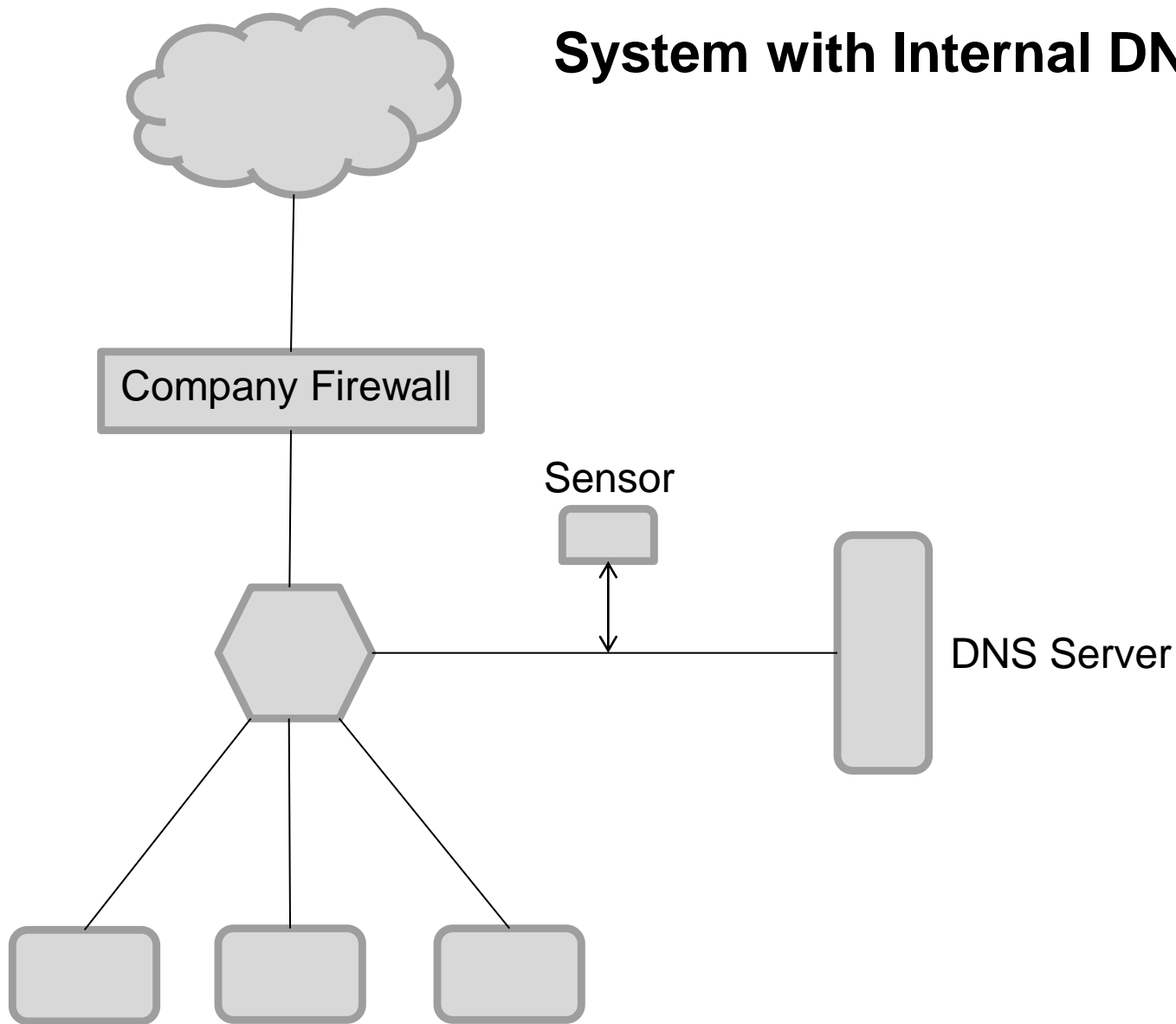
## Examples:

- Conficker
  - 50 thousand domain names an hour
- Torpig
  - Randomly generated from Twitter
- Kraken



oubezepg.com.mx  
gzyh.nl  
ktmybhf.sg  
psexy.com.mx  
cvipcw.com.ve  
vcos.com.gt  
jbdtoou.in  
iceorofq.com.hn  
nxeir.as  
gdeeuphz.ch  
yqzzag.yi.org  
yqcoqgmmb.yi.org  
lghuuuvwoct.yi.org  
ufizpvq.yi.org  
hicsac.dyndns.org  
hjviglrwd.dyndns.org  
hmhauqssekw.dyndns.org  
hnkoso.dyndns.org  
putv.moou.com  
tuoswxcoclw.moou.com  
tzvhyc.moou.com  
ufdwpgvj.moou.com

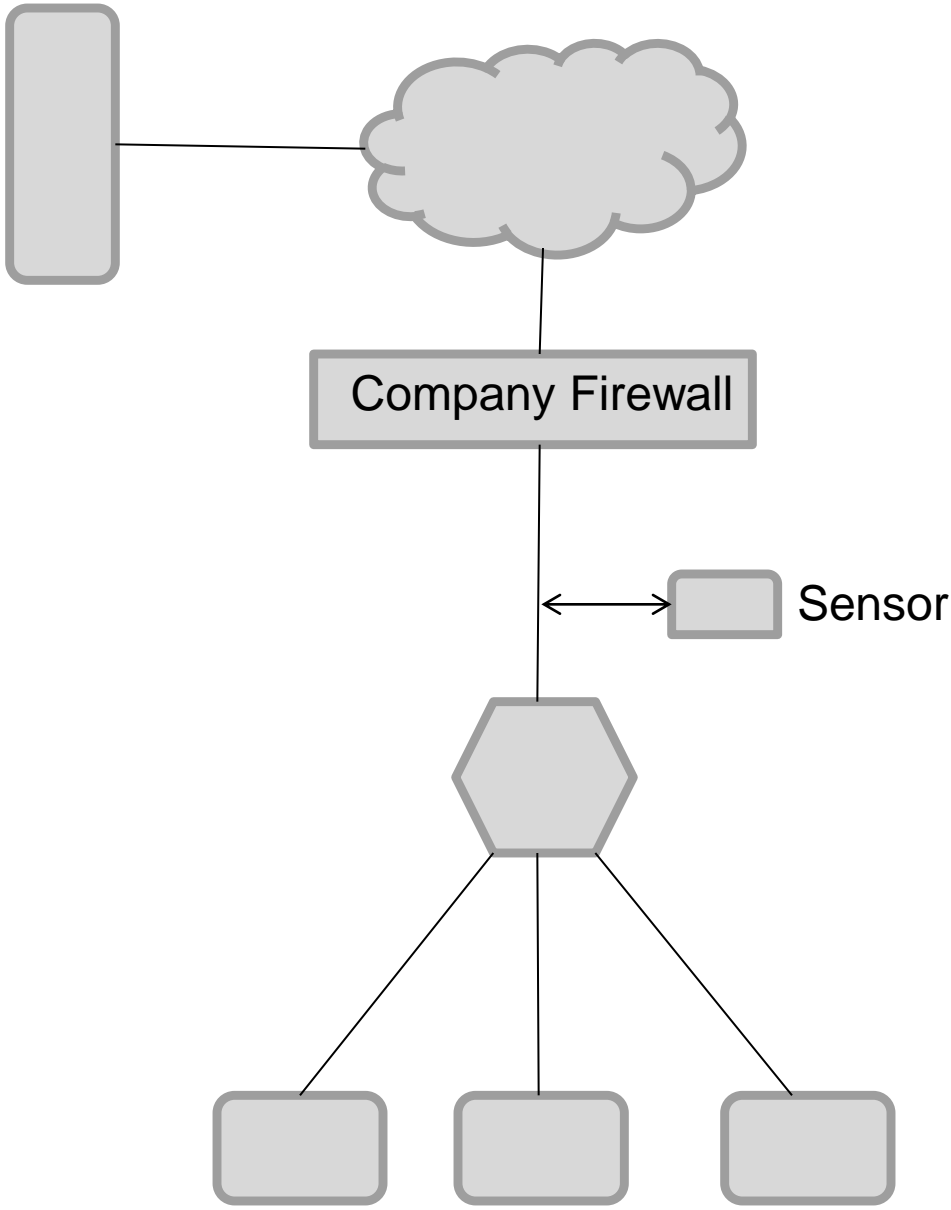
# System with Internal DNS Server



Infected Host

# System with External DNS Server

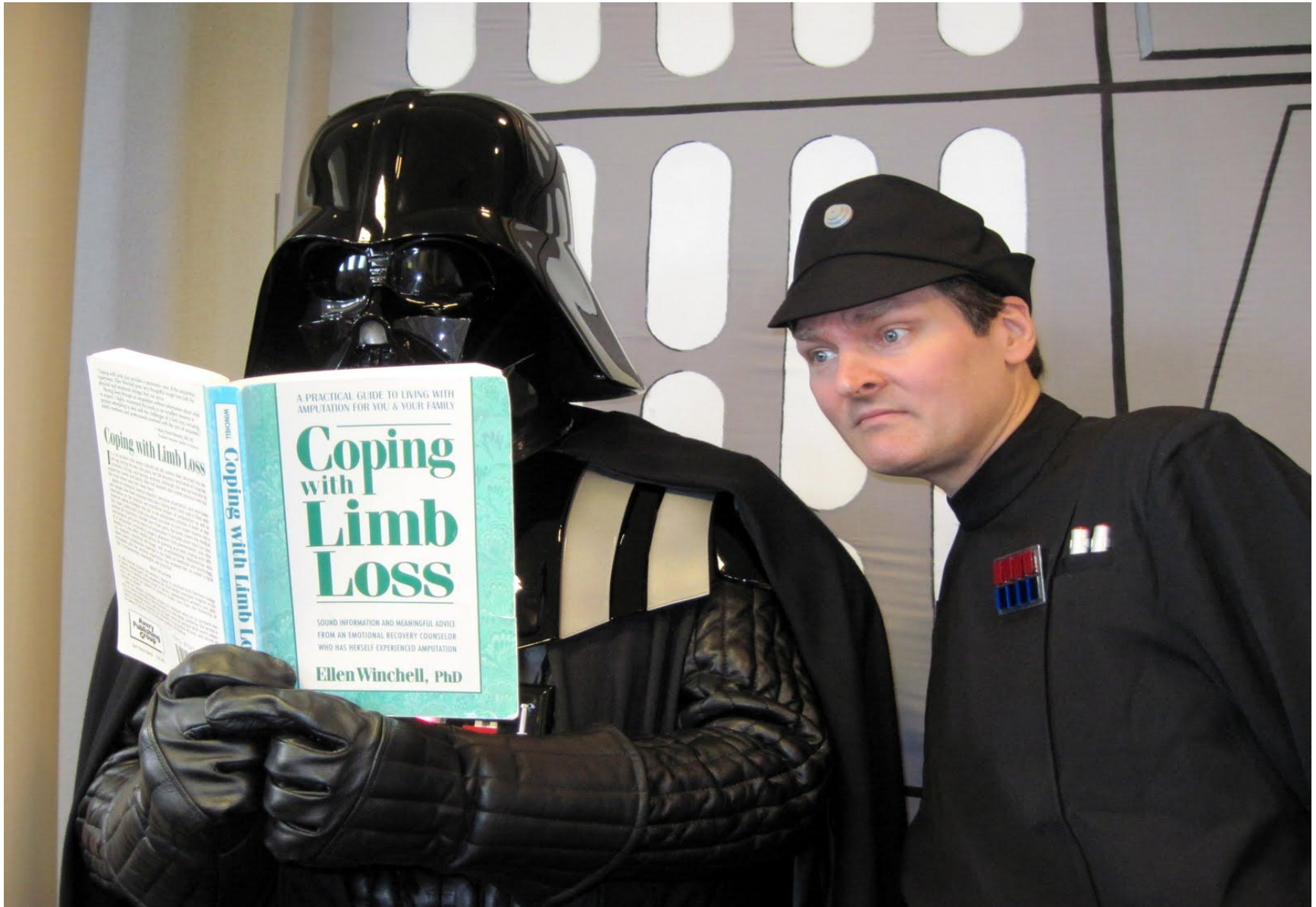
DNS Server



Infected Host

- Infected host sends DNS Query
- DNS Server returns Query Response
- Sensor reads Query Response
- if Malicious:
  - #send empty response to host or drop
  - #add domain to block-list
  - #notify administrator
- else:
  - #forward query response
  - #host establishes connection as normal

# DNS Query Response Analysis



## DNS Metrics Examined

- Short TTL
- Multiple Addresses (A Records)
- Different IP ranges
- Multiple Autonomous System Numbers (ASNs)
- Number of Name-servers (NS Records)
- Name-servers in different network ranges



## Average values observed for each metric

|            | A Records | NS Records | Number of IP Ranges | Number of ASNs | TTL      |
|------------|-----------|------------|---------------------|----------------|----------|
| Fast-flux  | 2.090032  | 3.916399   | 2.180064            | 3.70418        | 594.9968 |
| Legitimate | 1.730769  | 3.87574    | 0.1538462           | 1.094675       | 14885.42 |

# Classifier Options

- Manual inspection
  - Why when it can be automated?
- Signature
  - Needs to be kept up to date
- C5.0 decision tree
  - Similar to signature
  - A bit more “Intelligence” behind it
- Naïve Bayesian classifier
  - Evolve along with botnets

$$\ln \frac{P(F|D)}{P(\neg F|D)} = \ln \frac{P(F)}{P(\neg F)} + \sum_i \ln \frac{P(t_i|F)}{P(t_i|\neg F)}$$

## Where

- $\ln \frac{P(F|D)}{P(\neg F|D)}$  logarithmic probability ratio that a domain is Fast-flux or not
- $P(t_i|F)$  probability that a token appears in fast-flux DNS query

# \$ dig fanarm.net

QUESTION

```
fanarm.net.      IN  A
```

;ANSWER

```
fanarm.net. 300 IN A 71.35.101.107
fanarm.net. 300 IN A 71.37.48.123
fanarm.net. 300 IN A 195.214.238.241
fanarm.net. 300 IN A 219.95.36.17
fanarm.net. 300 IN A 41.222.11.122
```

;AUTHORITY

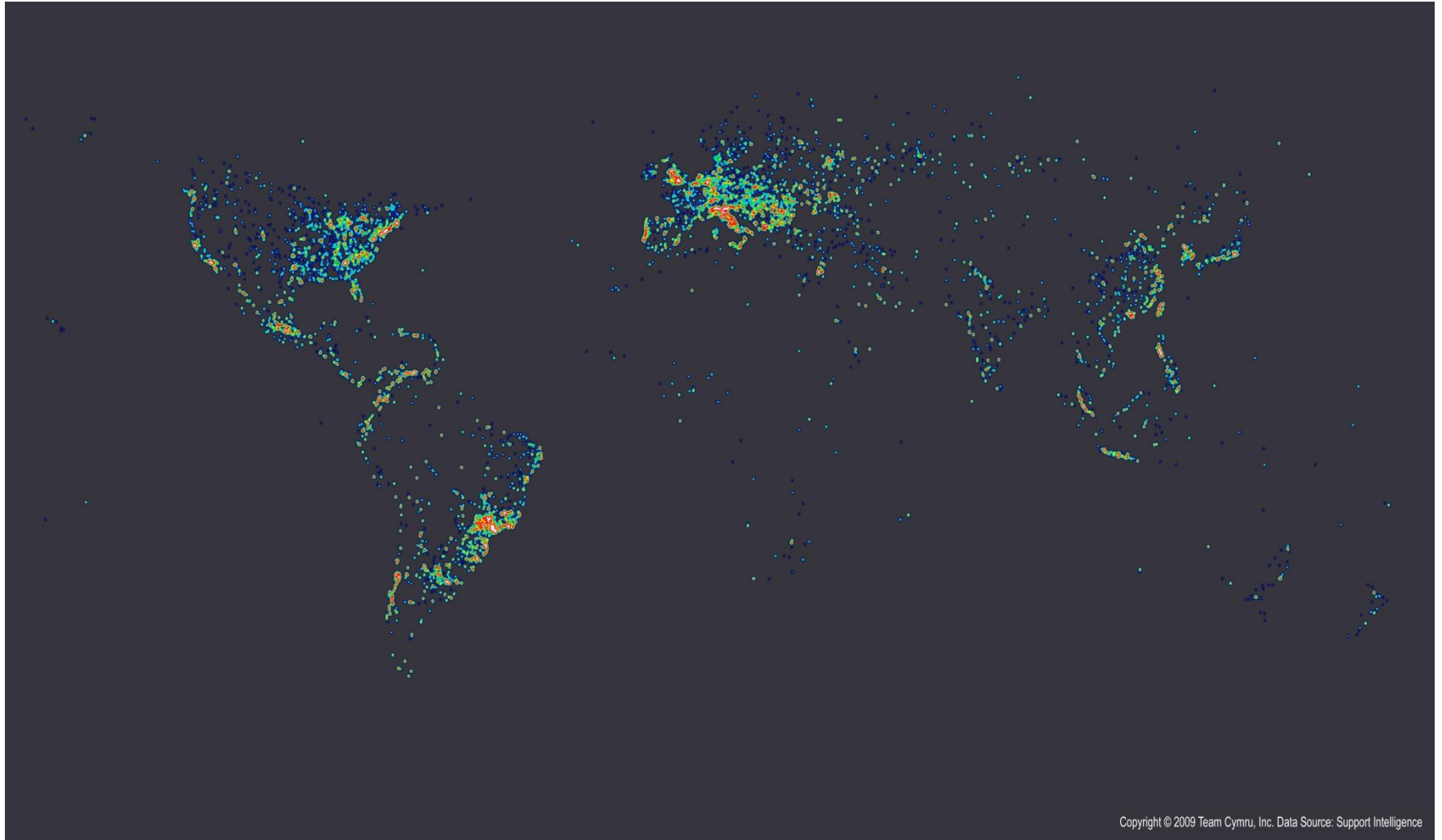
```
fanarm.net. 300 IN NS ns1.flickingers.net.
fanarm.net. 300 IN NS ns2.flickingers.net
```

| ASN   | Net-block        | Country | Registrar |
|-------|------------------|---------|-----------|
| 209   | 71.32.0.0/13     | US      | arin      |
| 209   | 71.32.0.0/13     | US      | arin      |
| 24881 | 195.214.236.0/22 | UA      | ripencc   |
| 4788  | 219.95.0.0/17    | MY      | apnic     |
| 36866 | 41.222.8.0/21    | KE      | afrinic   |

- Outputs a likelihood ratio
  - Malicious or not
  - Combined with other classifiers
  - Learns as botnet DNS queries change

| Domain            | Safe Score   | Malicious Score | Classification |
|-------------------|--------------|-----------------|----------------|
| gingerbucksea.com | 0.005304578  | 0.3550235       | fast-flux      |
| pearlrumor.ru     | 3.059976e-14 | 7.490562e-13    | fast-flux      |
| wordpress.com     | 1.536894e-08 | 4.250896e-10    | legitimate     |
| champiogogo.ru    | 3.395984e-09 | 1.723838e-06    | fast-flux      |
| yahoo.com         | 1.940412e-15 | 1.509179e-69    | legitimate     |

# Domain Name Analysis

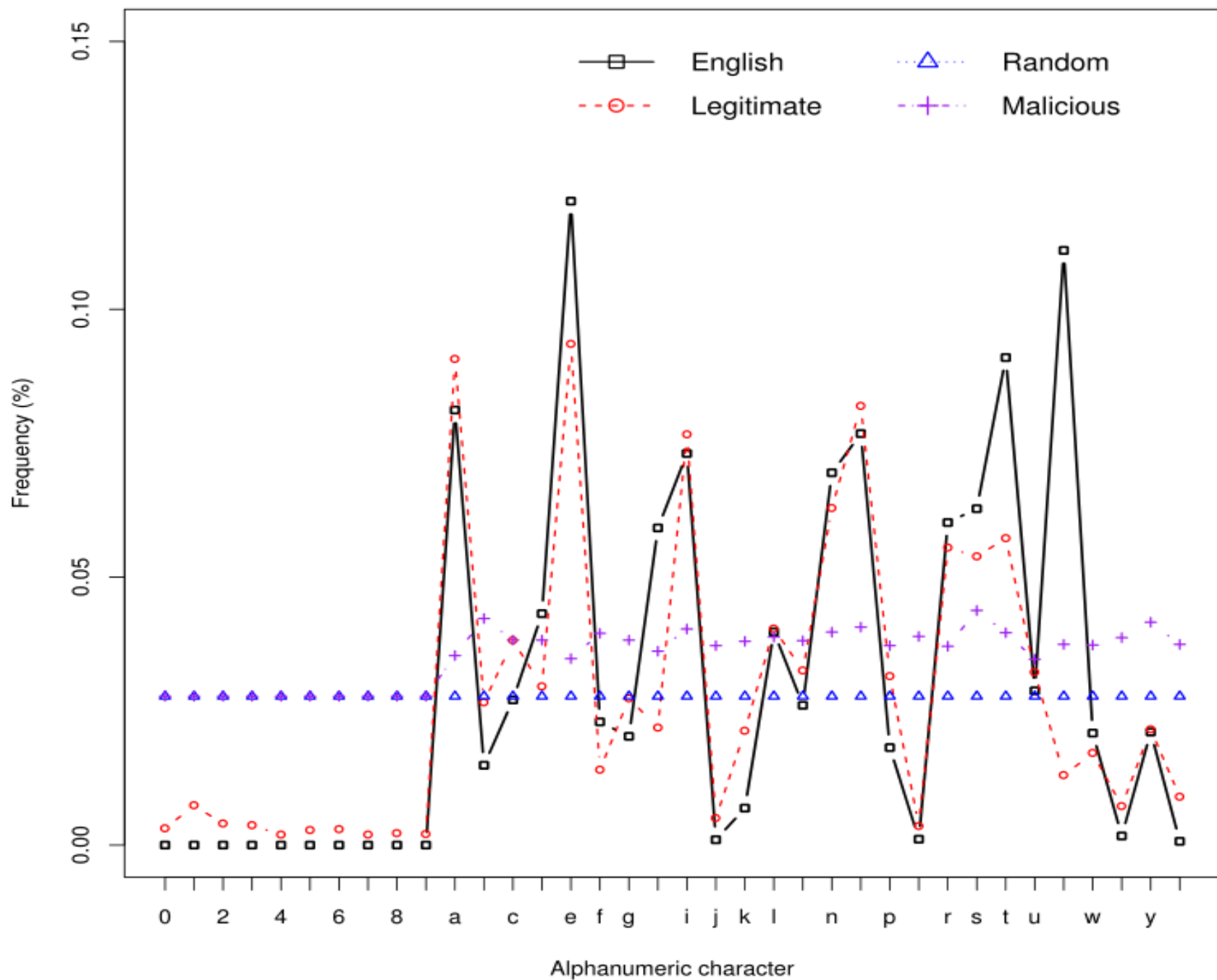


Copyright © 2009 Team Cymru, Inc. Data Source: Support Intelligence

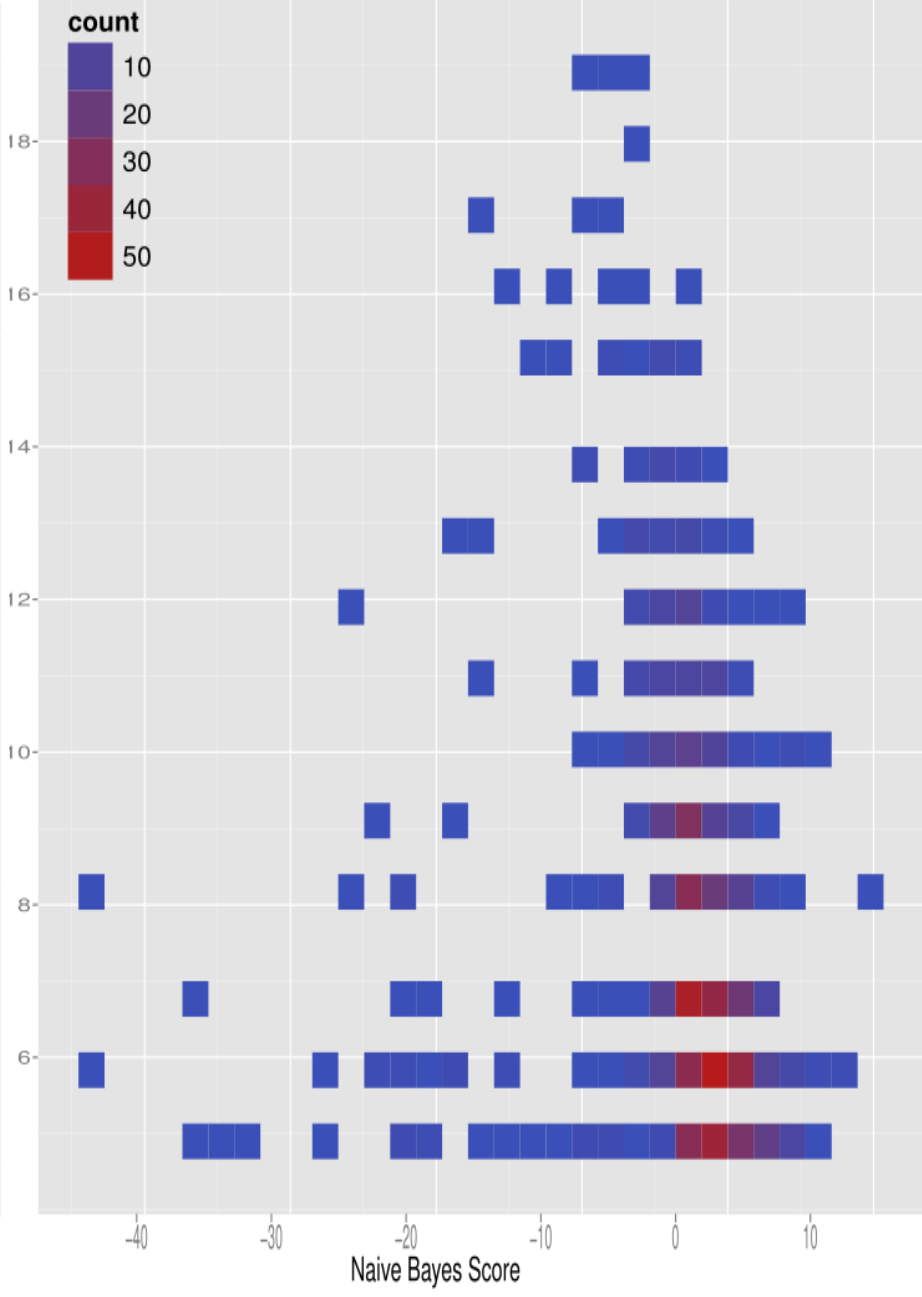
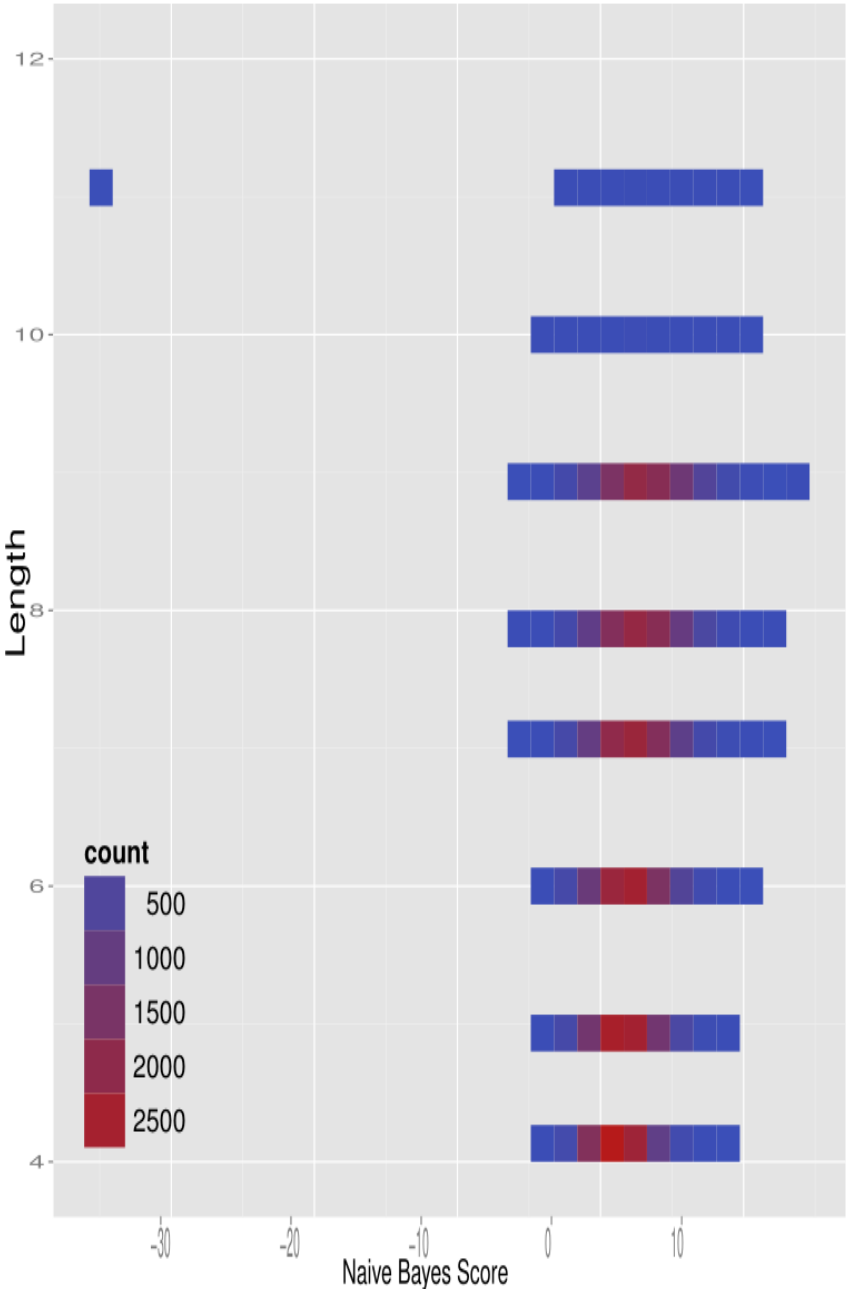
<http://www.team-cymru.org/images/conficker-2009-01-29-dark-full.jpg>

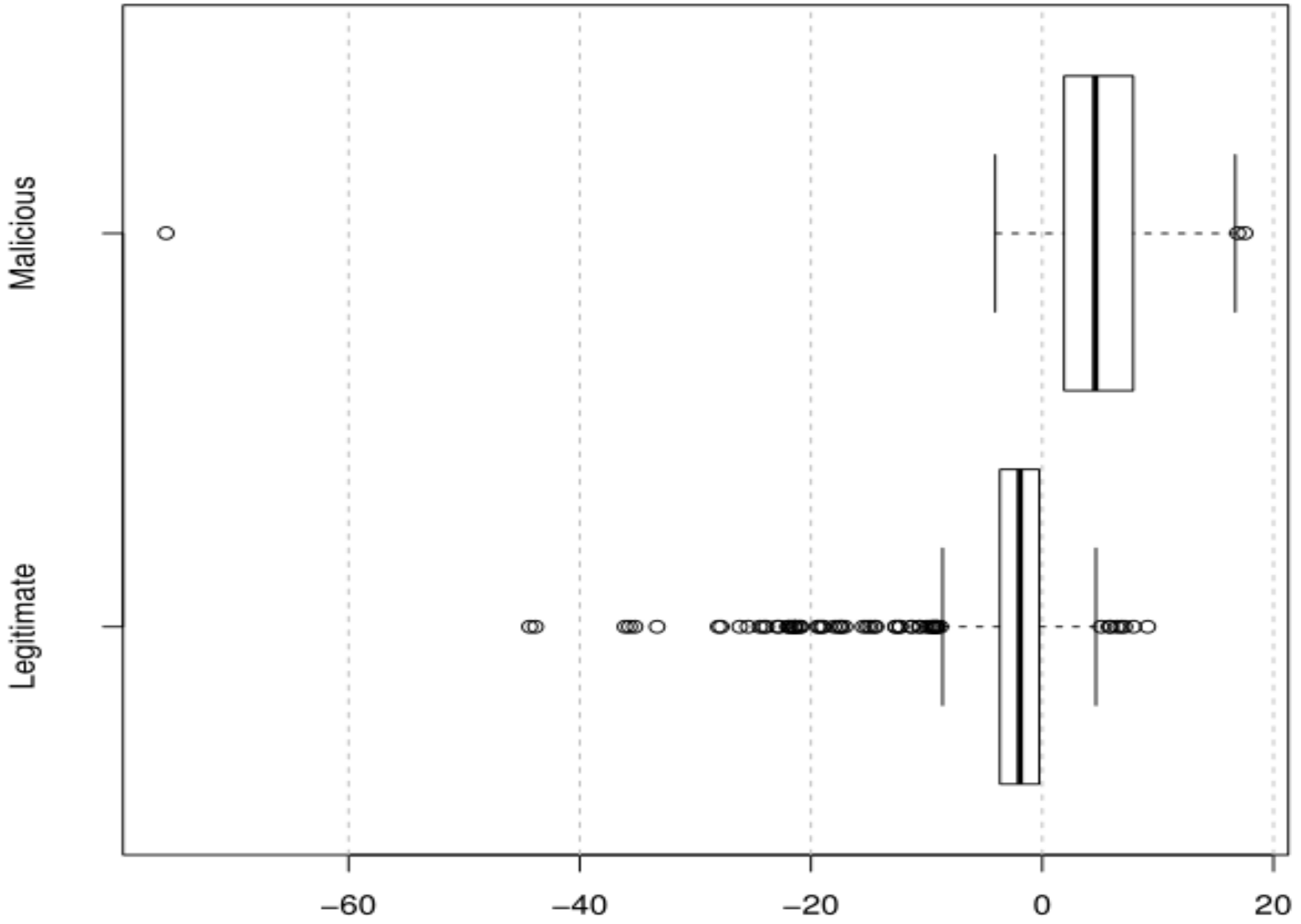
- Detect algorithmically generated domain names
- English words/domain-names have different character distribution from algorithmically generated words/domain-names
- [ighuuvwoct.yi.org](http://ighuuvwoct.yi.org)
  - Examine ***ighuuvwoct***
  - Use frequency distribution of characters
  - Bayesian statistics creates a learning system

# CHARACTER DISTRIBUTION



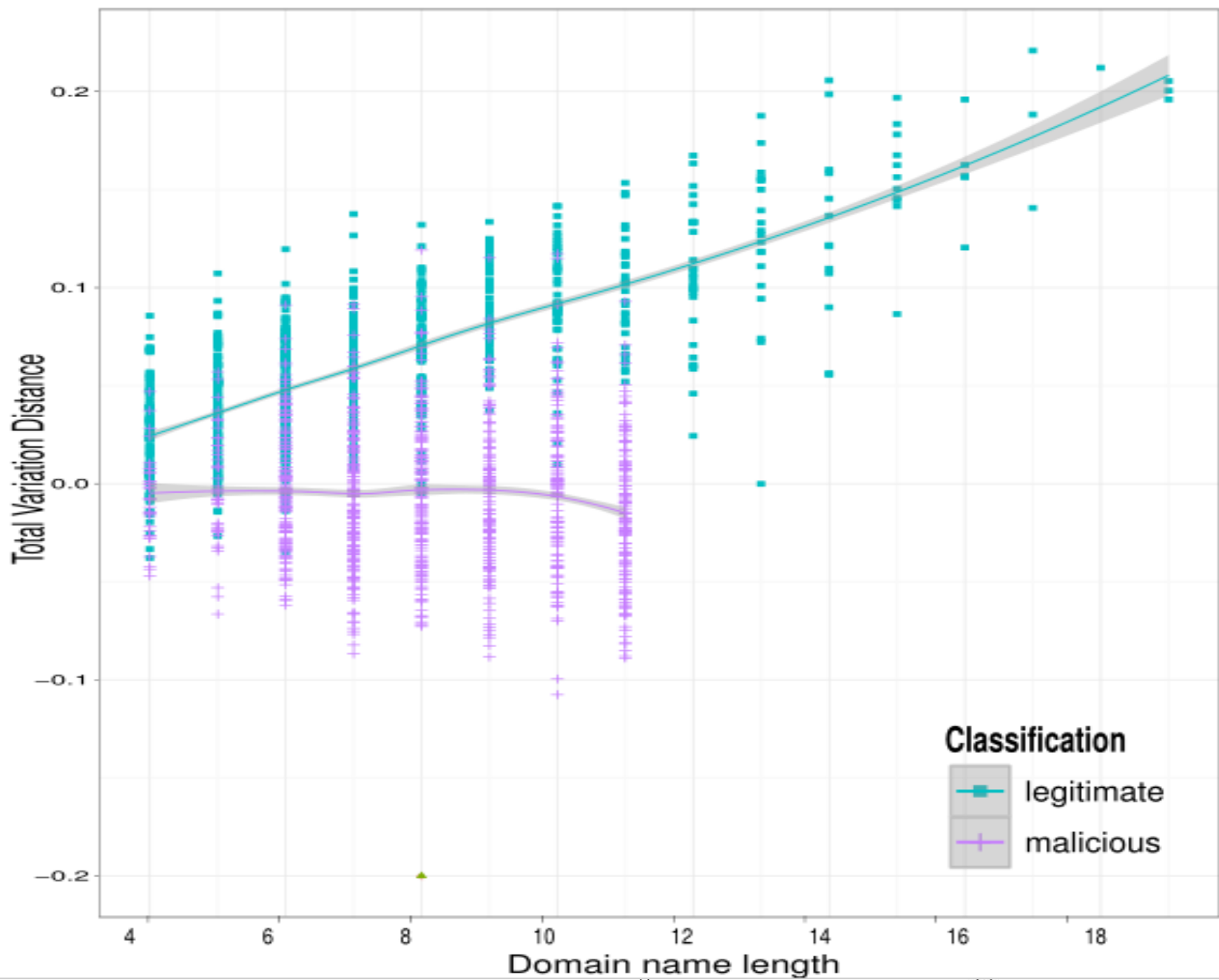
- Probability
  - Standard statistics
  - Add up all probabilities of characters appearing in malicious and legitimate domain
- Total Variation Distance
  - maximum possible difference between two probability distributions
- Naïve Bayesian
  - Assumes no prior bias
- Bayesian
  - Use result from other classifiers to set bias





| Domain name     | Output    | Classification | Correct |
|-----------------|-----------|----------------|---------|
| facebook.com    | -1.06400  | Legitimate     | Yes     |
| allrecipes.com  | -4.25654  | Legitimate     | Yes     |
| twitter.com     | -2.398181 | Legitimate     | Yes     |
| buzzel.com      | 2.47540   | Malicious      | No      |
| nhk.or.jp       | 0.64375   | Malicious      | No      |
| nbhkxkjh.com.fj | 6.61512   | Malicious      | Yes     |
| pveufjtm.com.bo | 3.25285   | Malicious      | Yes     |
| rrxwigqj.am     | 5.24226   | Malicious      | Yes     |
| ljtkrinq.com    | 2.75078   | Malicious      | Yes     |

# TOTAL VARIATION DISTANCE



# Classifiers Compared

| Classifier     | Accuracy | TPR | FPR |
|----------------|----------|-----|-----|
| Naïve Bayesian | 87%      | 82% | 8%  |
| Variation      | 82%      | 80% | 17% |
| Probability    | 84%      | 86% | 17% |
| Bayesian       | 85%      | 81% | 11% |

TPR = True Positive Rate

FPR = False Positive Rate

- Test how well the system scales
  - ISP level?
- Use artificial neural network to train new classifiers
- Examine bi-gram character distributions
- Ideas?



