DNS Anomaly Detection

superDAD

Nick Barendregt
Hidde van der Heide

DAD

Agenda

- Introduction
- Methods
- Results
- Conclusion
- Questions and Discussion

Introduction

"Examine the feasibility of detecting malware infected systems using DNS log data and develop a scheme for detecting these anomalies in DNS traffic.

Develop a simple proof of concept capable of processing text based output from our DNS logger."

Methods

- Non-DNS packets on port 53
- MX requests
- Keyword detection
- Blacklists

- Covert channel (DNS tunnel) detection
- Character frequency analysis
- Fast-flux detection
- Timing analysis
- Scoring mechanism

DNS Tunnel Detection

Characteristics

- Non DNS data
- Large number of packets
- Large packets
  - Long domain names
  - Large strings in NULL or TXT records
- Random data when compressed or encrypted

DNS Tunnel Detection

- Configure Iodine (tunnel DNS software)
- Downstream modes:
  - Raw UDP
  - NULL (experimental)
  - TXT
  - CNAME
  - A
  - etc.
- Encoded Base32/64/128

Character Frequency Analysis

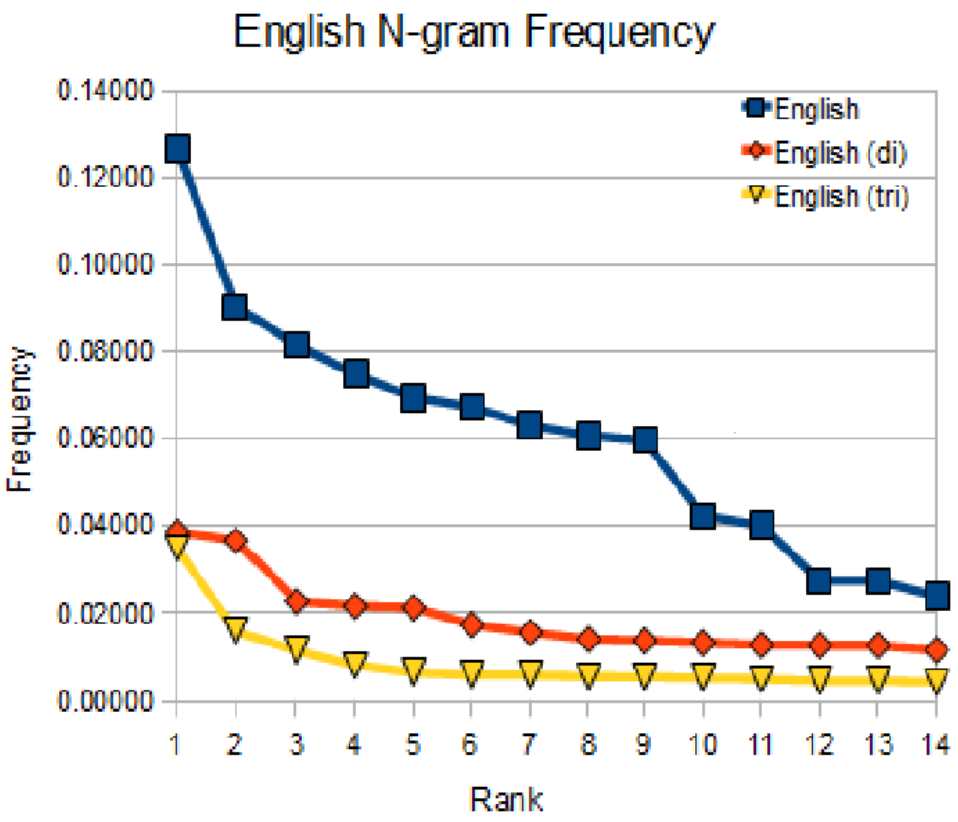
English Unigrams	
LETTER	FREQUENCY
e	0.12702
t	0.09056
a	0.08167
o	0.07507
i	0.06966
n	0.06749
s	0.06327
h	0.06094
r	0.05987
d	0.04253
l	0.04025
c	0.02758
u	0.02758
m	0.02406

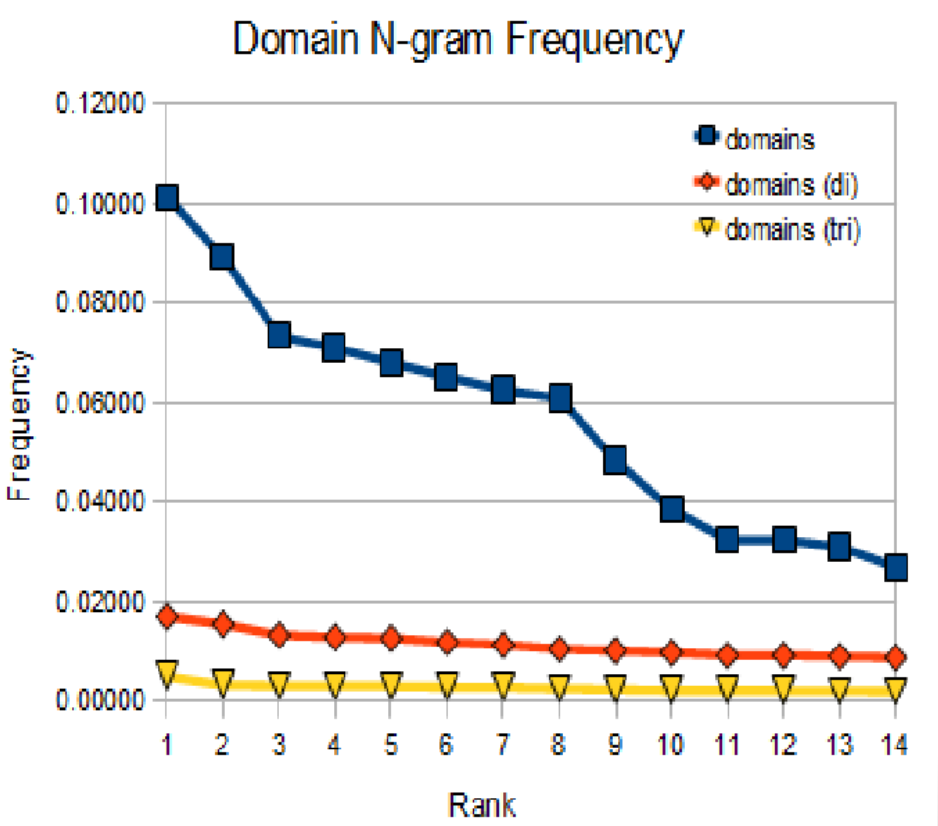
Domain Unigrams	
LETTER	FREQUENCY
e	0.10139
a	0.08935
i	0.07346
o	0.07105
s	0.06804
r	0.06524
t	0.06263
n	0.06094
l	0.04849
c	0.03861
m	0.03249
d	0.03247
u	0.03105
p	0.02689

English Bigrams	
LETTER	FREQUENCY
th	0.03883
he	0.03681
in	0.02284
er	0.02178
an	0.02141
re	0.01749
nd	0.01572
on	0.01418
en	0.01383
at	0.01336
ou	0.01286
ed	0.01276
ha	0.01275

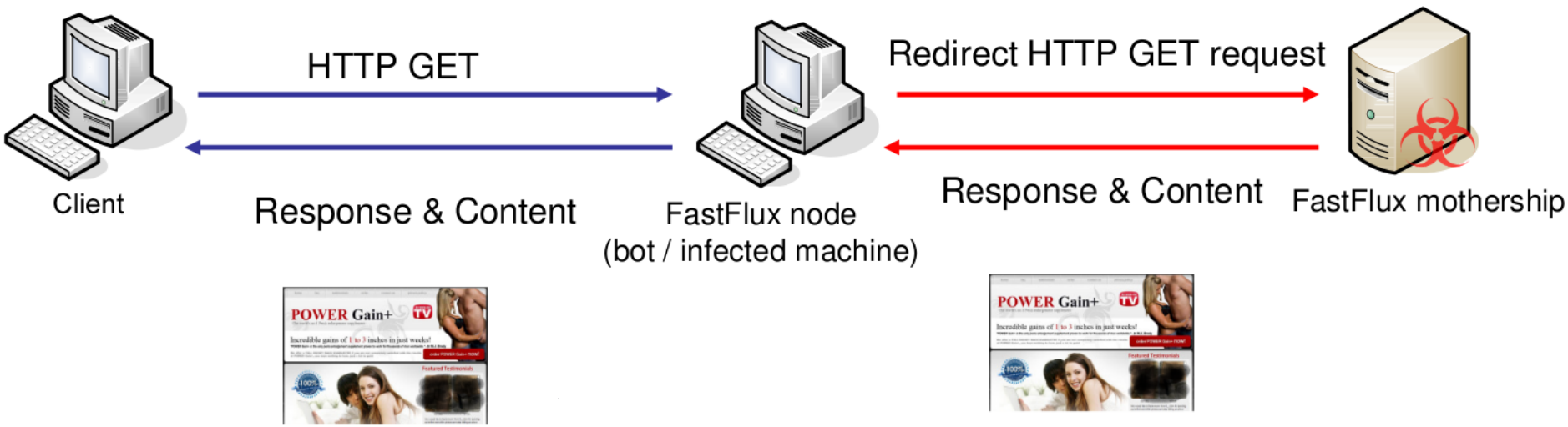
Domain Bigrams	
LETTER	FREQUENCY
in	0.01702
er	0.01550
an	0.01333
re	0.01290
es	0.01271
ar	0.01188
on	0.01135
or	0.01051
te	0.01017
al	0.00976
st	0.00921
ne	0.00921
en	0.00897

Character Frequency Analysis





Fast-Flux Detection



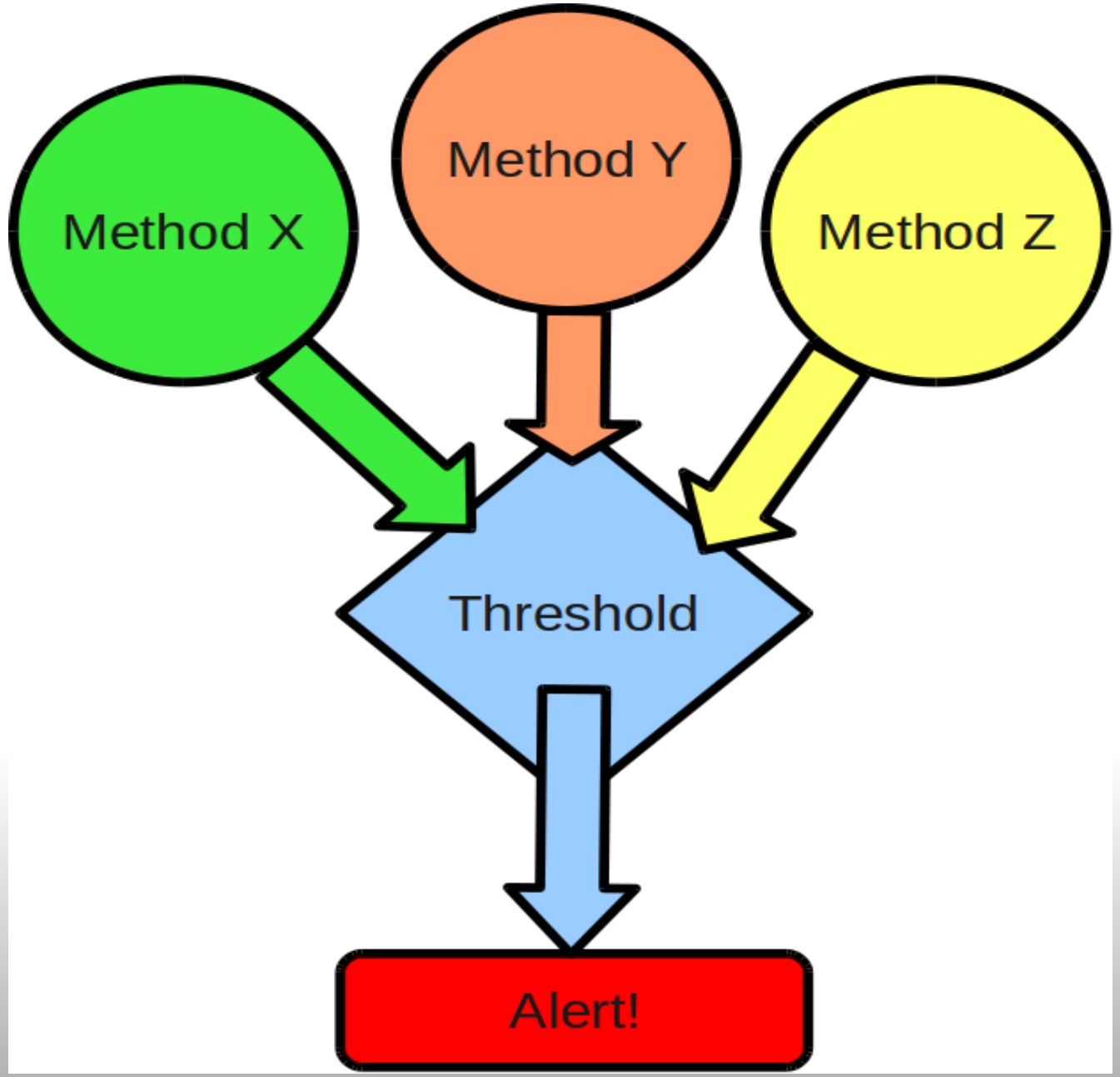
Fast-Flux Detection - Example

```
$ dig naughtydateingsite.net
;; ANSWER SECTION:
naughtydateingsite.net.    300    IN    A    77.127.166.235
naughtydateingsite.net.    300    IN    A    82.228.65.61
naughtydateingsite.net.    300    IN    A    84.109.81.176
naughtydateingsite.net.    300    IN    A    92.253.40.134
naughtydateingsite.net.    300    IN    A    94.54.254.3
naughtydateingsite.net.    300    IN    A    94.228.118.59
naughtydateingsite.net.    300    IN    A    114.33.131.22
naughtydateingsite.net.    300    IN    A    118.101.225.28
naughtydateingsite.net.    300    IN    A    201.167.15.123
naughtydateingsite.net.    300    IN    A    203.99.233.142
;; AUTHORITY SECTION:
naughtydateingsite.net.    172318    IN    NS    ns1.7418391.com.
naughtydateingsite.net.    172318    IN    NS    ns2.7418391.com.
naughtydateingsite.net.    172318    IN    NS    ns3.7418391.com.
naughtydateingsite.net.    172318    IN    NS    ns4.7418391.com.
naughtydateingsite.net.    172318    IN    NS    ns5.7418391.com.
naughtydateingsite.net.    172318    IN    NS    ns6.7418391.com.
; ADDITIONAL SECTION:
ns1.7418391.com.    85917    IN    A    173.212.75.160
ns2.7418391.com.    85917    IN    A    79.119.188.9
ns3.7418391.com.    85917    IN    A    88.87.251.45
ns4.7418391.com.    85917    IN    A    82.228.65.61
ns5.7418391.com.    85917    IN    A    79.117.122.25
ns6.7418391.com.    85917    IN    A    186.114.80.139
```

DNS Timing Analysis

- Group activity
- Regular queries (polling)
- Outside office hours

Scoring Mechanism



Results

- DNS Tunnel Detection
- Single Flux Detection
- Double Flux Detection

DNS Tunnel Detection

- Configured DNS tunnel software
- Captured stream of scp 10Mb random data
- Loaded in memory with Python Scapy
- Created frequency distribution graphs with NLTK toolkit
- Compare:
  - Other tunnel software
  - Frequency distribution for top sites
  - Frequency distribution for language

DNS Tunnel Detection - Base 32

Tunnel DNS Dump Unigrams Base32	
Letter	Frequency
d	0.09367
a	0.08899
m	0.07194
q	0.07153
r	0.06279
b	0.05572
g	0.04647
w	0.04637
h	0.04044
y	0.03982
c	0.03971
f	0.03951
t	0.03909
2	0.02817

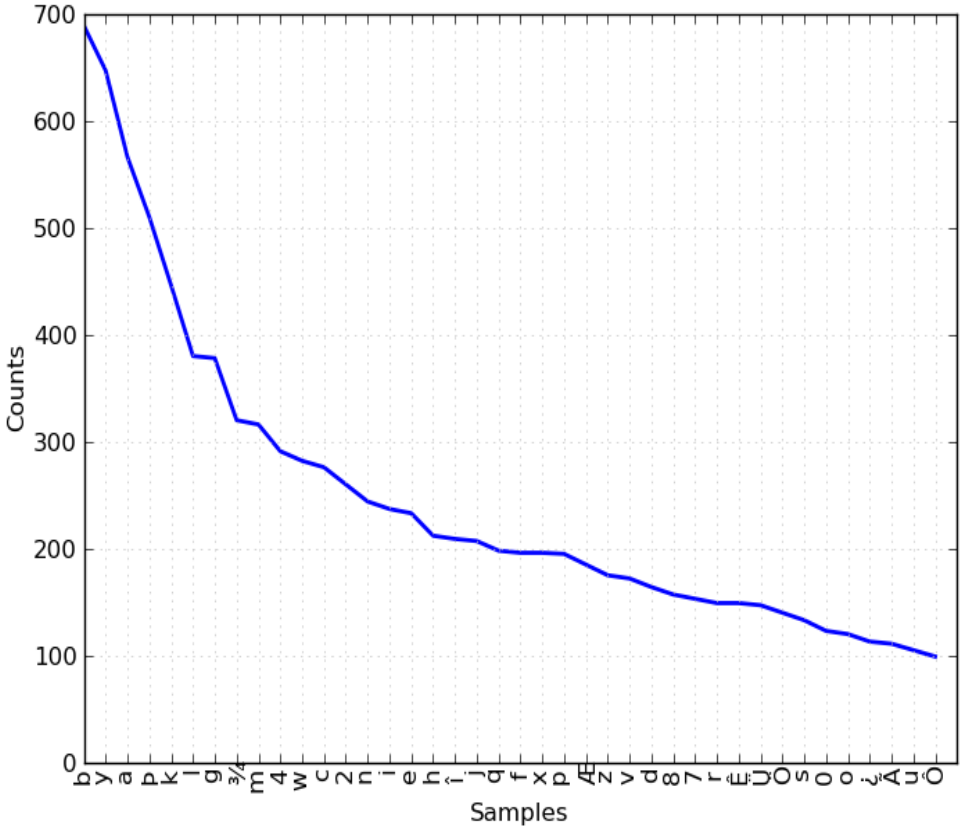
Domain Unigrams	
LETTER	FREQUENCY
e	0.10139
a	0.08935
i	0.07346
o	0.07105
s	0.06804
r	0.06524
t	0.06263
n	0.06094
l	0.04849
c	0.03861
m	0.03249
d	0.03247
u	0.03105
p	0.02689

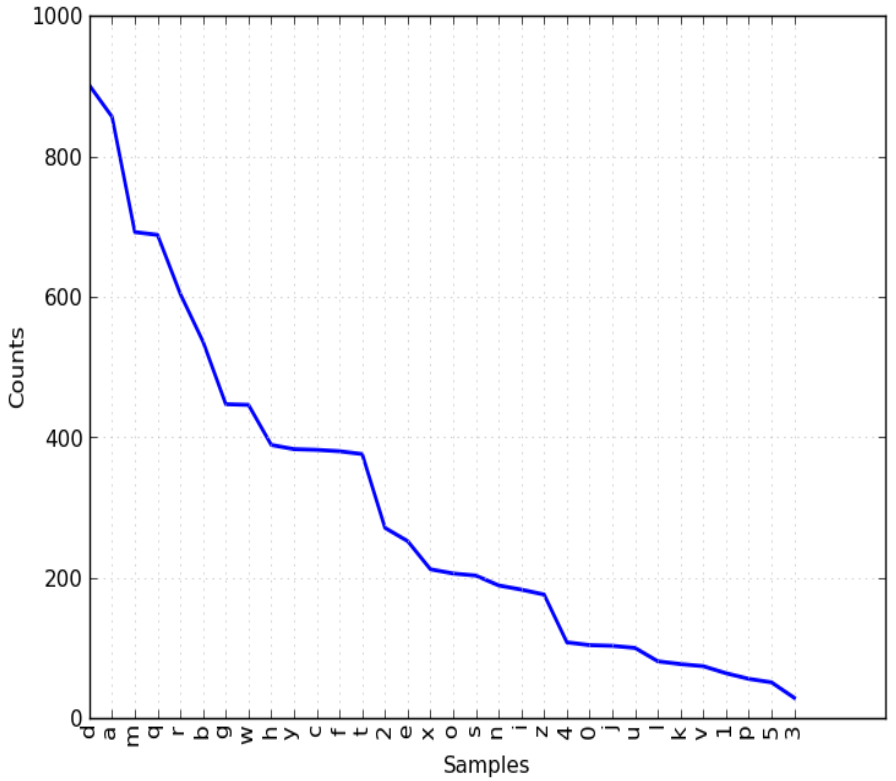
DNS Tunnel Detection - Base 128

Tunnel DNS Dump Unigrams Base128	
Letter	Frequency
b	0.05615
y	0.05273
a	0.04613
Þ	0.04156
k	0.03635
l	0.03097
g	0.0308
¾	0.02608
m	0.02575
4	0.02371
w	0.02298
c	0.02249
2	0.02119
n	0.01988

Domain Unigrams	
LETTER	FREQUENCY
e	0.10139
a	0.08935
i	0.07346
o	0.07105
s	0.06804
r	0.06524
t	0.06263
n	0.06094
l	0.04849
c	0.03861
m	0.03249
d	0.03247
u	0.03105
p	0.02689

DNS Tunnel Detection

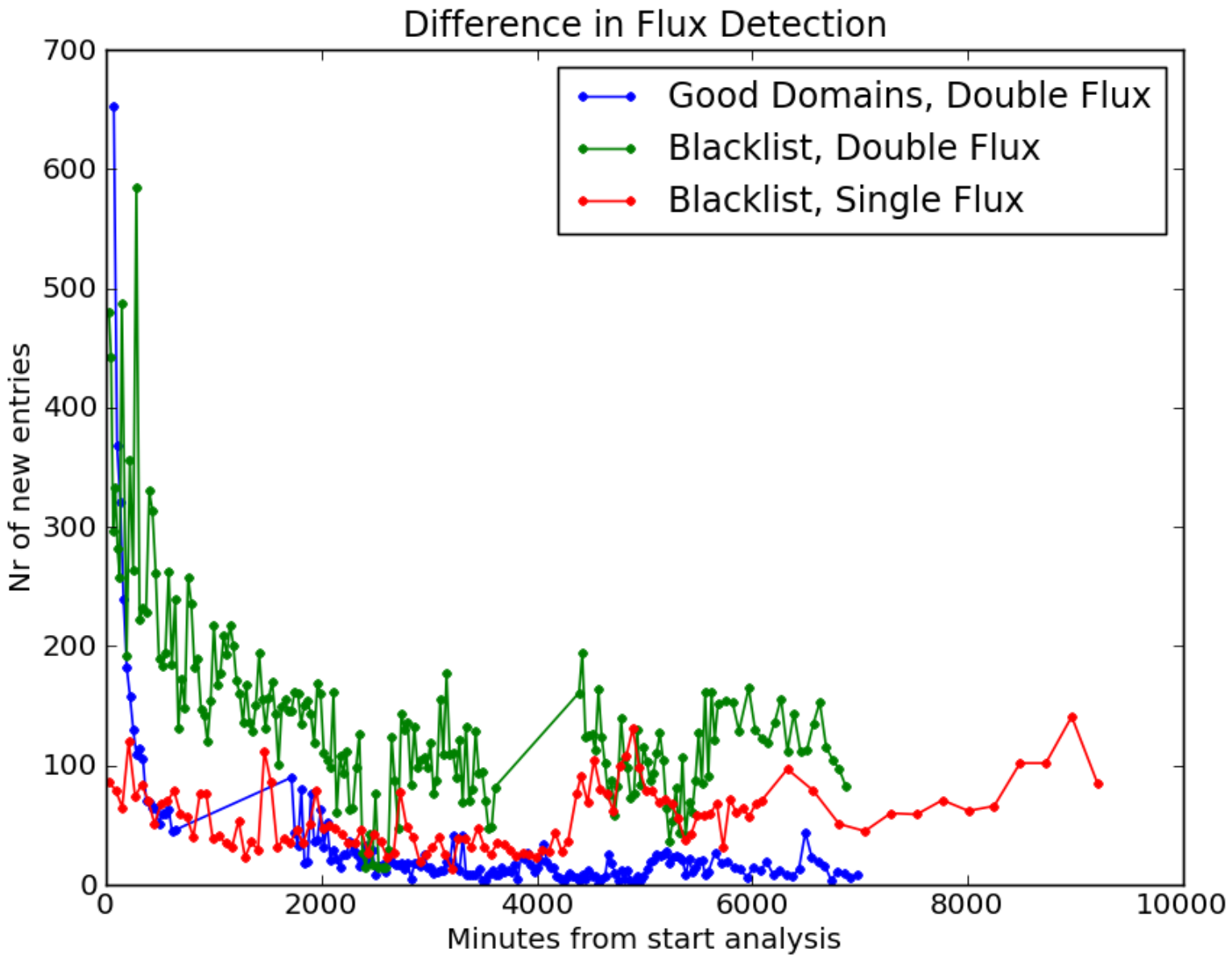




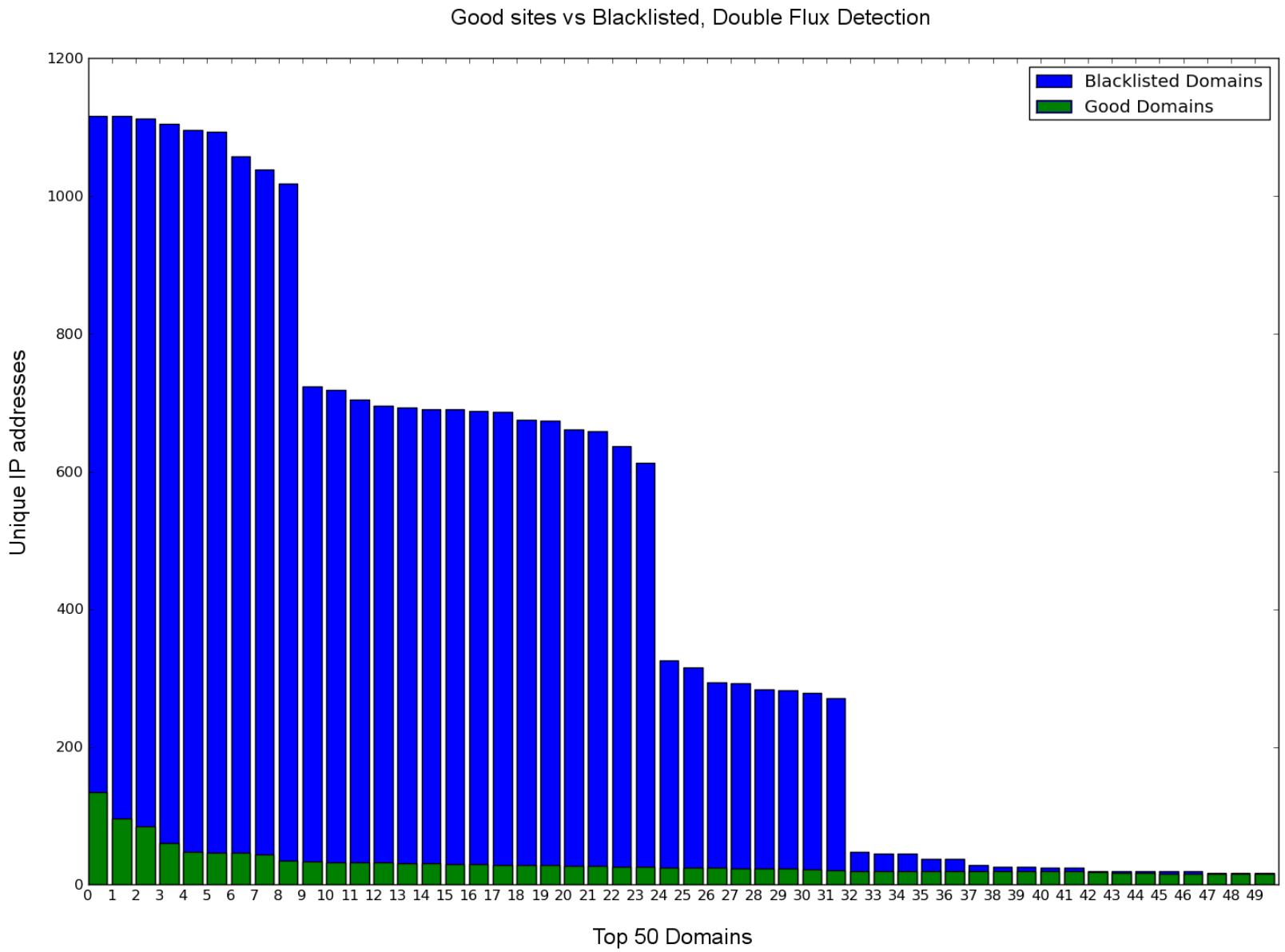
Fast-flux Detection

- Single Flux Detection
  - Simple bash system nslookup
  - Threaded python nslookup
- Double Flux Detection
  - DNS library
  - SOA Record
  - A Record
  - NS Record
    - ANY Record
- Database
  - Lookup previous entries
  - Takes time with more data

Fast Flux Detection



Fast Flux Detection



Conclusion

- Promising methods need to be done off-line
- The amount of data needed for proper time analysis becomes problematic
- Best probe position would be at the network border since TTL is unreliable
- Good results for methods, better when combined

- Yes!

Future Work

- Create full working tool
- Research best scoring mechanism
- Timing analysis
- Live data

Fun Facts

```
Single: 116 x 1 x 10.728 = 1.244.448
Double: 174 x 3 x 10.728 = 5.600.016
Good  :  22 x 3 x 10.000 =   660.000 +
Total domain queries:      7.504.464
Extra 48 hour run:        ~2.400.000

Tracked domains:             10.728
Unique IP addresses:         32.466
Total amount of time spend: ~5.000 minutes

Lines of code:              ~1500
Cups of coffee:              2 x 20 x ~4 = ~160
Research papers read:       ~30
```

Questions and Discussion