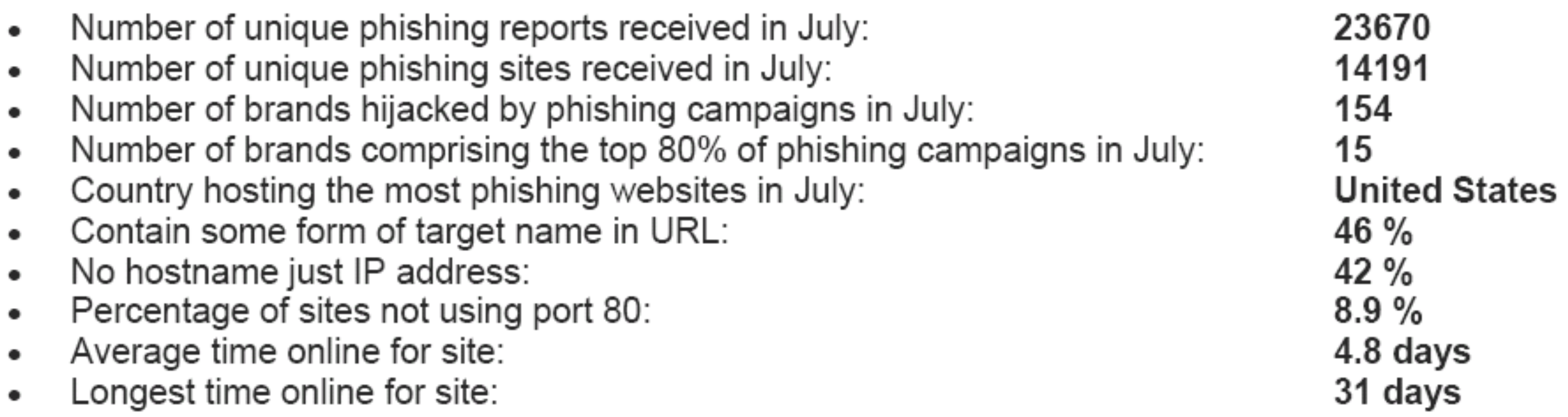
Analyzing the Effectiveness of Phishing at Network Level

Sagar Mehta, Nitya Sundareswaran, Kevin D. Fairbanks, Nick Feamster

Motivation



• Number of unique phishing reports received in July:	23670
• Number of unique phishing sites received in July:	14191
• Number of brands hijacked by phishing campaigns in July:	154
• Number of brands comprising the top 80% of phishing campaigns in July:	15
• Country hosting the most phishing websites in July:	United States
• Contain some form of target name in URL:	46 %
• No hostname just IP address:	42 %
• Percentage of sites not using port 80:	8.9 %
• Average time online for site:	4.8 days
• Longest time online for site:	31 days

- **Source - Phishing Activity Trends Report July, 2006 , Anti-Phishing workgroup**

- **Our work done from Jan 07 – Apr 07**

Related Work

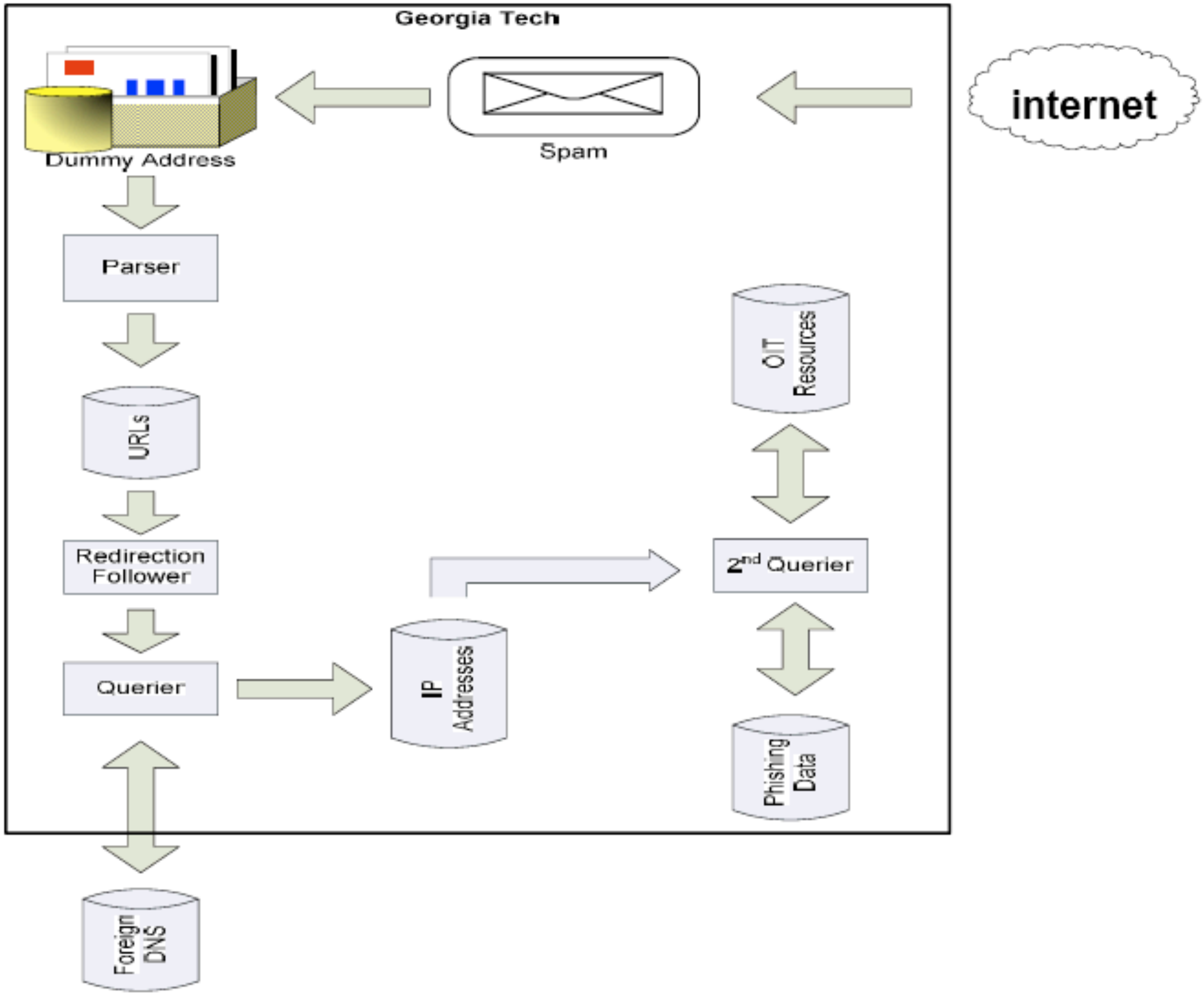
Mostly at application layer

- Why phishing works ? – Dhamija et al
- The Battle Against Phishing:Dynamic Security Skins - Dhamija et al
- Detection of Phishing pages based on visual similarity - Liu et al
- Phoney: Mimicking User Response to Detect Phishing Attacks -Chandrasekaran et al
- A Framework for Detection and Measurement of Phishing Attacks - Doshi et al
- Anti-Spam Techniques

Problem Statement

- Looking at the effectiveness of Phishing from network level = Complementary approach to application layer analysis
  - Correlate Phishing mails to outgoing traffic
  - Analyze traffic destined to Phishing sites

System Architecture



Data sources

- Spam Trap data
- Netflow Records
- DNS cache

Parsing script

- Parsing script to obtain urls from spam

- Filter using heuristics to obtain phishing urls
  - anchor text and actual link disagree
  - redirection – http 302, meta keyword
  - presence of certain keywords
  - presence of ip address in place of domain name

- Caveats:
  - Human intervention for correct interpretation of URL
    - http://www.example-com, Replace “-”with “.” In the above link
    - http://www.example .com, Remove space in the above link
  - Attached .jpg images that provide the URL address – no OCR
  - Deceptive user names e.g. ‘www.example1.com@example2.com’

Querying Script

Querying script to map phishing domains to IP addresses

Simulating HTTP client to follow redirects

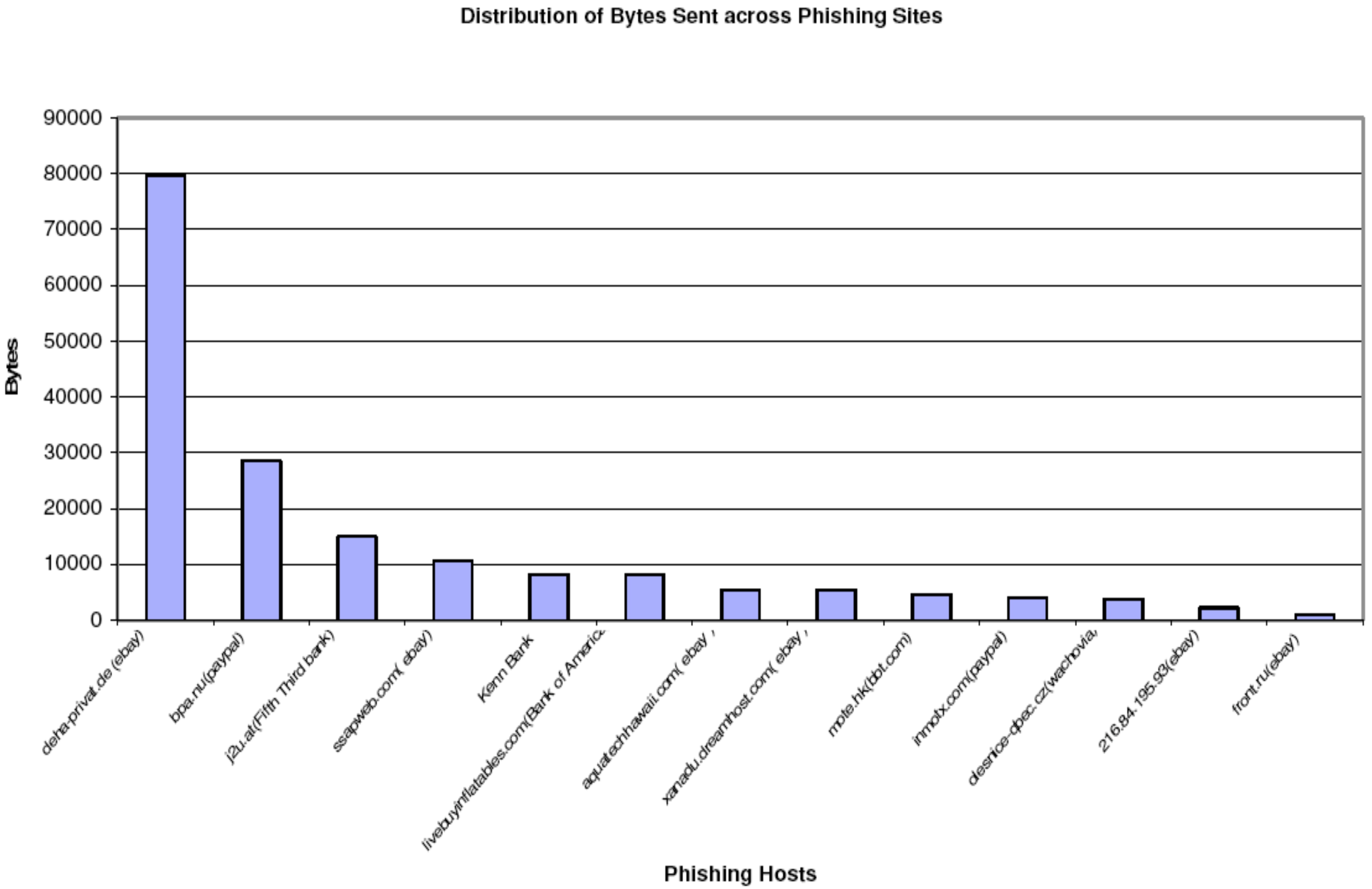
- Status code 300-307 in HTTP response
- Meta redirects

Caveat

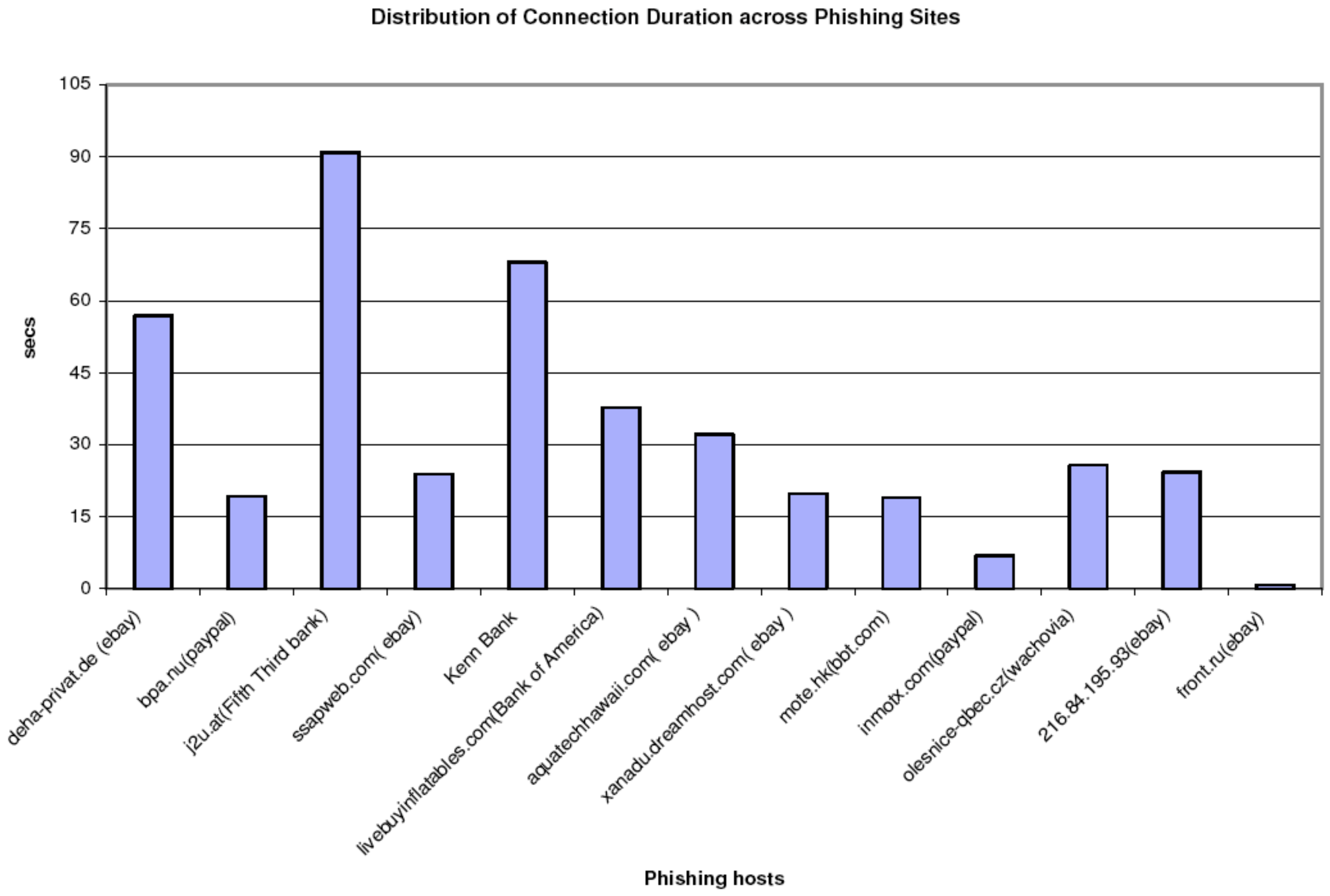
- Avoid corrupting the trace while mapping phishing domains to IP addresses by directing queries to a foreign name server

Extracted ip addresses to further query netflow data from GTRNOC to get netflow tuples using src ip, src port , dest ip, dest port as ‘key’

Interaction with known phishing Sites from PhishTank – wide varation in byte distribution even when interacting with sites imitating the same website

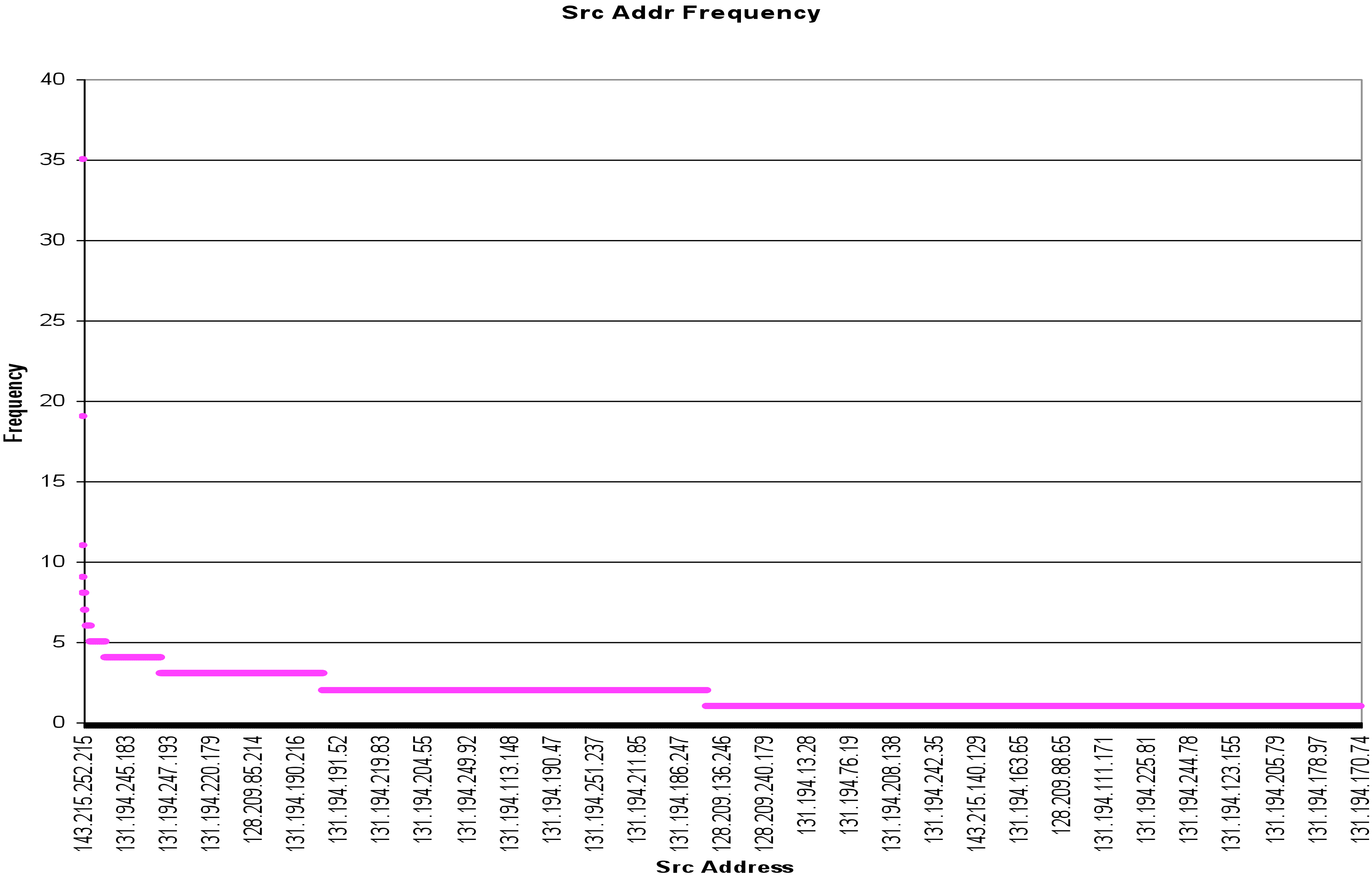


Similar variation in connection time distribution even when interacting with sites imitating the same website

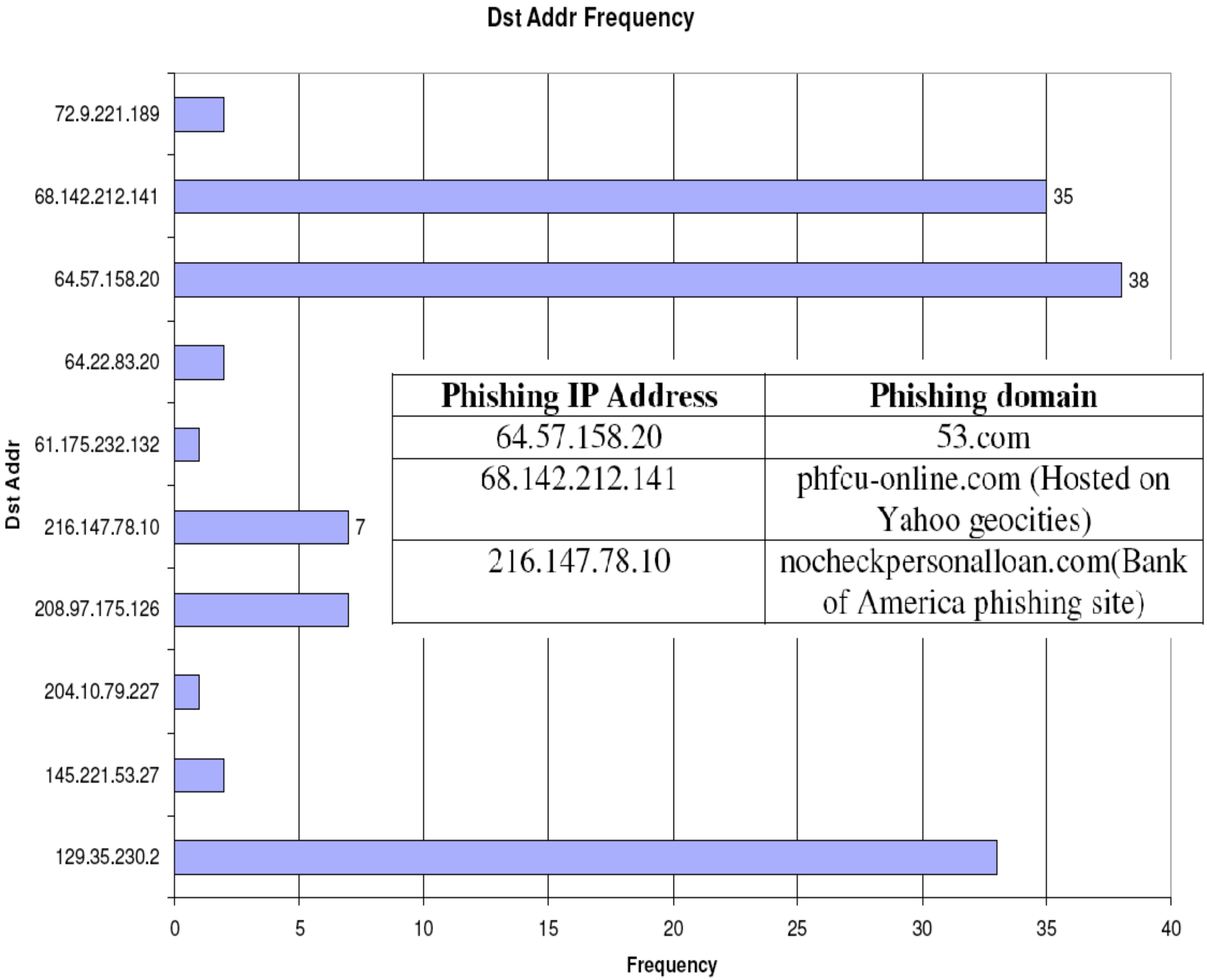


Georgia Institute of Technology

How many unique phishing sites did a source address visit ?

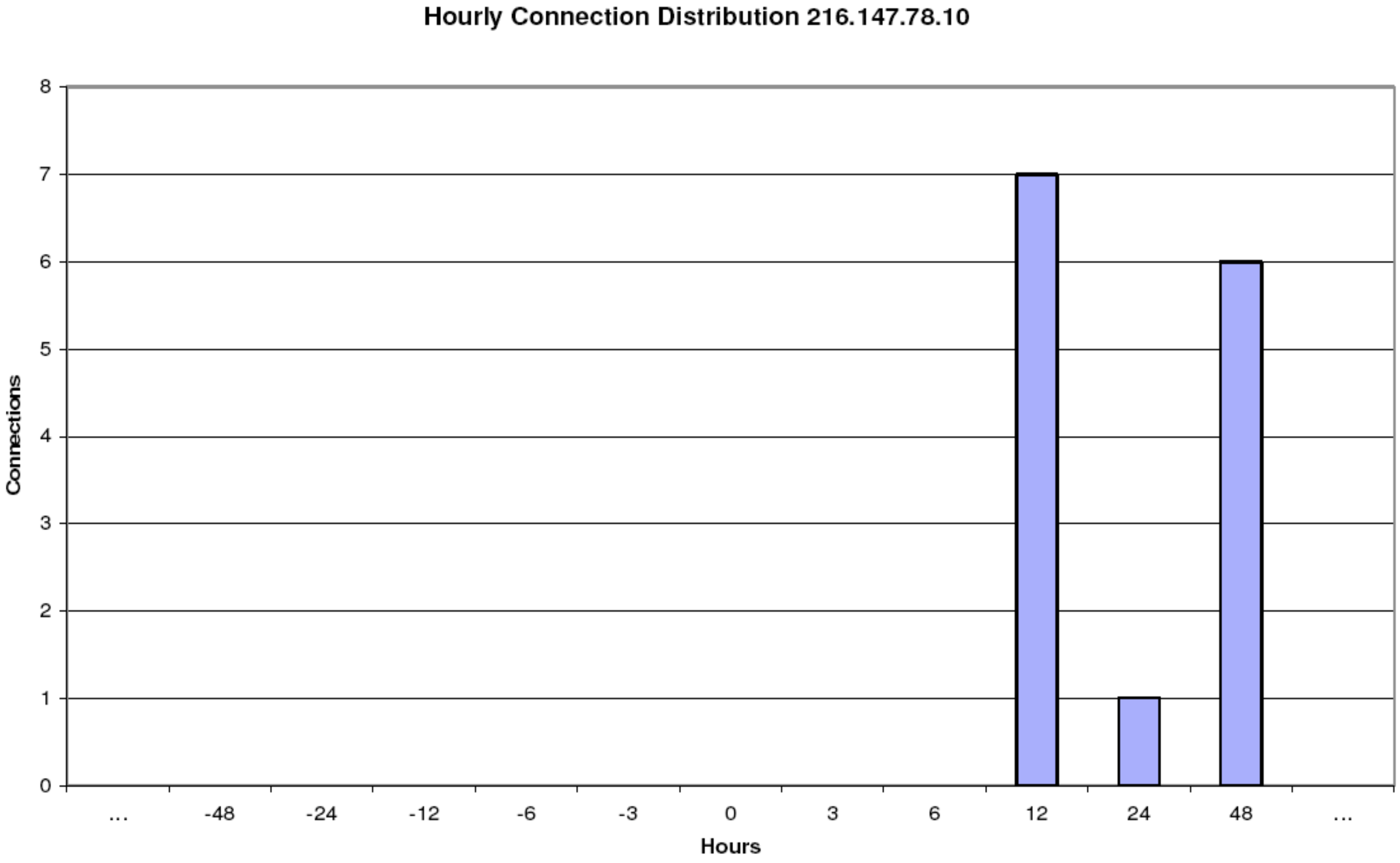


How many times a connection was made to a phishing site ?



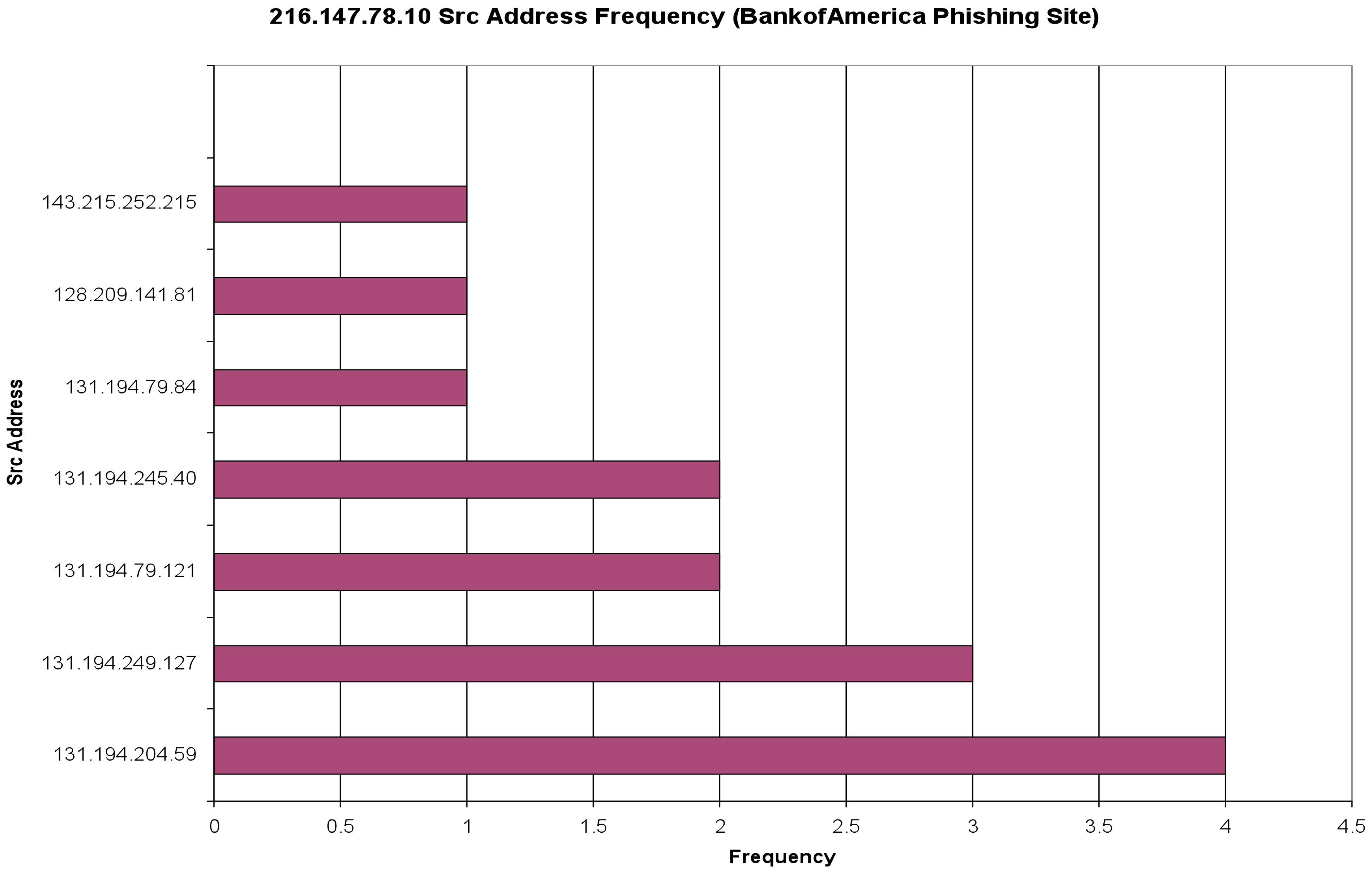
Phishing IP Address	Phishing domain
64.57.158.20	53.com
68.142.212.141	phfcu-online.com (Hosted on Yahoo geocities)
216.147.78.10	nocheckpersonalloan.com(Bank of America phishing site)

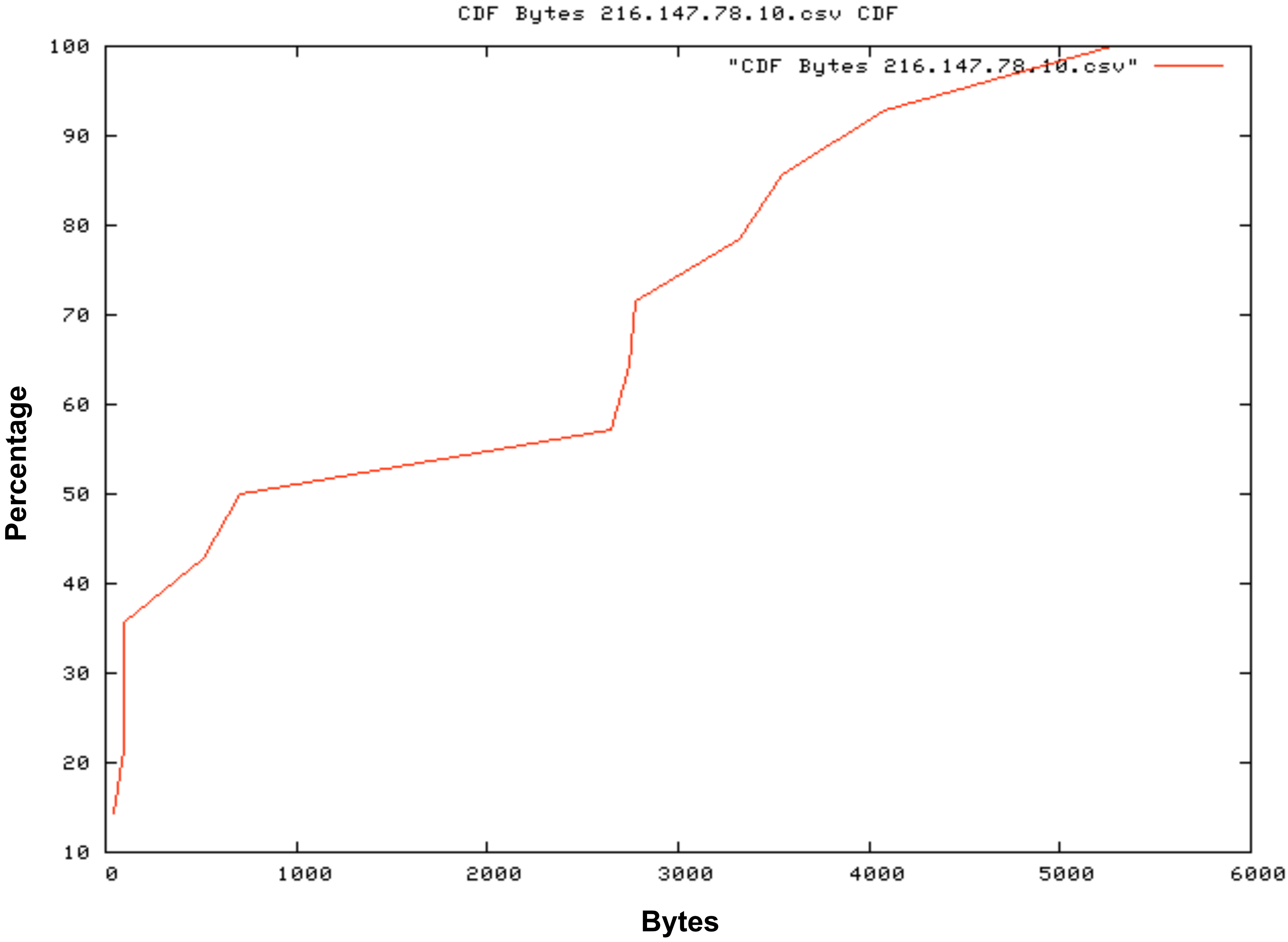
96 hour window around the receipt of Bank of America phishing email in the spam trap

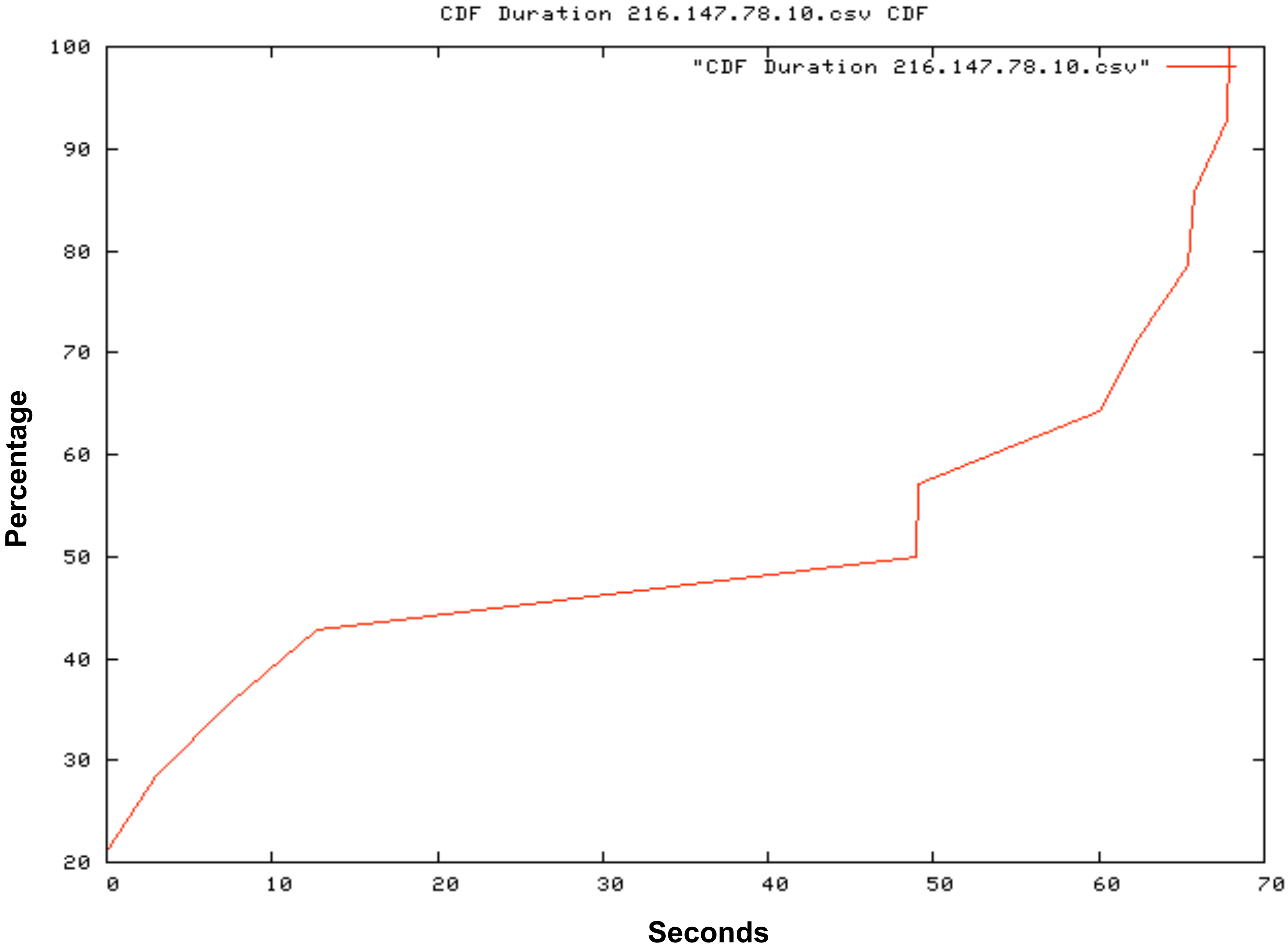


Connections made by diff src addresses to Bank of America phishing site –

Observations in line with “persistent connection behavior of browsers” by wang et al




CDF Bytes 216.147.78.10.csv CDF
"CDF Bytes 216.147.78.10.csv"
Percentage
100
90
80
70
60
50
40
30
20
10
0
1000
2000
3000
4000
5000
6000
Bytes



CDF Duration 216.147.78.10.csv CDF
"CDF Duration 216.147.78.10.csv"
Percentage
100
90
80
70
60
50
40
30
20
0
10
20
30
40
50
60
70
Seconds

Challenges while analyzing phishing at network level

- Lack of application layer context
- Not everybody sees the same set of spam/phishing emails
- Redirection Techniques
- Avg lifetime of a phishing site typically very small
- Timing differences
- Multiple Domain Hosting
- Other researchers on the same network

Recommendations and Future Work

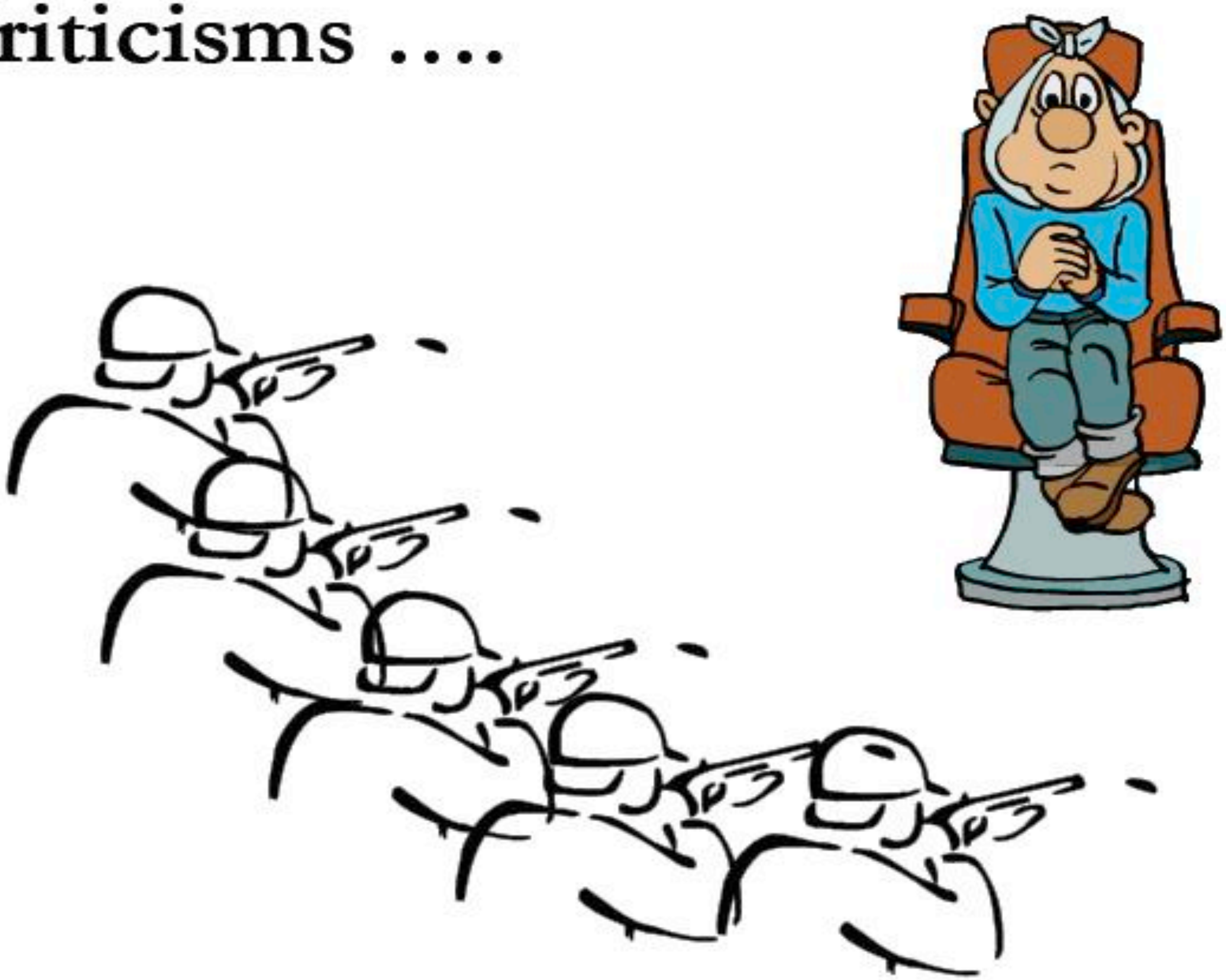
- Combined Data Sources
  - Application Level Sources
  - DNS Traces

- Multiple Vantage Points - Different Universities with Spam Traps
  - Can help address questions about -
    - Targeted Phishing
    - Percentage Phishing Mails per Spam Trap

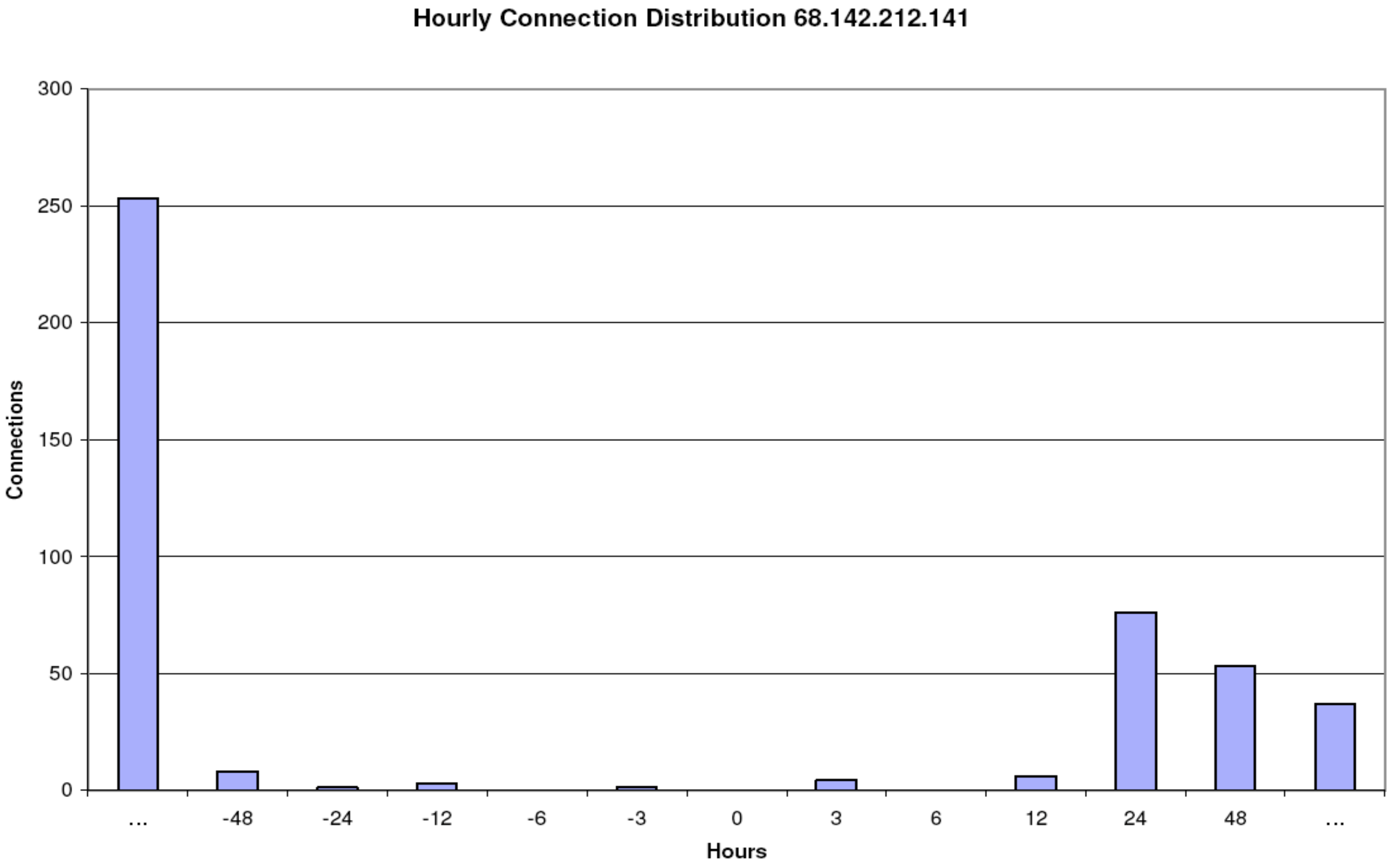
Acknowledgements

- "The logs and netflow traces used in this work were made available by the Georgia Tech Research Network Operations Center (www.rnoc.gatech.edu)

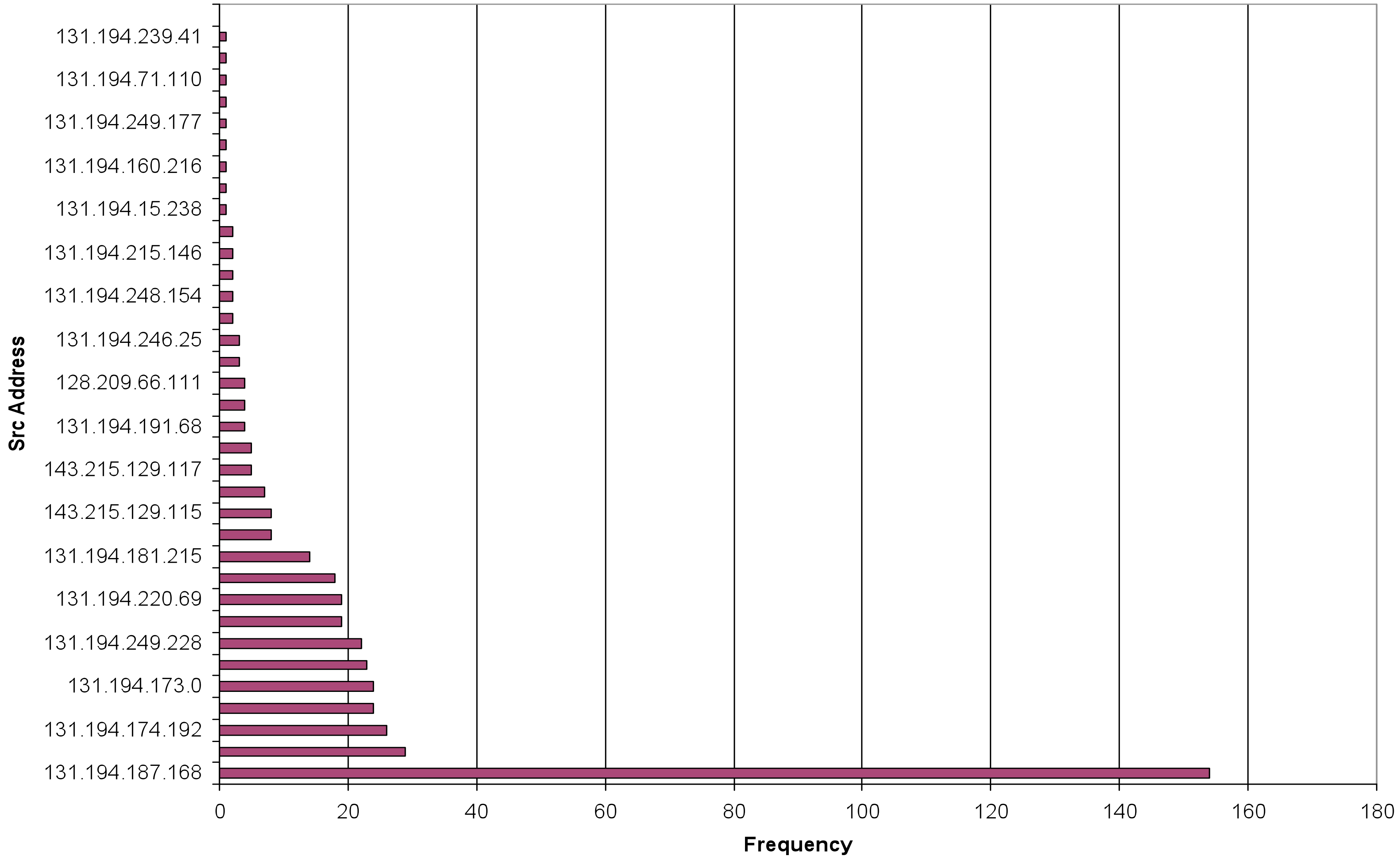
Kuestions, Komments, Kuriosities, Kriticisms ….

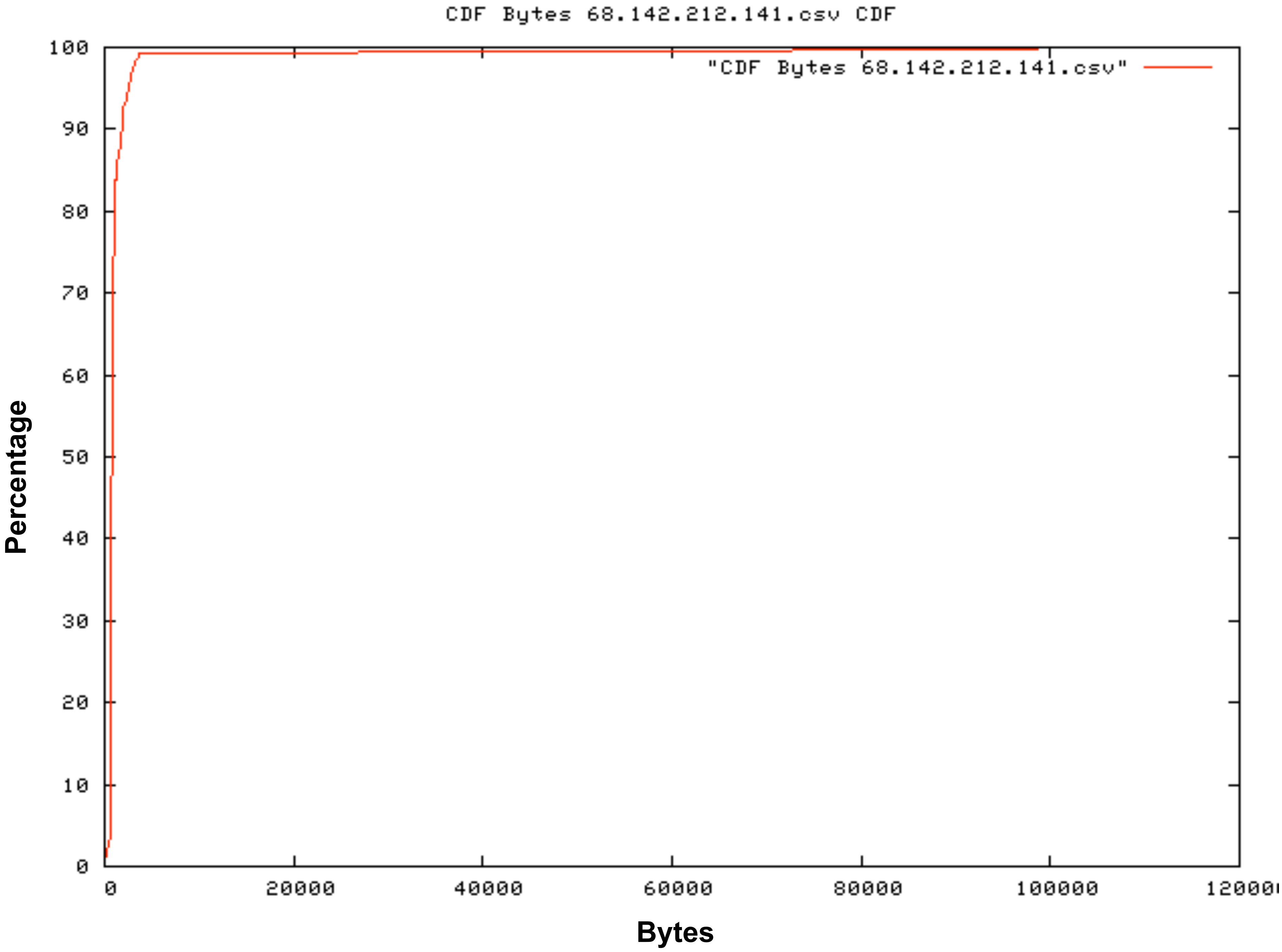


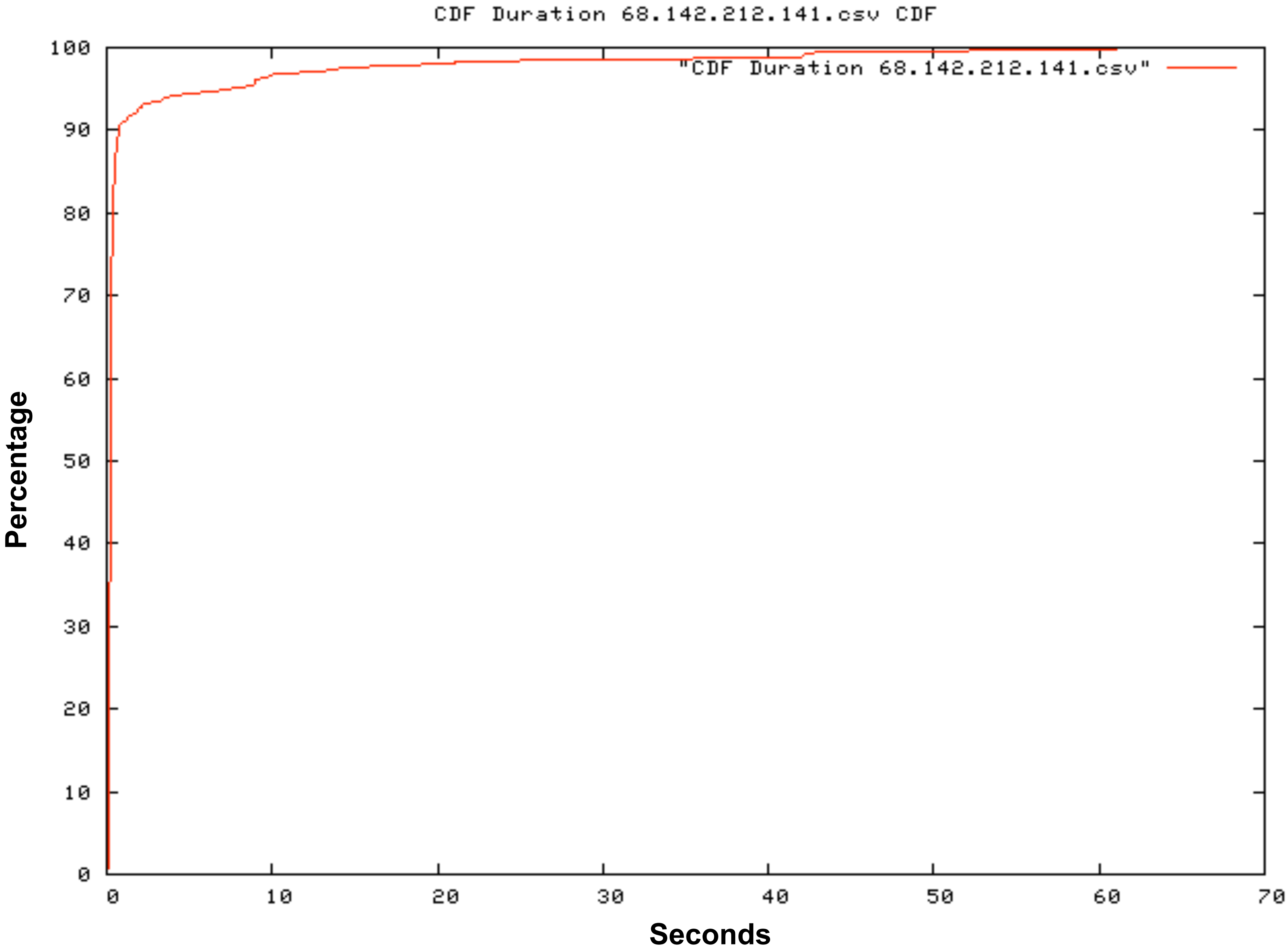
96 hour window around the receipt of phishing email about site hosted on yahoo geocities in the spam trap



68.142.212.141 Src Address Frequency (Hosted On Yahoo)




CDF Bytes 68.142.212.141.csv CDF
"CDF Bytes 68.142.212.141.csv"
Percentage
Bytes
0
10
20
30
40
50
60
70
80
90
100
0
20000
40000
60000
80000
100000
12000


CDF Duration 68.142.212.141.csv CDF
"CDF Duration 68.142.212.141.csv"
Percentage
Seconds
0
10
20
30
40
50
60
70
80
90
100
0
10
20
30
40
50
60
70