The Reputation of Networks – ARIN Region

Manish Karir, Kyle Creyts

(Merit Network Inc)

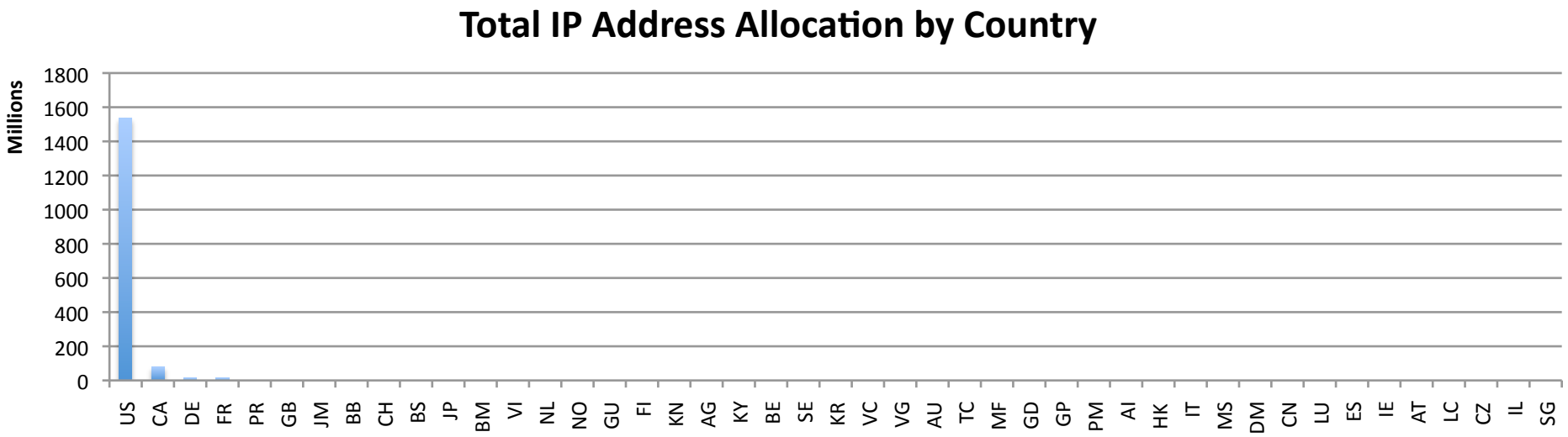
Outline

- Goal
- Background: IPv4 address allocation distribution in ARIN, commonly used blocklists
- Analysis
  - foreach(country, asn, bgp prefix)
    - SPAM Lists Distribution
    - Malware/Phishing Lists Distribution
    - Active Malicious Activity Lists
    - Highlight points of interest in data
- Network Reputation Discussion

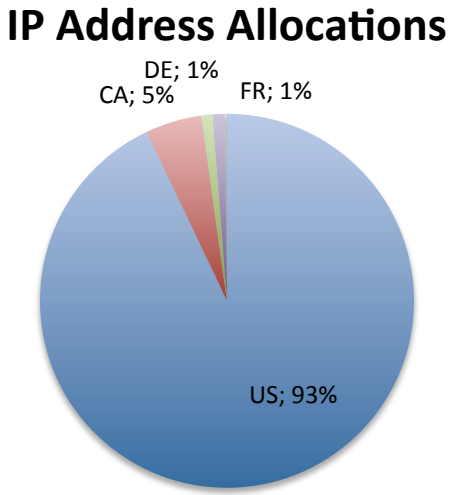
Common Reputation Block Lists (RBLs)

- RBLs are mostly lists of IP addresses of domains that have been observed to participate in suspicious behavior
- RBLs can be clustered by type of activity on which it is based:
  - SPAM Lists: SPAMHAUS(CBL), BRBL, SpamCop, wpbl, UCEPROTECT
  - Malware/Phishing hostsing: SURBL (multi), phishtank, hpHosts
  - Active Attack Behavior: Darknet Scanner (merit), Dshield, ssh brute-force (fail2ban, denyhosts)
- Our goal is to analyze relative distribution of hosts on these lists to determine if there are some common traits that can broadly characterize the observed relative malicious activity originating from a country, ASN, and prefix

ARIN Address Space Distribution by Country



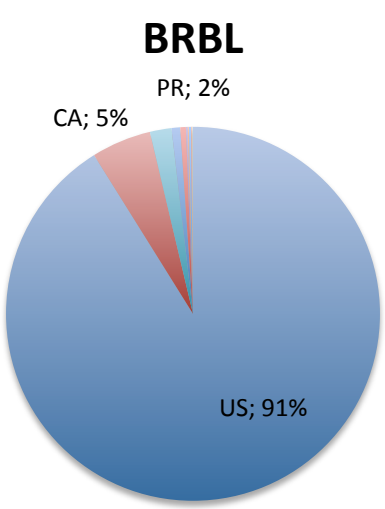
- Roughly 6.4M/24 blocks allocated ~ 1.65B IP addresses
- US accounts for 93% of all IP address allocations, Canada is 5%

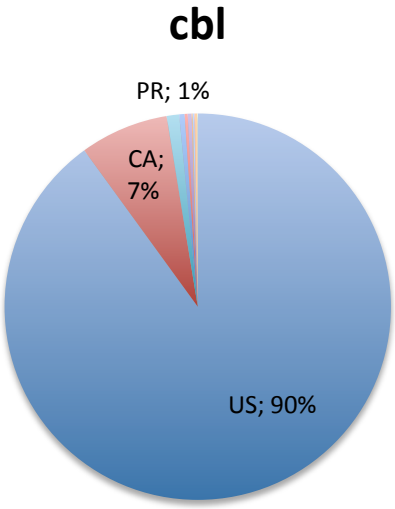


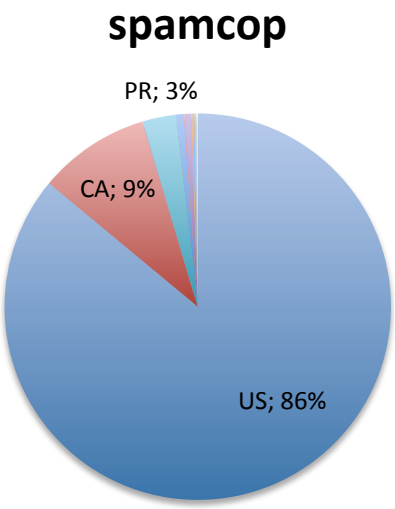
SPAM Lists Distribution Analysis

- Consider 3 largest/most popular SPAM Lists:
  - Barracuda BRBL
  - SPAMHAUS – CBL
  - SpamCop
  - Other SPAM data sources as well such as weighted private block list (wpbl), UCEPROTECT also analyzed but omitted here due to similarity
- Determine portions of those lists relevant to the ARIN region
- Determine relative distribution by country within ARIN region

SPAM Lists Distribution by Country





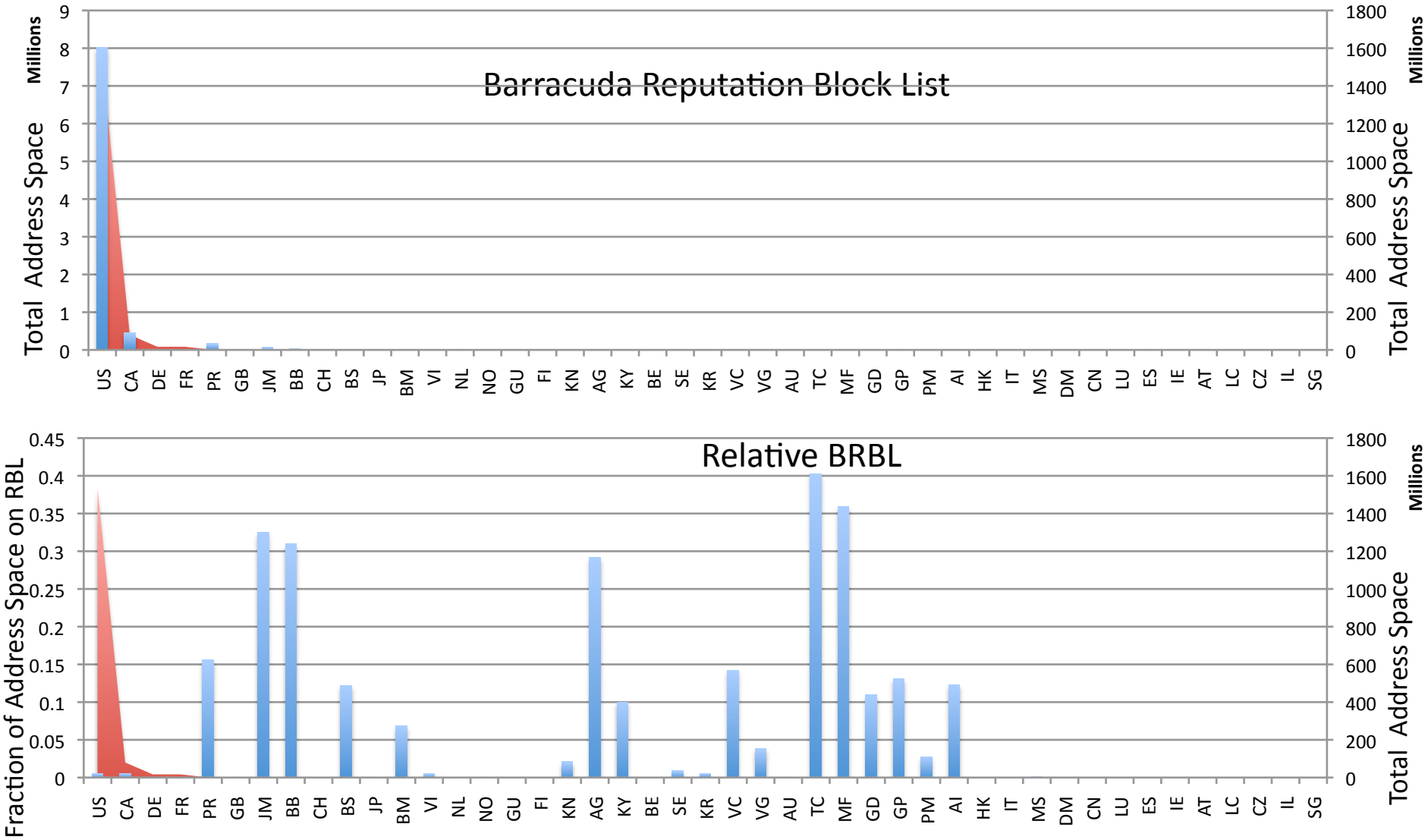


List	Total IPs	ARIN IPs	LACNIC IPs	RIPE IPs
Barracuda	128M	8.8M (6.8%)	22.7M (17%)	65M (51%)
SPAMHAUS CBL	8.1M	122K (1.5%)	1M (12%)	2.6M (32%)
SpamCop	325K	3.2K (1%)	28K (8%)	66K (20%)

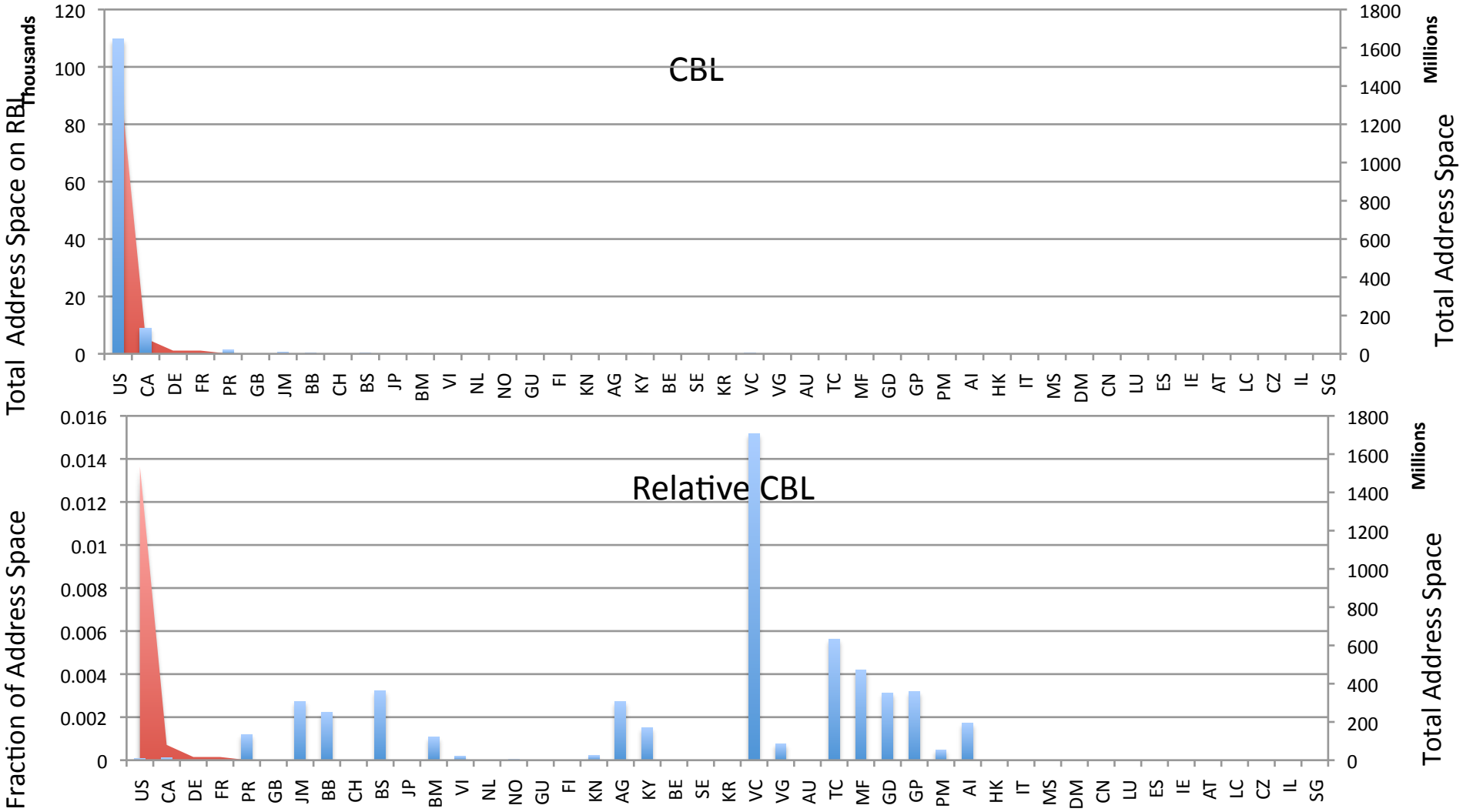
SPAM List Relative Distribution

- In general: larger allocations/blocks have more entries in block lists – expected if you assume infection rates are a steady fact of life and on average x% of any given IP address range will be on a block list
- But what happens when we look at block list entries relative to allocation sizes
- We should look at both the large and the small ends of allocation spectrum

Relative SPAM List Distribution by Country



Relative SPAM List Distribution by Country



SPAM List Discussion

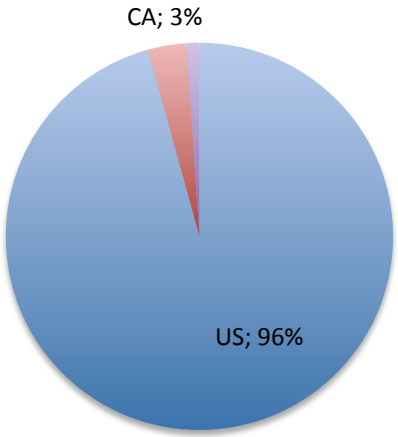
- All networks are not created equal when it comes to entries on a SPAM list
- In general ARIN region has much lower percentage of listings on SPAM RBLs
- Interesting things to notice:
  - US has 8M IPs on BRBL and 100K on CBL
  - Almost 15% of Puerto Rico is on BRBL
  - Most Caribbean islands have significantly higher percentage of IP address space on BRBL or CBL
- What accounts for these regional variations? Local policy? Connectivity? Network topology?

Malware/Phishing Lists Distribution Analysis

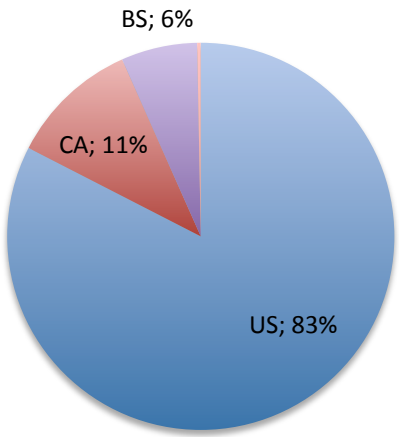
- Consider 3 common malware/phishing Lists:
  - SURBL
  - hpHosts
  - phishtank
  - Other popular data sources as well such as malwaredomains and malwaredomainsList are included in the SURBL-multi dataset.
- Determine portions of those lists relevant to the ARIN region
- Determine relative country distribution within ARIN region

Malware/Phishing Lists by Country

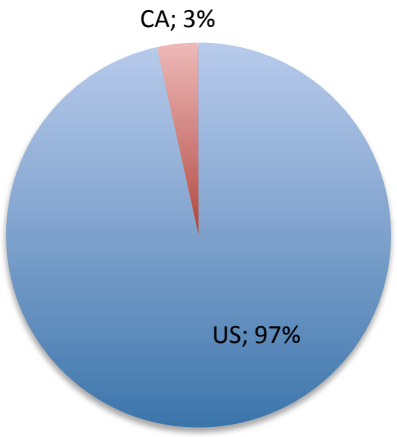
surbl



hphosts



phishtank



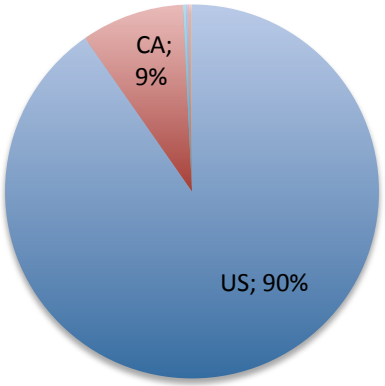
List	Total IPs	ARIN IPs	LACNIC IPs	RIPE IPs
SURBL	360K	194K (54%)	3K (<1%)	107K (30%)
Hphosts	185K	94K (51%)	2K (<2%)	71K (38%)
Phishtank	4700	2627 (56%)	124 (< 3%)	1700 (36%)

Malware/Phishing Discussion

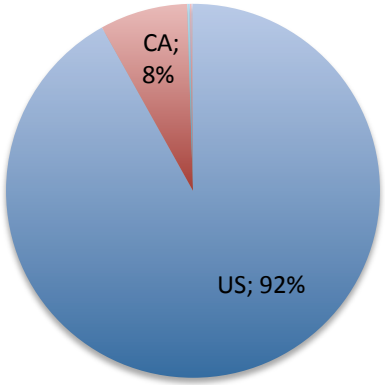
- In general, ARIN region activity on malware/ phishing lists is uncharacteristically high as a percentage of total listings as compared with RIPE and LACNIC
- US accounts for over 80% of all domains on the various malware/phishing list for the ARIN region.
- Bahamas is 6% of ARIN region entries on hphosts list

Active Malicious Activity by Country

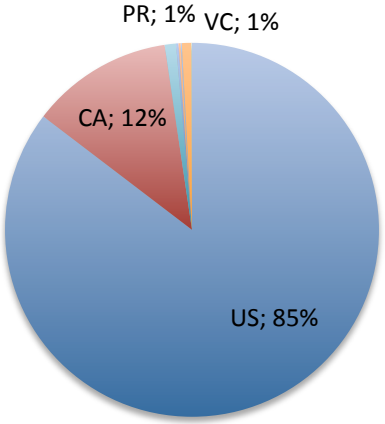
dshield



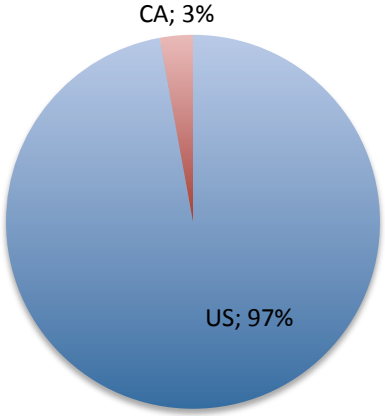
ssh brute-force



Darknet Scanning



zeus



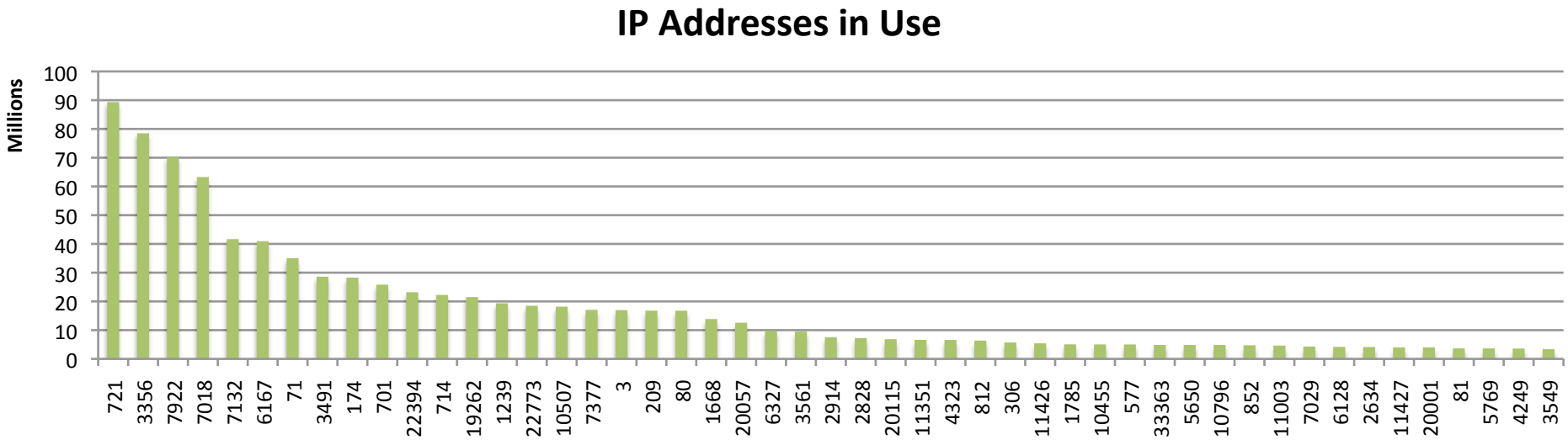
Active Malicious Global Comparison

List	Total IPs	ARIN IPs
ssh brute-force	68K	11K (16%)
Dshield	754K	128K (17%)
Darknet Scanning	156K	7.8K (5%)
Zeus	215	35 (16%)

List	Total IPs	LACNIC IPs
ssh brute-force	68K	11.6K (17%)
Dshield	754K	61K (8%)
Darknet Scanning	156K	28K (17%)
Zeus	215	1 (0%)

List	Total IPs	RIPE IPs
ssh brute-force	68K	22K (32%)
Dshield	754K	314K (42%)
Darknet Scanning	156K	83K (53%)
Zeus	215	161 (75%)

Address Distribution by ASN

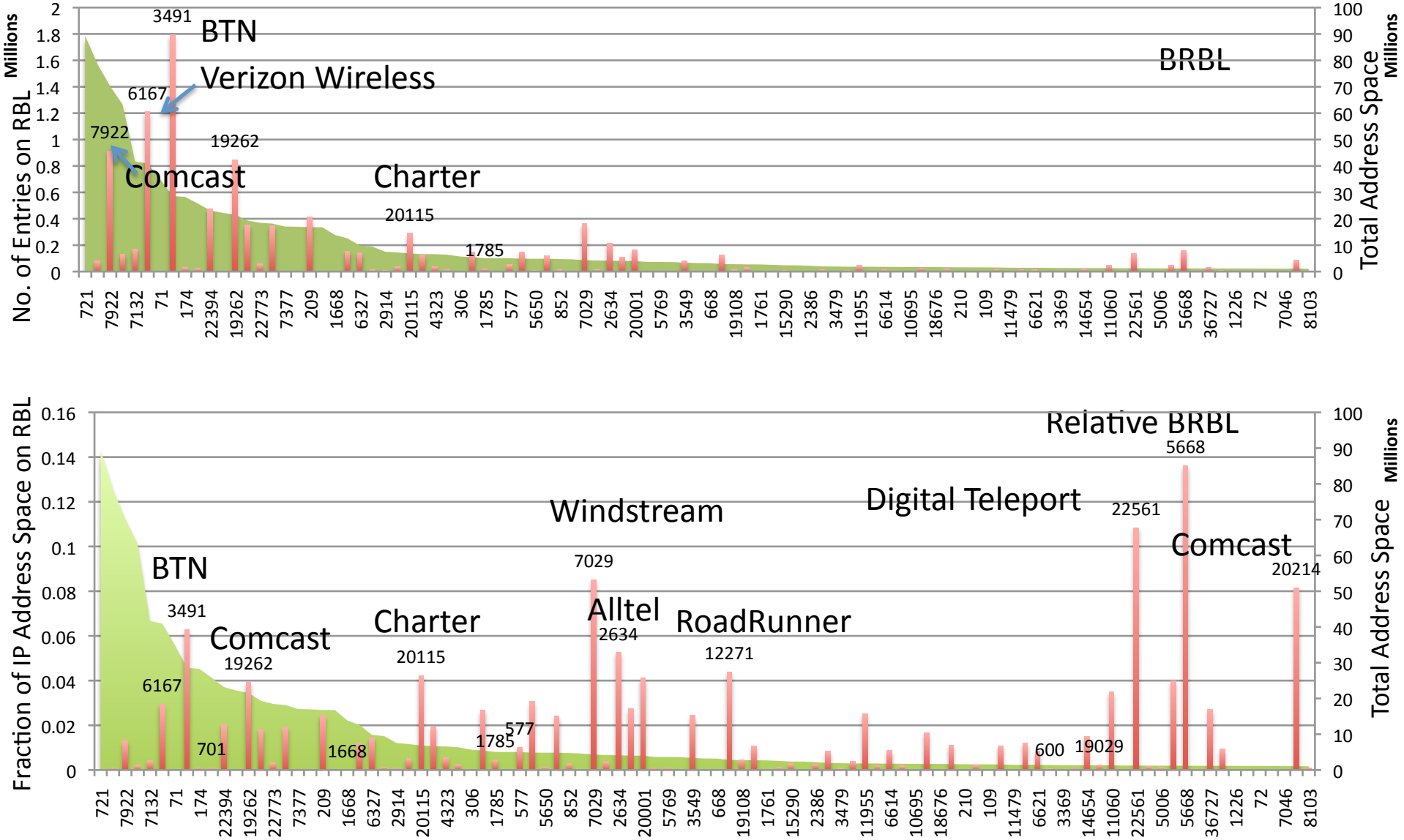


- Roughly 21K ASNs in use in ARIN region
- They account for roughly 140K of prefixes in the BGP routing table (total 370K entries)
- A total of 1.2B IPs
- We focus on the largest 50 ASNs

Top 10 ASNs by Size

ASN	Name	IP Addresses
721	DNIC	12M (9%)
3356	LEVEL3	12M (9%)
7922	Comcast	7M (5.3%)
7018	ATT	6M (4.6%)
7132	SBIS-ATT	4.8M (3.7%)
6167	Verizon Wireless	3.7M (2.8%)
71	HP	3.3M (2.5%)
3491	BTN	3.2M (2.4%)
174	Cogent	2.8M (2.1%)
701	MCI/Verizon Business	2.6M (2%)

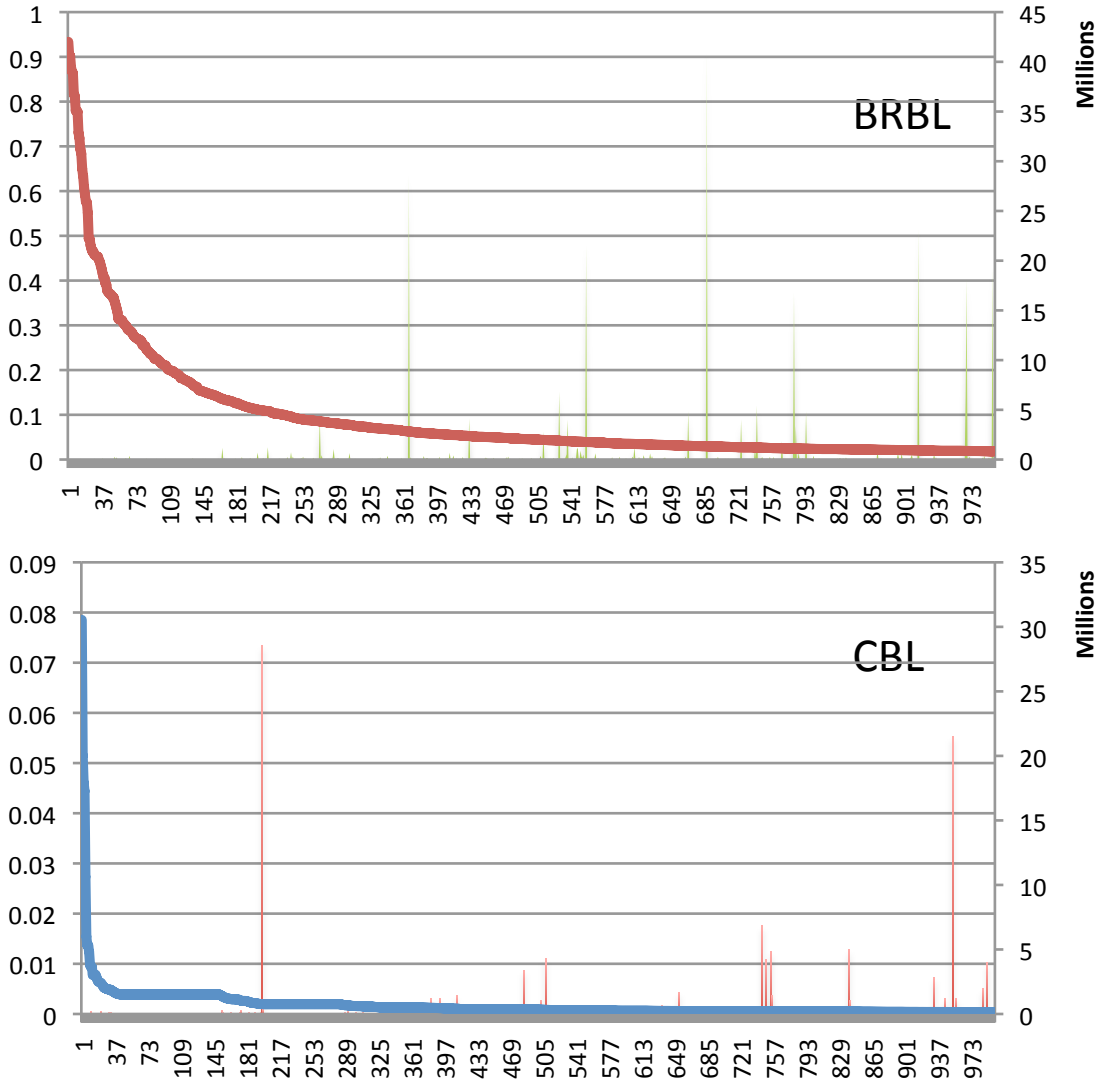
SPAM List IP Distribution by ASN



SPAM List IP Address Distribution by ASN Discussion

- AS 7922 - Comcast has almost 900K IPs on BRBL, AS3491 – BTN has highest count 1.8M but only 6% of its total address space, much lower fraction for Comcast
- 5 of the top 50 largest ASNs have more than 5% of their address space on BRBL
- Absolute numbers are lower for other lists but general trends are similar
- Important to not only pay attention to large networks but also networks with large fractions

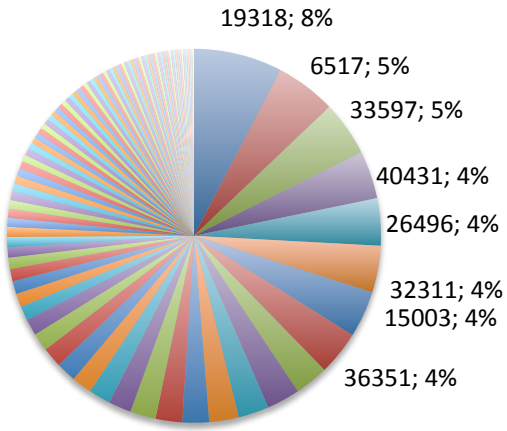
ASN IP Blocklisting Distribution



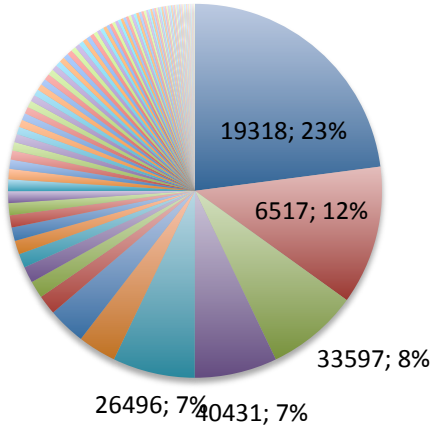
- Top 1000 ASNs with largest percentage of their networks on SPAM blocklists
- Almost 200 ASNs have atleast 10% of their IPs on BRBL
- Less than 10 ASNs have atleast 1% of their IPs on CBL

Malware/Phishing Domains Distribution by ASN

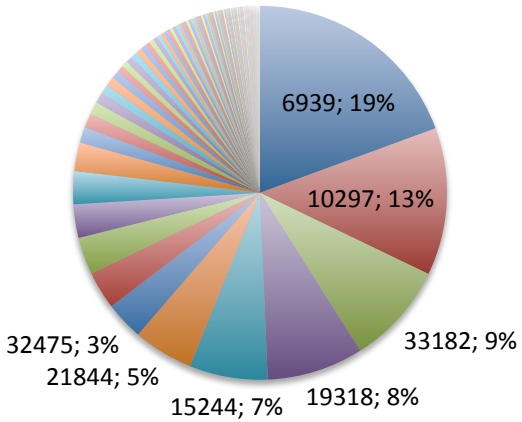
surbl



hphosts



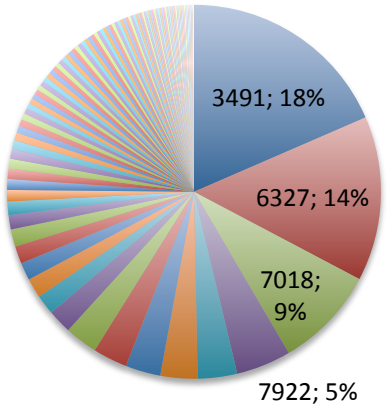
phishtank



- Top 5 ASNs account for almost 50% of entries on lists, SURBL shows greater distribution of entries across ASNs
- AS 19318 – NJIIX is almost 23% of entries on hpHosts list and 8% of SURBL
- AS6517 – Reliance Globalcom represents 5% of SURBL entries, and 12% of hpHosts
- AS6939 – HE is almost 20% of entries on phishtank list, AS10297- eNET, and AS33182 – HostDime account for another 21%

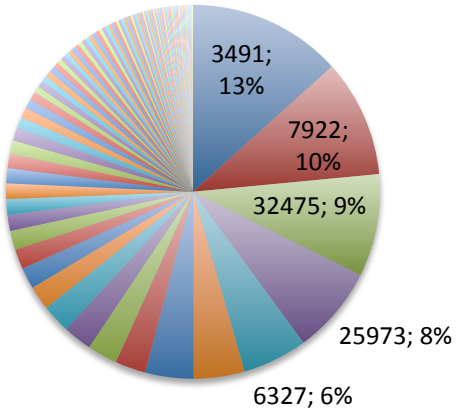
Active Malicious Activity by ASN

Darknet Scanning

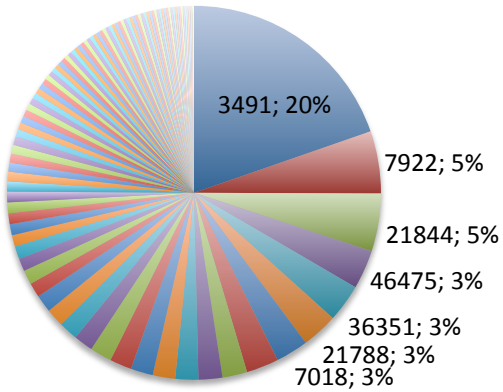


- Top 5 ASNs account for almost 50% of entries
- ssh brute-force shows more even distribution with only 1 ASN accounting for 20%

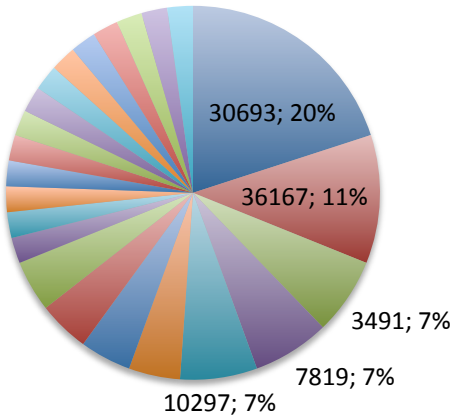
dshield



ssh brute-force



zeus

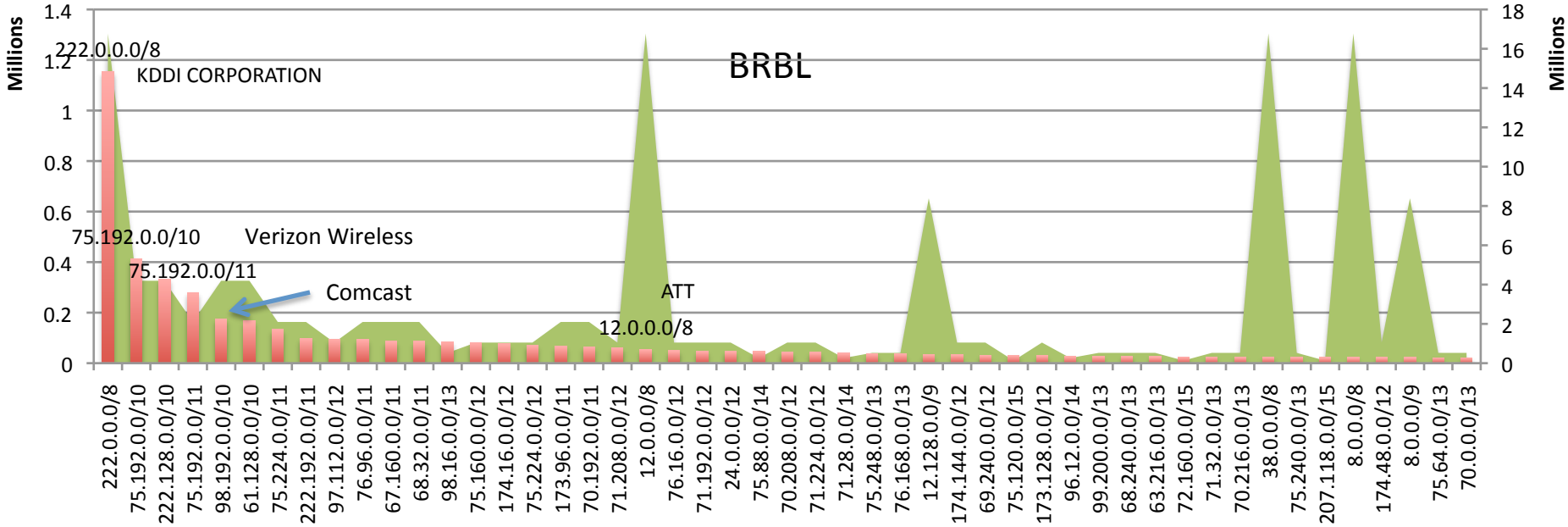


Active Malicious Activity Discussion

List	Total IPs	ARIN IPs
ssh brute-force	68K	11K (16%)
Dshield	754K	128K (17%)
Darknet Scanning	156K	7.8K (5%)
Zeus	215	35 (16%)

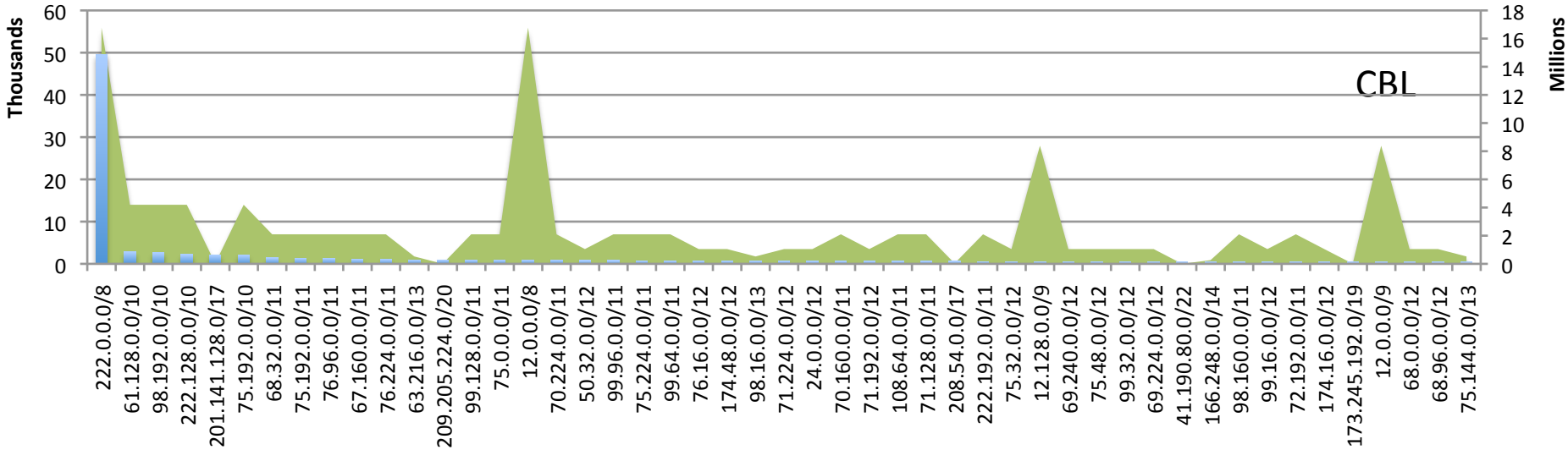
- AS3491 – BTN is 18% of observed scanning activity from ARIN region and 20% of ssh brute-force attempts, 7% of zeus list
- AS7922 – Comcast is 10% of Dshield activity and 5% of ssh brute-force attempts
- AS30693 – Eonix is 20% of Zeus entries for ARIN region – (but small number)
- AS6327 – Shaw represents 14% of scanning activity

BGP Prefix SPAM List IP Distribution



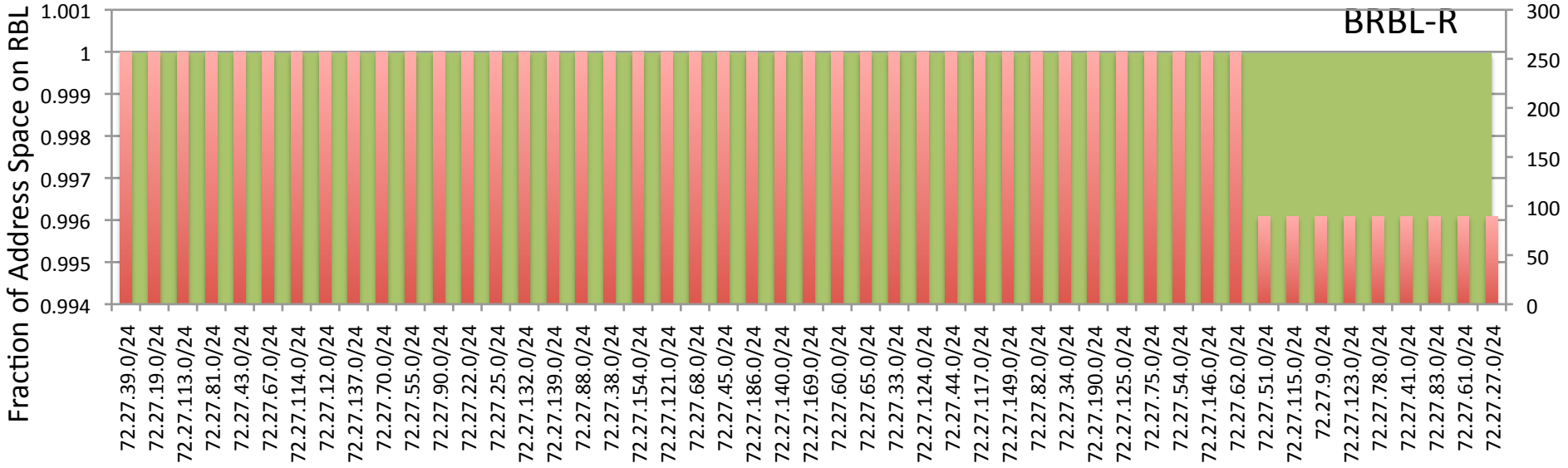
- BGP ARIN region prefixes 138K out of total routing table of ~370K
- No surprise that some large prefixes have large numbers of IPs in BRBL but not all
- BUT – only 1 prefix (in largest 50 prefixes) has over 1M IPs in the BRBL
- Top 5 prefixes have 200K or more IPs each on BRBL

BGP Prefix SPAM List IP Distribution



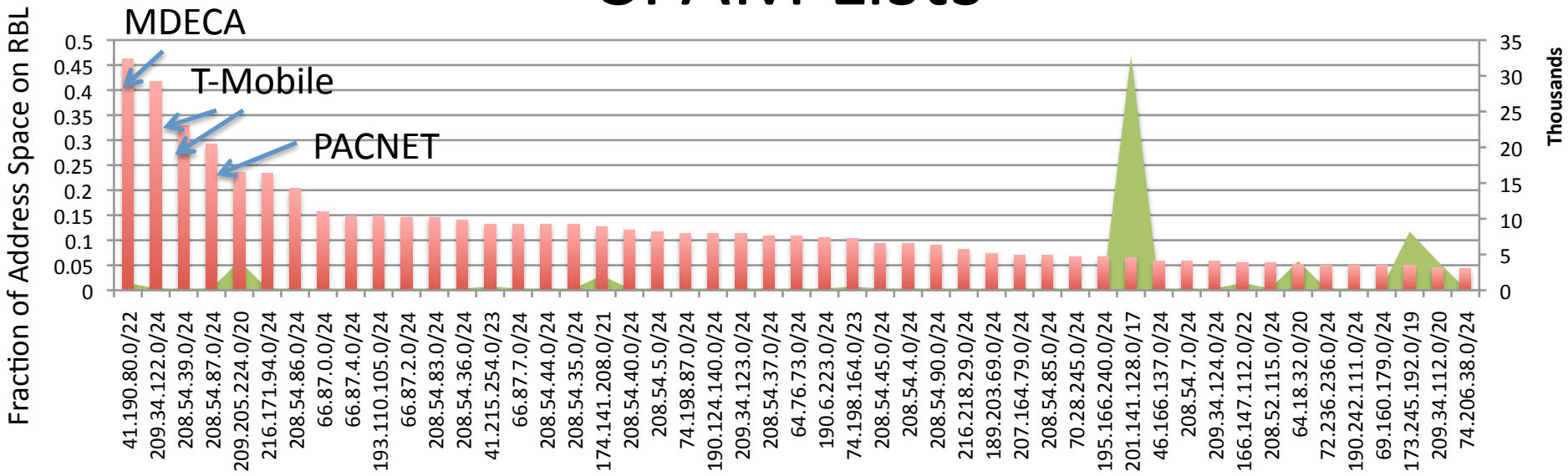
- Only 10 of the prefixes shown above have more than 1K or more IPs listed
- 222/8 – KDDI appears to be outlier with almost 50K entries on the CBL

Relative Amounts of IP addresses in SPAM lists



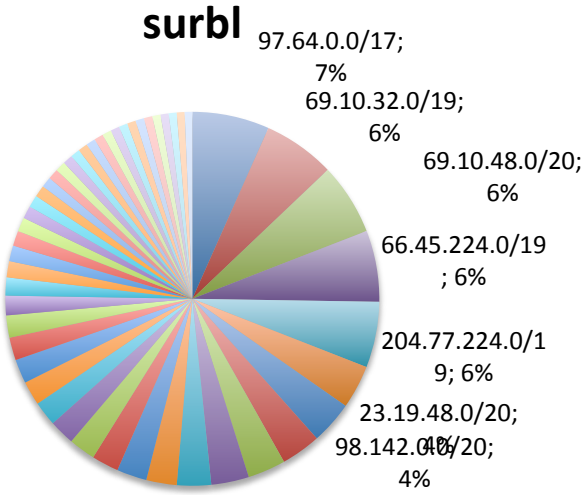
- 1300 ARIN region prefixes have over 50% of their address space included in the BRBL mostly small prefixes
- Over 2500 prefixes out of all ARIN region prefixes have more than 25% of their IP address block listed in the BRBL
- Cable and Wireless Jamaica appears to be heavily listed

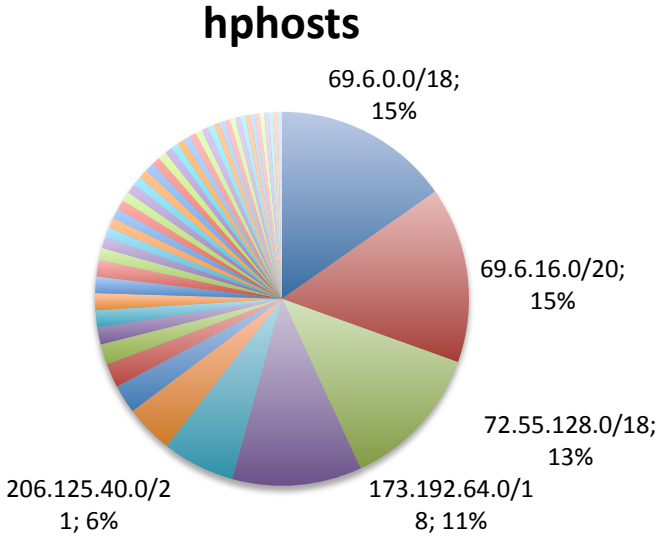
Relative Amounts of IP Address in SPAM Lists



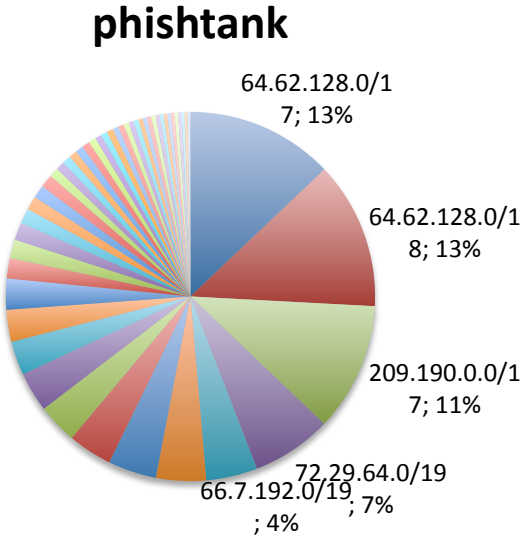
- 25 prefixes have atleast 10% of their IPs listed in CBL mostly /24s
- 209.34.122.0/24 - MDECA has over 40% of its space on CBL
- 209.205.224.0/20 – PACNET has 22% of its space on CBL
- All 50 prefixes shown have 5% or more of their space on CBL

Malware/Phishing IP Address Distribution



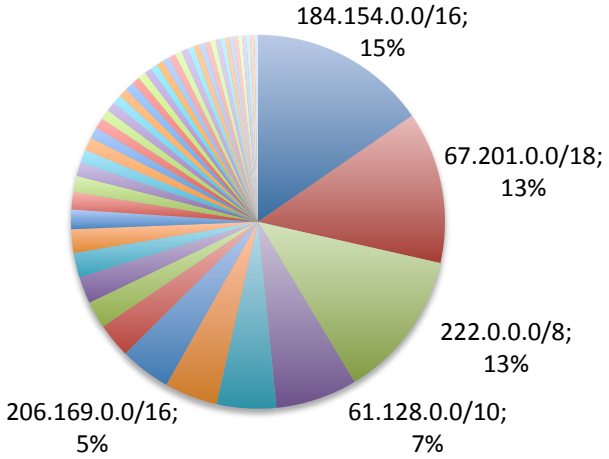


- Relative percentages of IPs for the top 50 prefixes for each data type are shown above
- 69.10.32.0/19– Interserver accounts for 6% of SURBL entries
- 72.55.128.0/18 – iWeb Technologies and 173.192.64.0/18 – SoftLayer Technologies together account for almost 25% of hpHosts entries
- 64.62.128.0/18 – Hurricane Electric is 13% of phishtank entries, 209.190.0.0/17 – eNET is 11% and 72.29.64.0/19 – HostDime is 7%

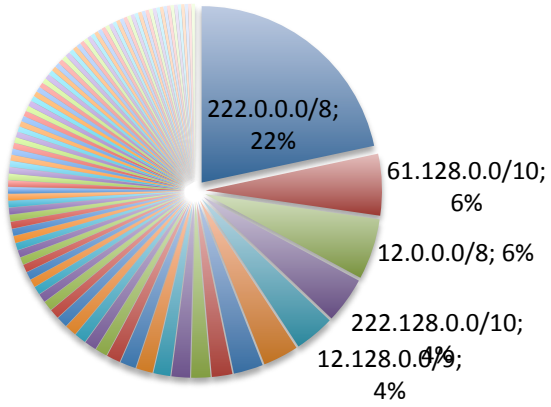


Active Malicious Activity List IP Distribution

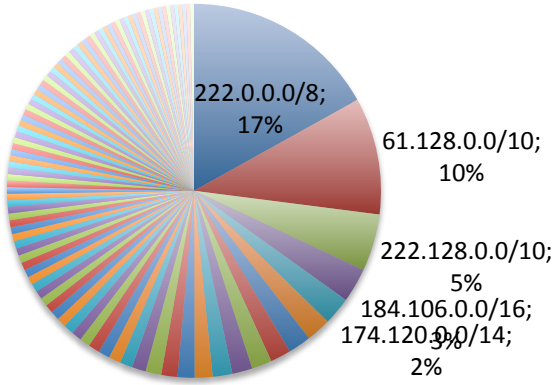
dshield



Darknet Scanners



ssh brute-force



- 184.154.0.0/16 – SingleHop is 15% of Dshield activity, 67.201.0.0/18 – PacketExchange is another 13%
- 222.0.0.0/8 – KDDI accounts for 22% of Darknet Scanning activity and 17% of ssh brute-force listings
- Some prefixes are APNIC region allocations being announced by ARIN region ASNs

Discussion

- Network reputation is an attempt to construct a metric or set of metrics that illustrate the collective reputation of all hosts in your administrative domain
- While infected hosts and botnets are a fact of life, how much of such activity represents an acceptable level of network pollution 1%? 10% of all hosts?
- Hosts that engage in malicious activity such as spam, phishing, malware, scanning in a network reduce the externally visible global network reputation of that network – it does not go un-noticed
- It can be seen that not all networks are equal when it comes to network reputation. What policies, topology, connectivity, other factors make some networks better than others? How can we learn from them?
- Reputation of hosts on your network has an impact on the usability of your network as portions might get blocked for various services

Using Network Reputation

- Network reputation is not just something other people know about you
- You can use it to craft flexible local policies that can better manage your risk profile
- Variable services can be offered to networks with different reputations
- You can control how much of your network and what services on your network are visible to networks with varying reputation levels
- Reputation information can even be a factor in BGP path selection algorithm

Network Reputation

- Our goal is to develop a comprehensive global network reputation system that computes for each prefix in the BGP routing table a reputation metric.
- Variations can allow arbitrary network boundaries not simply BGP boundaries but that is the starting point
- Data from common sources such as RBLs is the starting point for bootstrapping the reputation system, however in order to be successful the system must have data from many many vantage points
- Different networks have different views of reputations of other networks
- The more vantage points you have the closer to "true reputation you will get"
- The system must allow all networks to participate and contribute reputation information regarding all other networks while being resistant to collusion and false reporting
- Current project at Merit Network Inc is building such a system and an effort will soon be made to recruit participant networks on various mailing lists
- If you would like to participate please send email to: mkarir@merit.edu
- How reputable is your network?