Finding the Needle in the Haystack

Jonzy
Data Security Analysis, Sr.

Information Security Office

Finding the Needle in the Haystack

With all the information available via NetFlows, finding the "Needle in the Haystack" (the bad actor in NetFlows), can be somewhat difficult at best. Methods to discover illegitimate traffic can be as simple as looking at TCP flags, to more complex procedures such as defining thresholds for number of flows with ratios to unique destinations. There are other methods available, but I will be focusing on these thresholds and ratios and why this approach turns the needle into a goal post. The CPU cycles needed for this analysis are reduced by implementation of AVL trees (Balanced Binary Trees), and knowing the bottleneck to process the data is based on reading the data from disc. The algorithm used takes less then a second to process 3 million flows collected over a 5 minute time span. Both inbound and outbound, as well as local, traffic needs to be considered. Inbound analysis will help protect against external threats, outbound traffic protects yourself from external embarrassment, and local analysis identifies local problems that can lead to bigger problems.

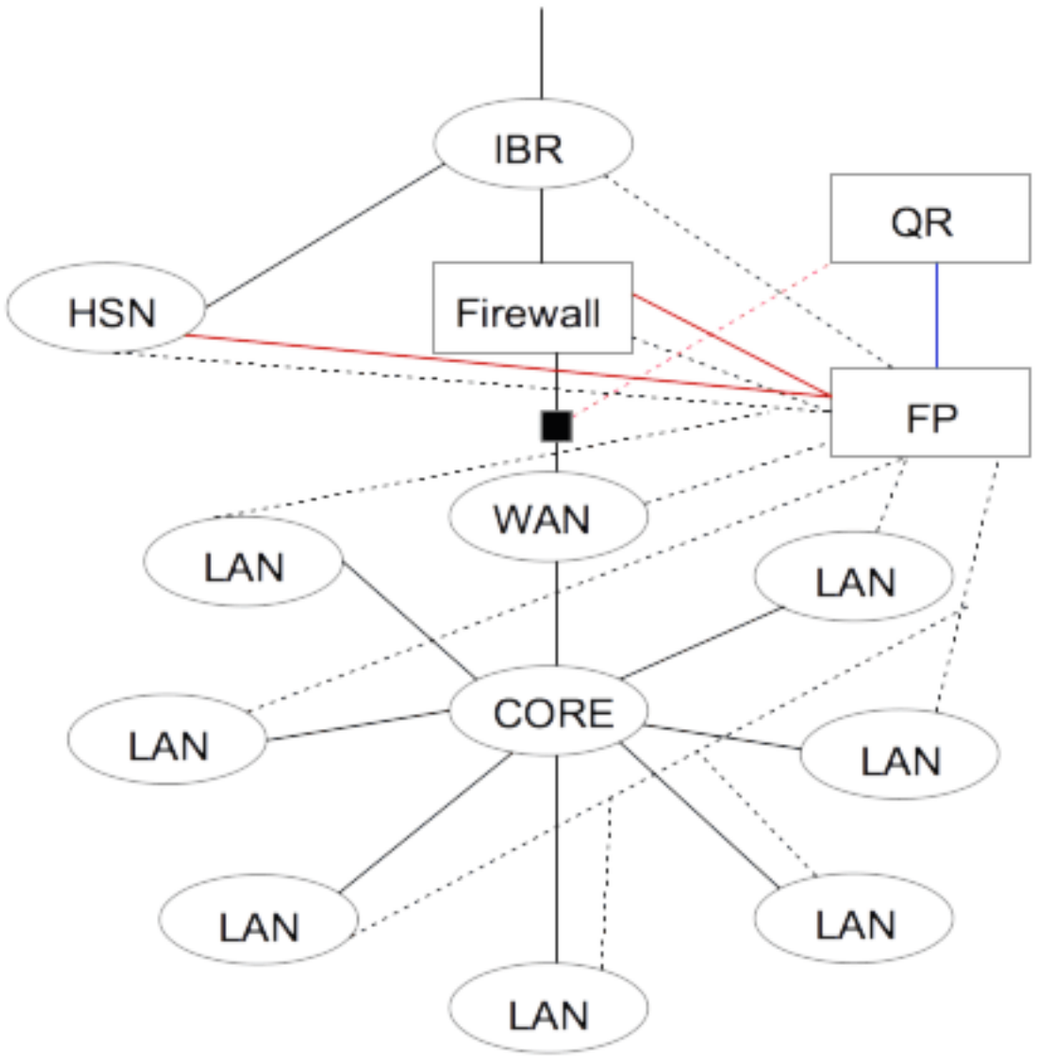
THE UNIVERSITY OF UTAH

Information Security Office

Network Layout / Flow Collection

IBR - 2 routers, with a 100 Gb/s channel to the Net
WAN - 2 routers, with a 40 Gb/s commodity network
LAN - 28 routers, with a 40 Gb/s internal network
HSN - 1 router, with a 100 Gb/s channel to the Net
FP - Flow Processor

- Null-route / Blockage
- Netflow Collection
- QR to FP link
- QR Tap



THE UNIVERSITY OF UTAH

Information Security Office

Flow Collection Hardware and Stats

The Collector

HP ProLiant DL380p Gen8
Processor: 2x Intel(R) Xeon(R) CPU E5-2640 0 @ 2.50GHz
6/6 cores; 12 threads
64-bit Capable
Memory: 98 GB DDR3 1333 MHz RAM
Storage: 12x HP 600GB 15K RPM 6GBs SAS Drives configured RAID 5
NIC: 3x 1Gbs copper NIC connected full duplex
Average Load: less then 1.5, but has been as high as 22.

Flow Collection Statistics

COLLECTOR	AVERAGE/DAY NUM_FLOW_RECORDS	AVERAGE_TIME TO_PROCESS_24_HOURS
IBR	719,521,466	13 seconds
WAN	711,442,717	12 seconds
LAN	1,945,181,346	32 seconds
HSN	14,065,862	less then a second

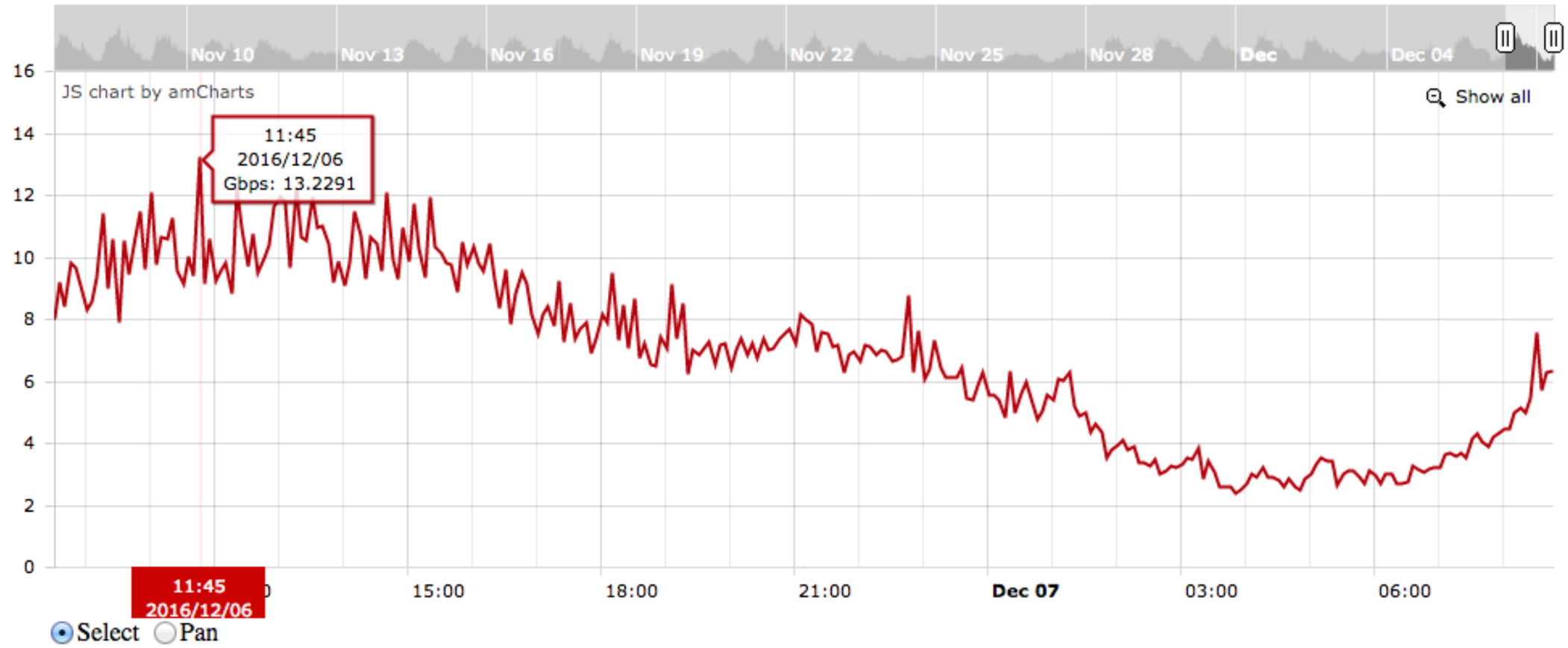
THE UNIVERSITY OF UTAH

Information Security Office

Average Gb/s per 5 minutes

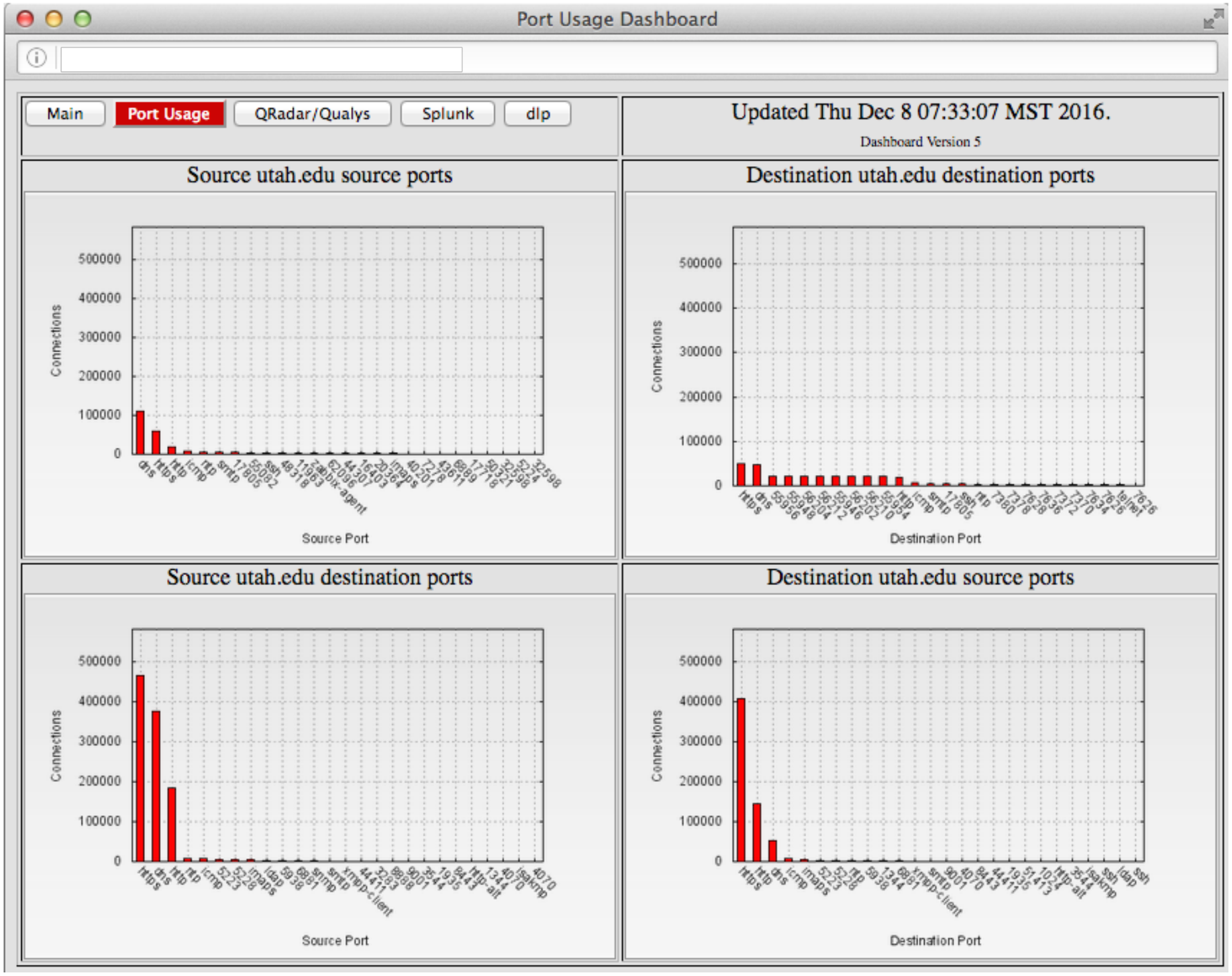
Average Gb/s per 5 minutes

Go Back



THE UNIVERSITY OF UTAH

Information Security Office



THE UNIVERSITY OF UTAH

Information Security Office

Destination Port Traffic

20161105	
53	24,215,863
443	15,243,092
80	9,355,812
55956	5,646,672
55948	5,638,853
56204	5,618,038
56212	5,614,816
55954	5,464,568
55946	5,441,662
56210	5,436,780
56202	5,433,117
27015	4,718,038
0	3,219,192
23	1,931,961
123	1,837,832
25	1,199,969
11963	1,036,750
22	732,815
6881	654,799
10050	644,902
8080	633,549
6889	598,315
7634	453,766
7380	435,086
7370	434,959
7636	434,537
7372	433,509
7628	433,333
7626	432,128
7378	425,950

20161023	
2323	48,782,842
53	22,773,508
443	16,836,694
80	6,184,346
0	3,435,316
35962	2,545,436
123	1,851,134
23	1,286,816
25	1,241,177
22	907,131
11963	719,472
10050	645,745
55956	634,299
55948	633,488
56204	631,845
56212	628,498
56202	619,404
55946	618,566
55954	618,278
56210	611,344
6889	551,449
6881	494,017
19709	478,056
52126	433,639
993	390,159
40201	374,789
55082	363,053
17718	294,917
8000	216,457
8080	210,625

20161126	
7547	63,109,855
53	17,999,985
443	14,574,230
80	6,396,821
55948	5,138,553
56204	5,124,958
55956	5,118,086
56212	5,094,511
55946	4,948,344
55954	4,940,829
56210	4,926,110
56202	4,920,184
26915	3,787,926
0	3,017,123
123	1,303,023
6881	1,228,031
25	1,090,313
57210	938,287
22	828,605
11963	699,102
23	696,571
10050	608,699
7275	561,205
6889	414,871
7628	397,209
7634	395,654
7372	395,443
7380	391,665
7636	390,305
7626	388,911

THE UNIVERSITY OF UTAH

Information Security Office

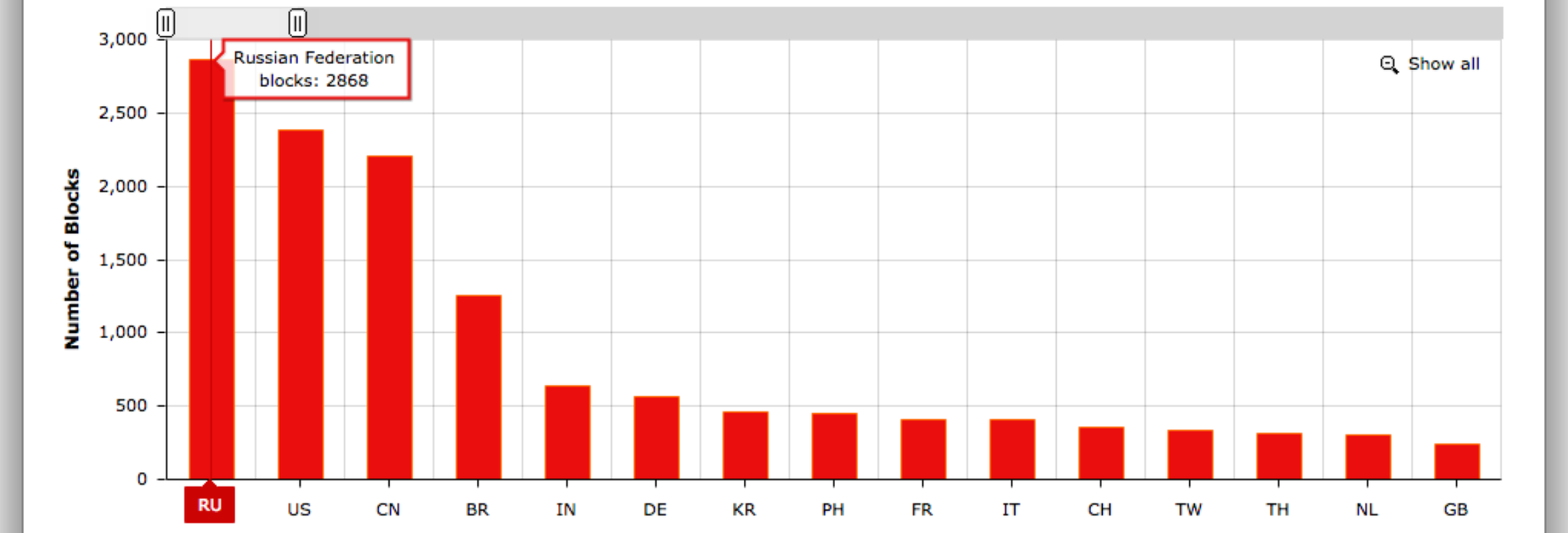
Destination Port Traffic

YYYYMMDD	port-666	port-7547	port-2323
20161008	14	487	503,131
20161009	22	469	430,642
20161010	13	541	465,448
20161011	12	613	522,745
20161012	11	638	503,572
20161029	10	613	284
…			
20161121	12	602	190
20161122	501	668	210
20161123	1255	670	254
20161124	1,343	562	290
20161125	1,676	634	307
20161126	2,052	1,948,000	205
20161127	1,813	1,958,141	198
20161128	1,001	1,292,690	133
20161129	905	1,105,433	170
20161130	1,200	269,421	168
20161201	1,246	236,270	193
20161202	1,772	188,490	203

Blocks per Country

[Go Back]

Updated: Wed Dec 7 08:52:06 MST 2016



THE UNIVERSITY OF UTAH

Information Security Office

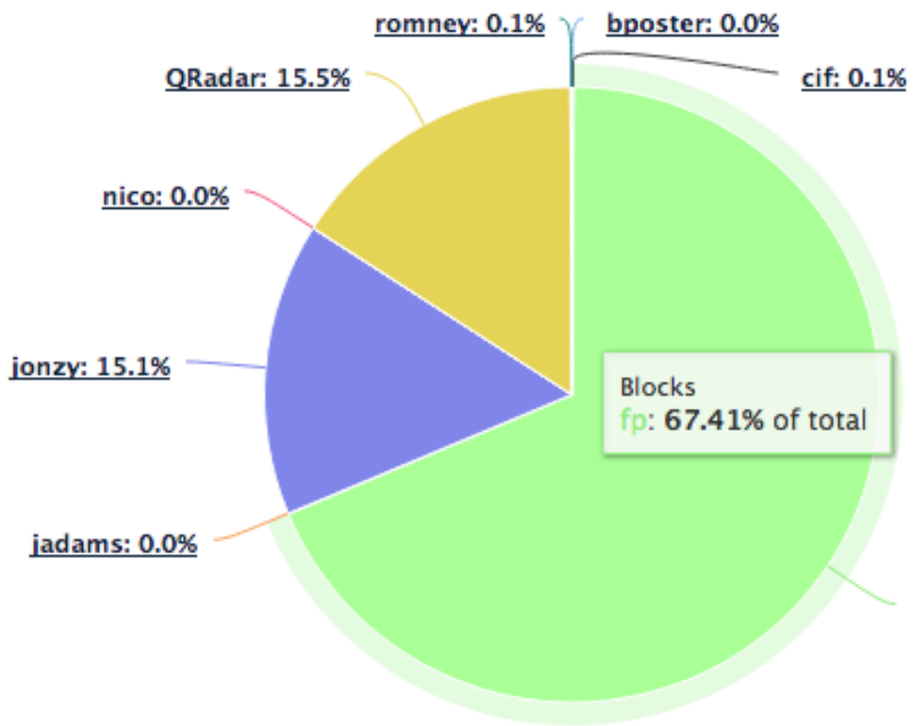
IP Blockage Breakdown

Go Back to Current Number of IP's Blocked

Updated: Wed Dec 7 08:48:32 MST 2016

Current IP Blockage.

Click the slices to view reason for Blockage.



Percentages represent the percentage of all IP's blocked.

THE UNIVERSITY OF UTAH

Information Security Office

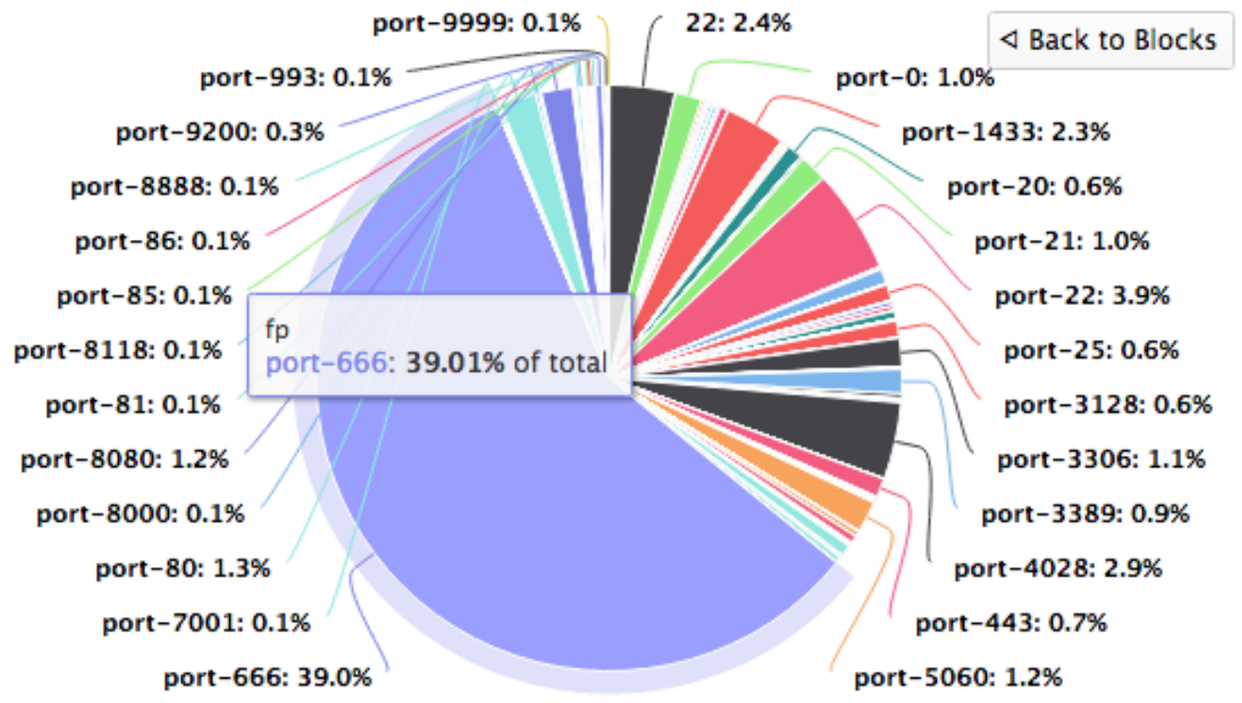
IP Blockage Breakdown

Go Back to Current Number of IP's Blocked

Updated: Wed Dec 7 08:48:32 MST 2016

Current IP Blockage.

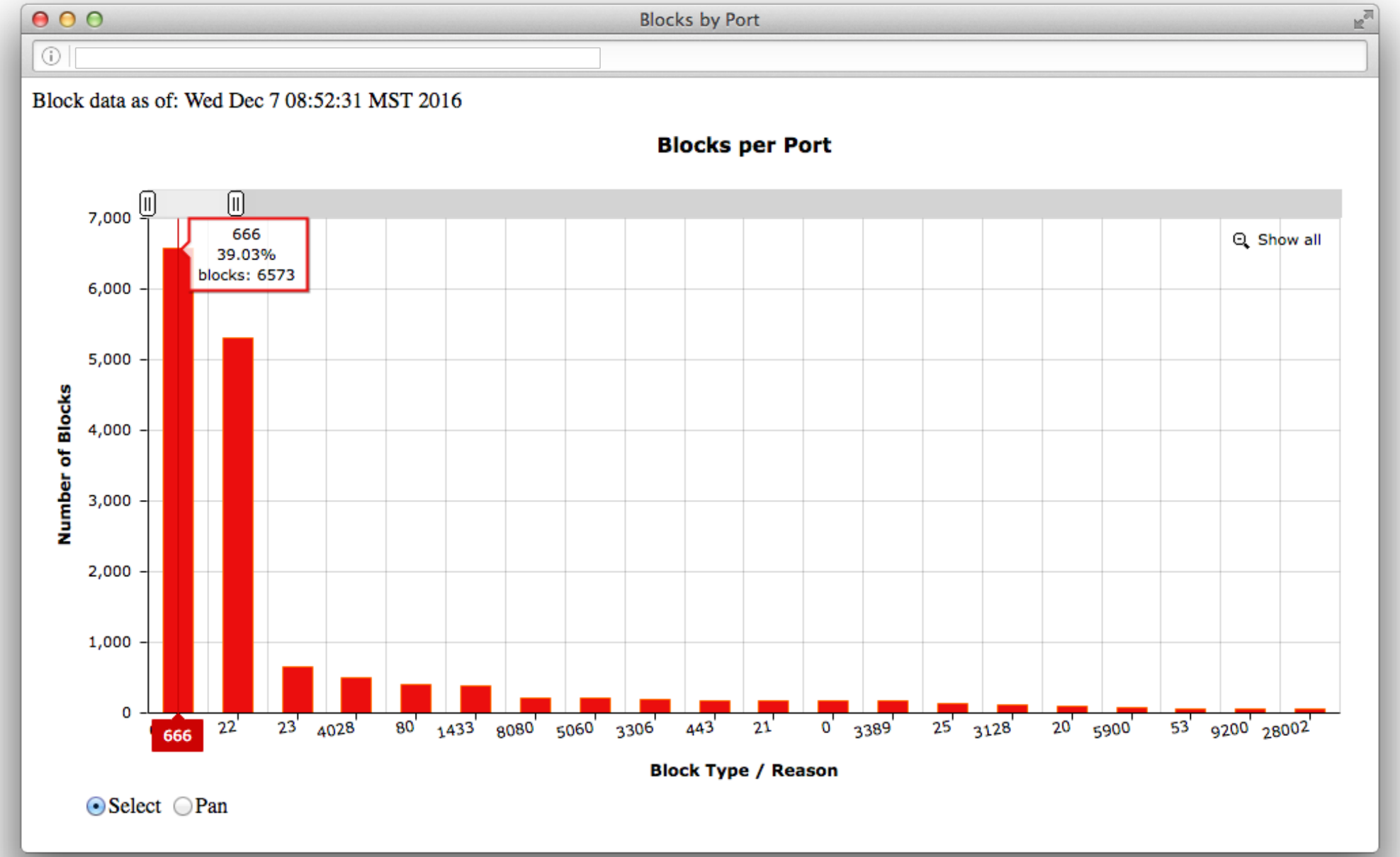
Click the slices to view reason for Blockage.



Percentages represent the percentage of all IP's blocked.

THE UNIVERSITY OF UTAH

Information Security Office



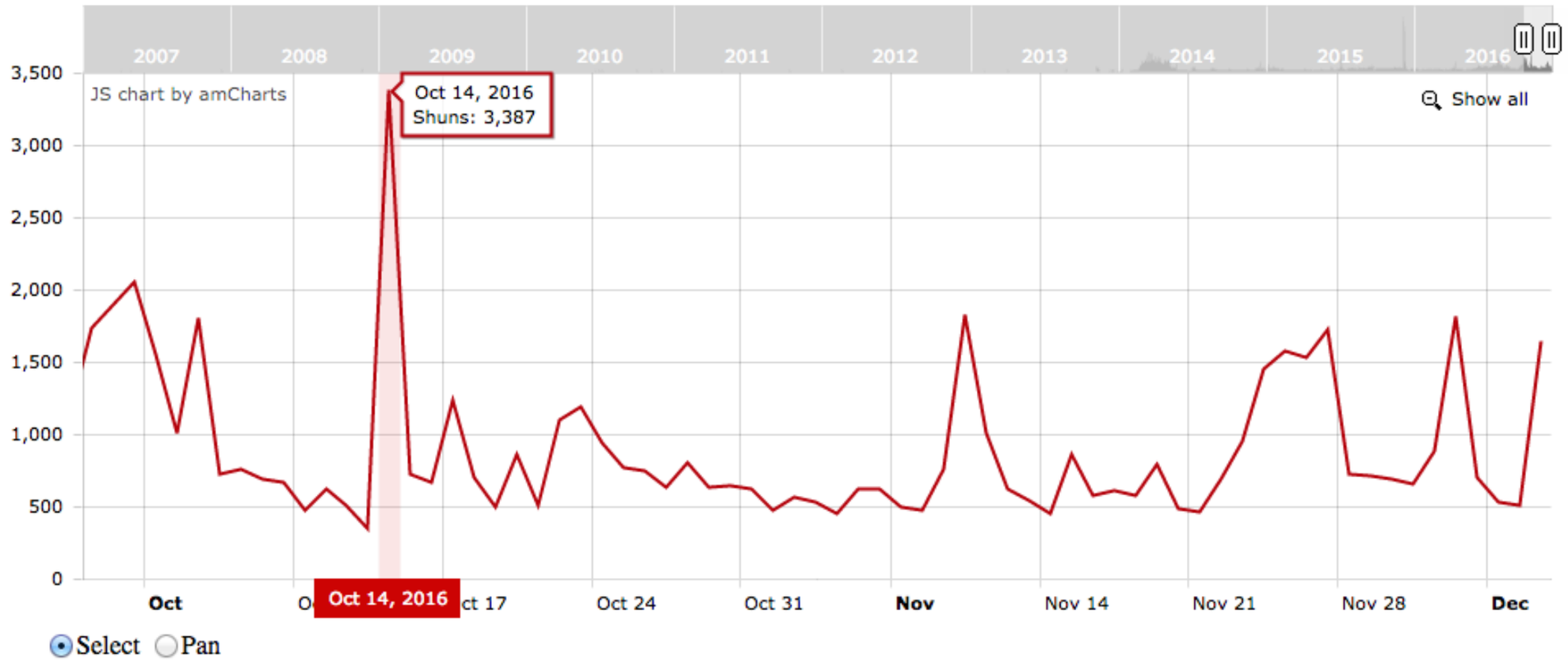
THE UNIVERSITY OF UTAH

Information Security Office

Shuns per Day

Shuns per Day

Go Back



THE UNIVERSITY OF UTAH

Information Security Office

Thresholds Identify Bad Actors

Any given IP generating X amount of flows per time period T, destined to N number of unique hosts is cause for alarm when:

X >= number of flows threshold, 128 for example
N >= unique destination IP's threshold, 75% or 96 for example

Caution: This does not guarantee a Bad Actor. Case in point, there may be a case where multiple local devices are accessing a 1 or more remote IP's for anything ranging from News, Patches, or a remote proxy.

Either way, looking are local responses, SYN flags, number of packets, and byte size can help identify problematic traffic.

Anything matching X >= 256 and N >= 75% or 192, where all the destination IP's reside in a /24 or contiguous set of Class-C Ciders, is almost 100% a remote probe.

Thresholds Identify Bad Actors

Another situation is a botnet probing your local network, where X is small, say16, and N is 16 – a possible probe. Additionally you may see a small X, say 16, with an N < 25% - a possible brute force attack.

The key with using thresholds, is determined by your environment.

Some thresholds will be different for different ports. Case in point, you may see a local host attempting to contact a remote host dozens of times a second, but this type of traffic would have X = ? but N = 1.

Thresholds Identify Bad Actors

Looking at remote traffic generating traffic where N=0 destined to any port over a 30 minute time frame sorted on TOTFLOWS shows the following:

#SOURCEIP__	TOTFLOWS	UNIQDIPS	UNQDPRTS	UNQSPRTS
172.217.5.78	121595	22754	33473	3
31.13.70.7	111641	14283	37774	2
64.27.28.157	102371	14	16368	1
31.13.70.1	101304	17615	30438	2
172.217.5.206	100658	20067	29119	3
31.13.70.36	92485	13073	33760	2
216.58.217.206	91262	19034	29163	3
216.58.193.206	81432	17576	27313	3
172.217.4.174	73586	17254	26665	3
172.217.4.142	72247	17311	26945	3
74.125.28.189	69199	6693	13536	1
31.13.70.52	67295	6092	23371	2
69.172.216.111	55571	4768	29061	2
64.39.105.128	47227	116	2504	30277
172.217.4.162	42821	9351	21540	3
31.13.70.14	42498	6269	23026	1
199.91.136.100	42330	475	22791	1
...				

Thresholds Identify Bad Actors

Looking at remote traffic generating traffic where N=0 destined to any port over a 30 minute time frame sorted on UNQDPRTS shows the following:

#SOURCEIP__	TOTFLOWS	UNIQDIPS	UNQDPRTS	UNQSPRTS
101.96.9.164	258	3	1	257
103.12.117.42	512	512	1	1
103.200.22.222	772	1	1	766
103.236.201.240	2247	2222	1	1
103.55.60.11	1023	1023	1	1009
104.254.231.71	134	1	1	128
107.191.0.122	131	2	1	116
107.23.48.136	442	1	1	441
107.23.49.116	479	1	1	476
107.77.227.153	350	1	1	323
107.77.227.199	312	3	1	298
107.77.228.42	273	1	1	263
107.77.228.66	1098	1	1	1032
107.77.228.76	188	2	1	188
107.77.229.11	1616	1	1	1592
107.77.230.40	148	7	1	147
108.51.19.160	198	1	1	198
...				

Thresholds Identify Bad Actors

Looking at remote traffic generating traffic where N=128 destined to port-22 over a 30 minute time frame shows the following:

#SOURCEIP__	TOTFLOWS	UNIQDIPS	UNQDPRTS	UNQSPRTS
31.3.245.23	492	492	1	1
64.39.105.128	170	62	1	150
103.55.60.11	1023	1023	1	1009
103.236.201.240	2247	2222	1	1
109.120.155.243	512	512	1	1
184.73.156.0	182	1	1	182
212.2.5.120	392	392	1	390

All of the above were auto-blocked due to thresholds with the exception of:

64.39.105.128
184.73.156.0

Thresholds Identify Bad Actors

Example where N=64 destined to port-7547 over 60 minute time frame:

#SOURCEIP__	TOTFLOWS	UNIQDIPS	UNQDPRTS	UNQSPRTS
107.170.103.171	83	83	1	83

Example where N=32 destined to port-666 over 60 minute time frame:

#SOURCEIP__	TOTFLOWS	UNIQDIPS	UNQDPRTS	UNQSPRTS
5.141.126.188	34	32	1	33
81.248.19.59	36	36	1	36
94.50.7.182	52	51	1	50
114.41.108.4	61	61	1	61

Example where N=16 destined to port-666 over 60 minute time frame:

#SOURCEIP__	TOTFLOWS	UNIQDIPS	UNQDPRTS	UNQSPRTS
5.141.126.188	34	32	1	33
64.39.105.128	22	19	1	22
81.248.19.59	36	36	1	36
94.50.7.182	52	51	1	50
114.41.108.4	61	61	1	61
117.93.26.42	31	31	1	31

Conclusion

Monitoring and tracking destination port usage is by no means a complete solution finding the “Needle in the Haystack”, but it definitely turns a needle into haystack.

Thresholds for the number of flows generated by remote IP's, per unique destination IP's also turns a needle into a haystack.

Using destination port analysis along with thresholds is one method for finding the Needle in the Haystack.