Going with the (s)Flow at 200G

Richard Yule<richard@linx.net>
Nigel Titley<nigel@titley.com>
Mike Hughes <mike@linx.net>

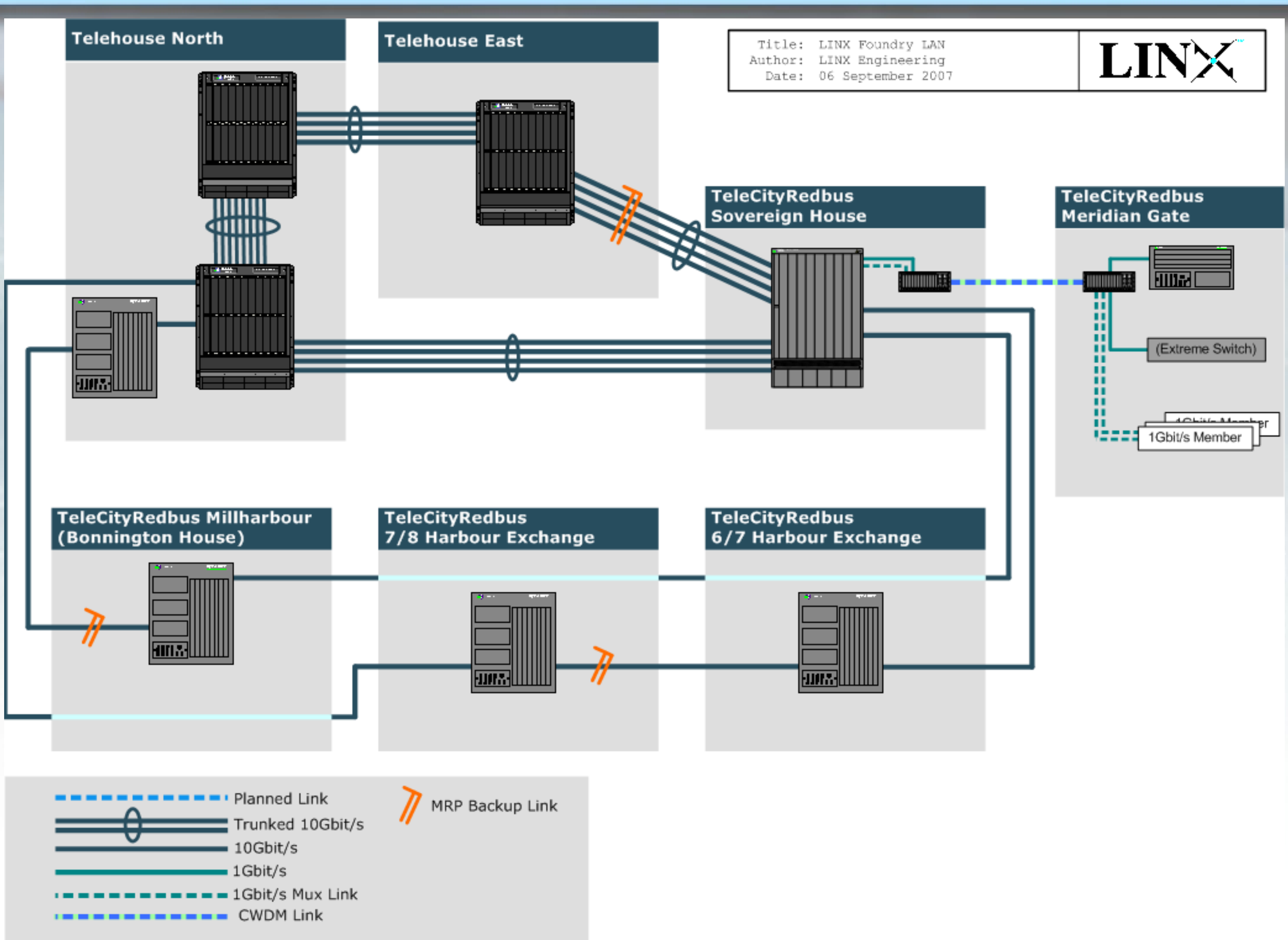
LINX Overview

- Founded in 1994 by 5 Members
- Non-profit, Mutual Ownership
- Now
  - 280 Members from 43 countries
  - 60% of the routing table peered
  - ~240Gbps peak (5 minute average)
  - Over 100 managed private interconnects
    - For large traffic flows

LINX Architecture

- Dual LAN Architecture
  - One LAN using Foundry switches
  - One LAN Extreme switches
- 7 sites in London Docklands
  - Connected by multiple diverse fibre rings
  - 8x10GE trunk ISL between top sites
  - 3 new sites to be connected in 2008
- Bi-lateral or multilateral peering
- 100M, 1GE and 10GE member ports

LINX Foundry Network



Title: LINX Foundry LAN
Author: LINX Engineering
Date: 06 September 2007

LINX
Telehouse North
Telehouse East
TeleCityRedbus Sovereign House
TeleCityRedbus Meridian Gate
(Extreme Switch)
1Gbit/s Member
1Gbit/s Member
TeleCityRedbus Millharbour (Bonnington House)
TeleCityRedbus 7/8 Harbour Exchange
TeleCityRedbus 6/7 Harbour Exchange
Planned Link
Trunked 10Gbit/s
10Gbit/s
1Gbit/s
1Gbit/s Mux Link
CWDM Link
MRP Backup Link

Engineering

Enter sFlow

- What is sFlow?
  - Defined in RFC 3176
  - A means of taking sampled traffic data from within a network
  - Works in Layer 2 networks (e.g. IXP)!
- sFlow agent on switch/router sends sFlow datagrams (UDP) to a sFlow collector
- sFlow collector runs analysis software

Purposes of LINX sFlow project

- Provide member to member statistics
  - For use by LINX members
  - For use by LINX engineering staff
- Provide engineering staff with tools such as
  - Traffic matrix (i.e. between nodes)
  - Peering matrix
  - Spot traffic anomalies
- More intelligence about our network

Challenges

- Most existing sFlow tools didn’t do what we wanted
  - Commercial: expensive, inflexible
  - Too Lightweight: couldn’t handle the data
  - Too Exhaustive: tried to extrapolate the data down to a specific L3 flow, scaling issues
  - Incomplete: only did part of the job
- Most importantly, little or no concept of “a member”
- We had to write our own tools

Phase 0.9: Proof of Concept

- First attempt was done using sflowtool feeding to pmacct
- Write the output into RRD files
- Problems
  - Constrained by disk I/O
  - Produced large unwieldy page of graphs
- Not very flexible
  - Borne out by minimal use

Switch Config

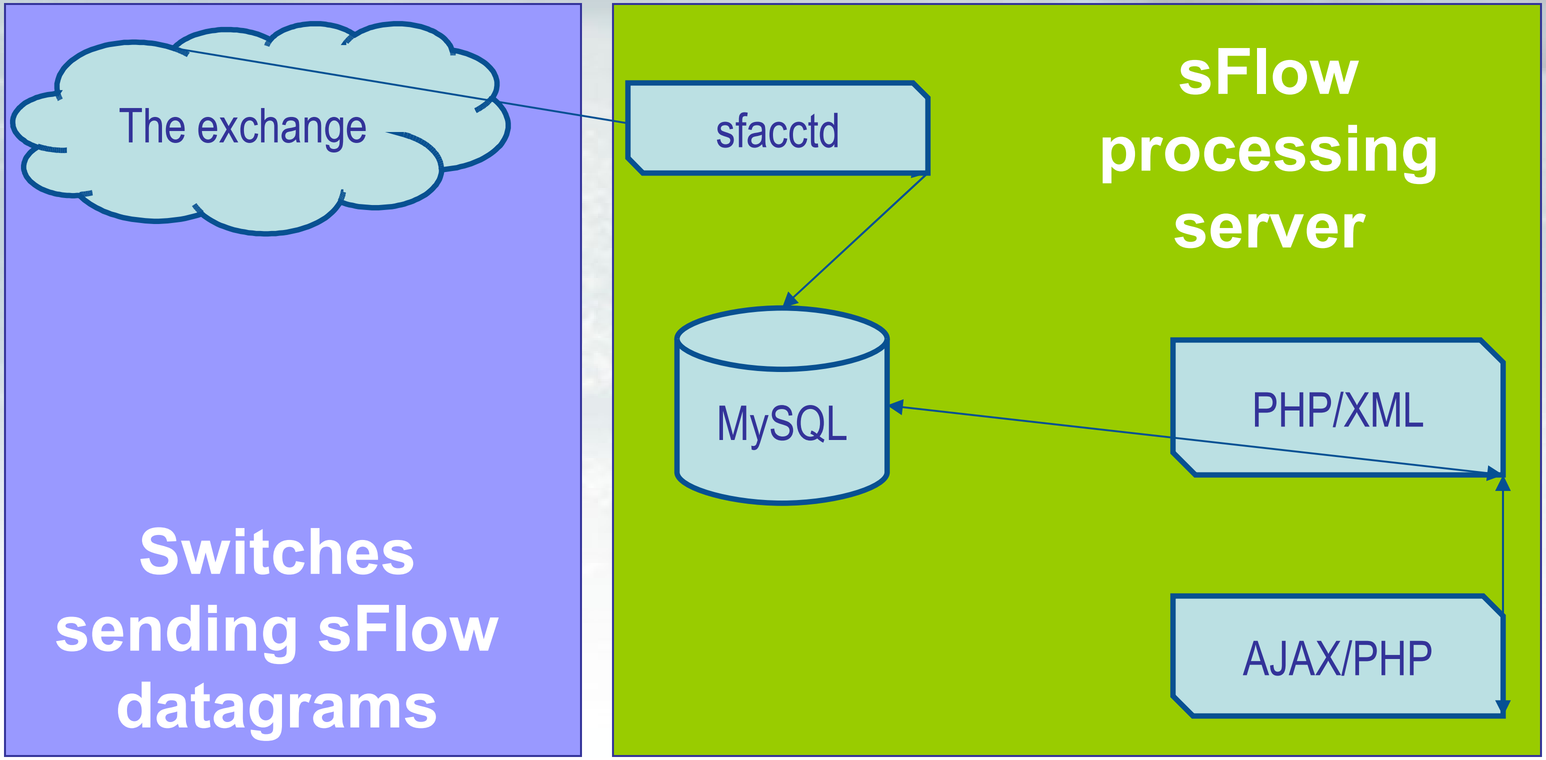
- All switches send sFlow packets over a VLAN interface in a specific sFlow VLAN – removes mgmt i/f concerns
- sFlow collector has interface to VLAN
- 1 in 2048 sample rate

```
sflow enable
sflow destination 172.22.0.90 6301
!
interface ethernet 4/4
 sflow forwarding
```

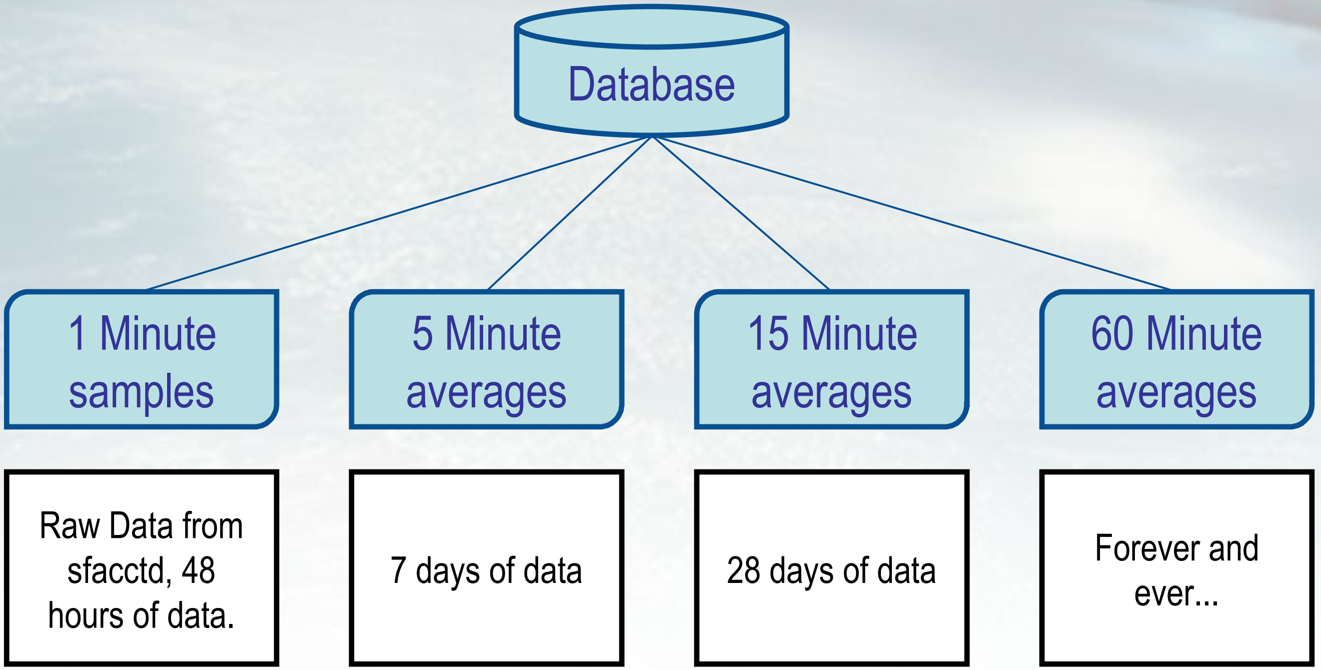
Phase 1: Cleansheet

- We knew what we wanted to achieve
- We brought in a programmer with good DB and web programming skills
- Allowed him to come at it from his own direction
  - No dictation about type of technology to use: Other than sFlow in, traffic data out.
- Took a fairly “minimalist” approach
  - Throwing away data we don’t need

Overview


The exchange
Switches sending sFlow datagrams
sfacctd
sFlow processing server
MySQL
PHP/XML
AJAX/PHP

Database Layout



Database

1 Minute samples

5 Minute averages

15 Minute averages

60 Minute averages

Raw Data from sfacctd, 48 hours of data.

7 days of data

28 days of data

Forever and ever...

Advantages and Limitations

- Upsides
  - We get more than just graphs from the same data sets
  - pmacctd gives us huge scope for functionality (pre-processing before insert)
- Downsides
  - Deleting data
  - Temporary tables
  - Joining tables

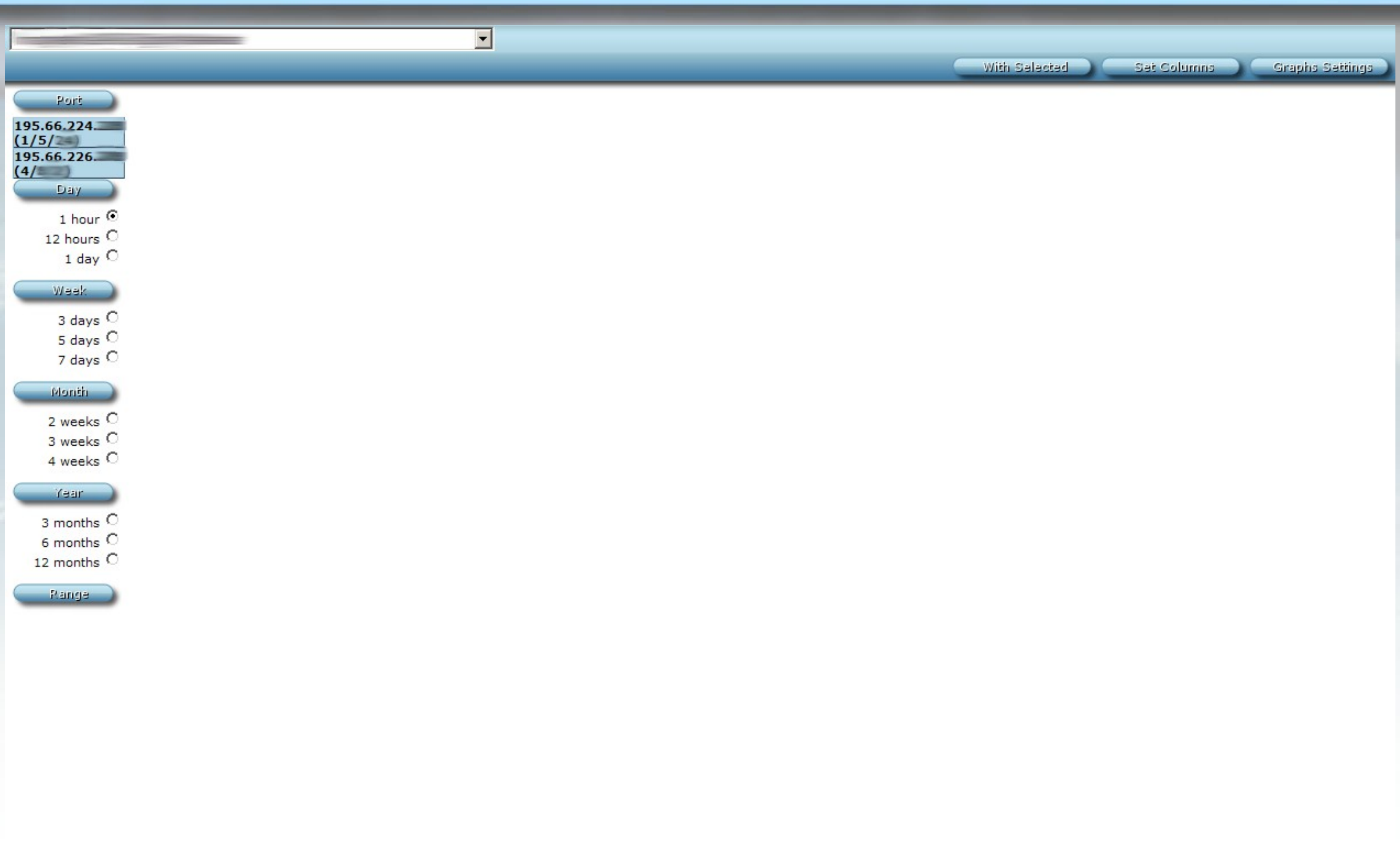
Hardware

- 2x dual-core CPU’s
- 16 GB of RAM
  - Can hold table indexes in memory
- 820GB RAID6 Array (8 x 146GB disks)
- Possible to scale hardware by running a distributed system
  - Mirror of system to allow for maintenance
  - Archive databases on different boxes

A few numbers

- 36 Million rows for 48 hours of 1 minute samples
- 32 Million rows for 7 days of 5 minute averages
- 43 Million rows for 4 weeks of 15 minute averages
- 40 Million rows for 3 months of 1 hour averages
- 147GB of data collected over the last 3 months

sFlow Portal Entry Screen



[redacted]

195.66.224.[redacted] (1/5/[redacted])
195.66.226.[redacted] (4/[redacted])

Day

1 hour ●
12 hours ○
1 day ○

Week

3 days ○
5 days ○
7 days ○

Month

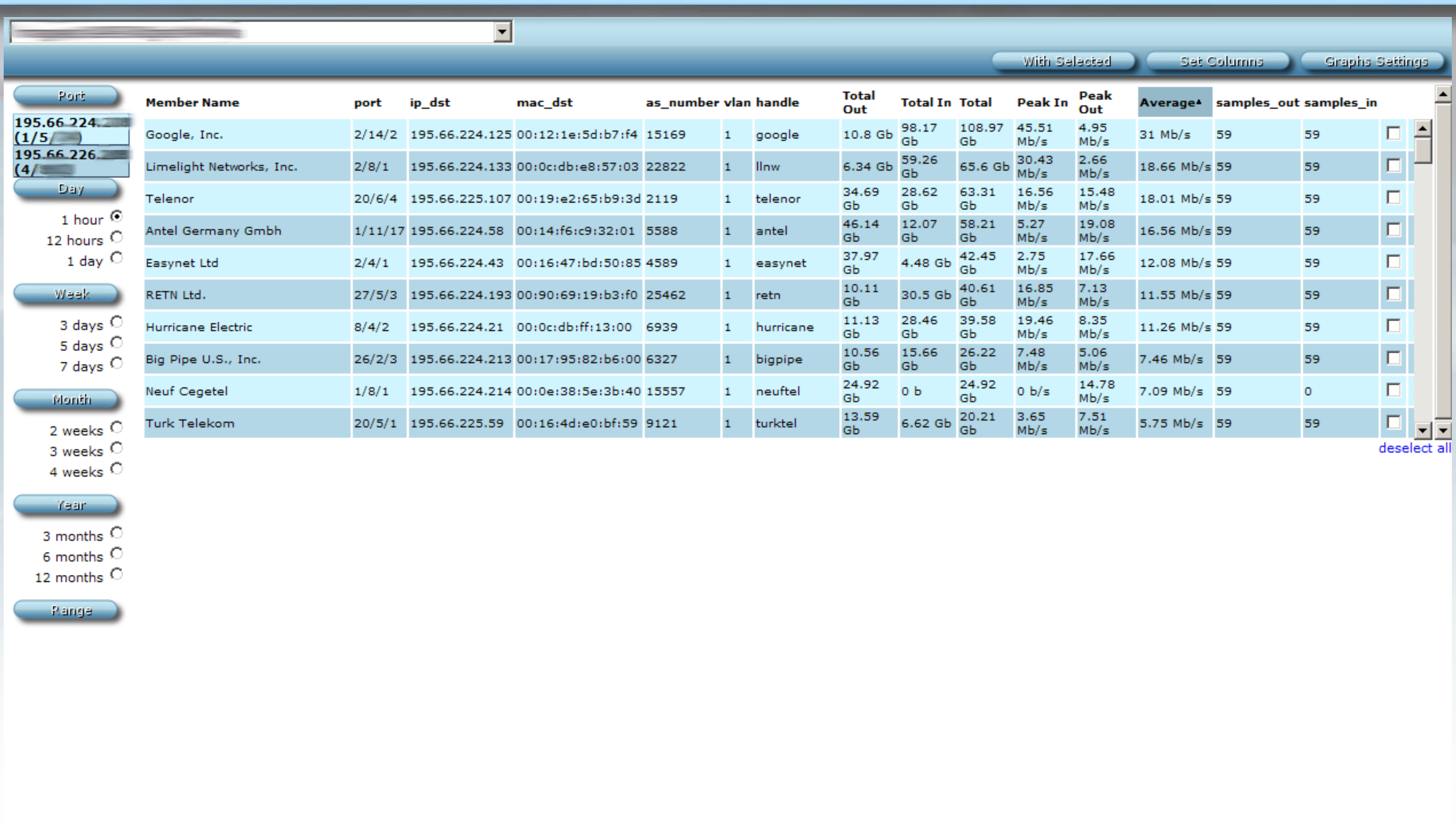
2 weeks ○
3 weeks ○
4 weeks ○

Year

3 months ○
6 months ○
12 months ○

Range

Select Your Switch Port



195.66.224.[redacted] (1/5/[redacted])

195.66.226.[redacted] (4/[redacted])

1 hour ●
12 hours ○
1 day ○

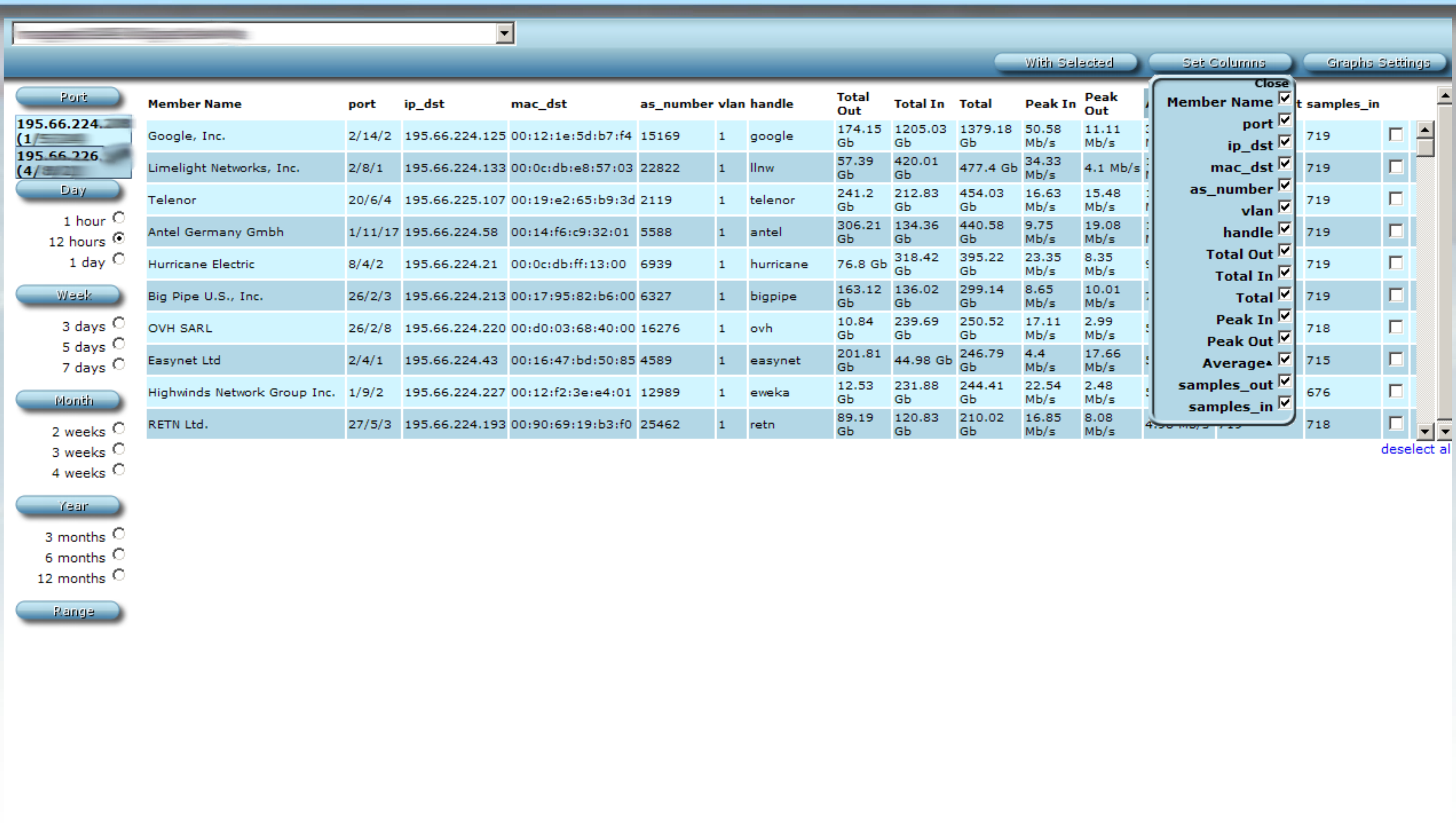
3 days ○
5 days ○
7 days ○

2 weeks ○
3 weeks ○
4 weeks ○

3 months ○
6 months ○
12 months ○

Member Name	port	ip_dst	mac_dst	as_number	vlan	handle	Total Out	Total In	Total	Peak In	Peak Out	Average▴	samples_out	samples_in	
Google, Inc.	2/14/2	195.66.224.125	00:12:1e:5d:b7:f4	15169	1	google	10.8 Gb	98.17 Gb	108.97 Gb	45.51 Mb/s	4.95 Mb/s	31 Mb/s	59	59	□
Limelight Networks, Inc.	2/8/1	195.66.224.133	00:0c:db:e8:57:03	22822	1	llnw	6.34 Gb	59.26 Gb	65.6 Gb	30.43 Mb/s	2.66 Mb/s	18.66 Mb/s	59	59	□
Telenor	20/6/4	195.66.225.107	00:19:e2:65:b9:3d	2119	1	telenor	34.69 Gb	28.62 Gb	63.31 Gb	16.56 Mb/s	15.48 Mb/s	18.01 Mb/s	59	59	□
Antel Germany Gmbh	1/11/17	195.66.224.58	00:14:f6:c9:32:01	5588	1	antel	46.14 Gb	12.07 Gb	58.21 Gb	5.27 Mb/s	19.08 Mb/s	16.56 Mb/s	59	59	□
Easynet Ltd	2/4/1	195.66.224.43	00:16:47:bd:50:85	4589	1	easynet	37.97 Gb	4.48 Gb	42.45 Gb	2.75 Mb/s	17.66 Mb/s	12.08 Mb/s	59	59	□
RETN Ltd.	27/5/3	195.66.224.193	00:90:69:19:b3:f0	25462	1	retn	10.11 Gb	30.5 Gb	40.61 Gb	16.85 Mb/s	7.13 Mb/s	11.55 Mb/s	59	59	□
Hurricane Electric	8/4/2	195.66.224.21	00:0c:db:ff:13:00	6939	1	hurricane	11.13 Gb	28.46 Gb	39.58 Gb	19.46 Mb/s	8.35 Mb/s	11.26 Mb/s	59	59	□
Big Pipe U.S., Inc.	26/2/3	195.66.224.213	00:17:95:82:b6:00	6327	1	bigpipe	10.56 Gb	15.66 Gb	26.22 Gb	7.48 Mb/s	5.06 Mb/s	7.46 Mb/s	59	59	□
Neuf Cegetel	1/8/1	195.66.224.214	00:0e:38:5e:3b:40	15557	1	neuftel	24.92 Gb	0 b	24.92 Gb	0 b/s	14.78 Mb/s	7.09 Mb/s	59	0	□
Turk Telekom	20/5/1	195.66.225.59	00:16:4d:e0:bf:59	9121	1	turktel	13.59 Gb	6.62 Gb	20.21 Gb	3.65 Mb/s	7.51 Mb/s	5.75 Mb/s	59	59	□

Select/Deselect Columns



[redacted]

195.66.224.[redacted]
(1/[redacted]
195.66.226.[redacted]
(4/[redacted]

1 hour ○
12 hours ●
1 day ○

3 days ○
5 days ○
7 days ○

2 weeks ○
3 weeks ○
4 weeks ○

3 months ○
6 months ○
12 months ○

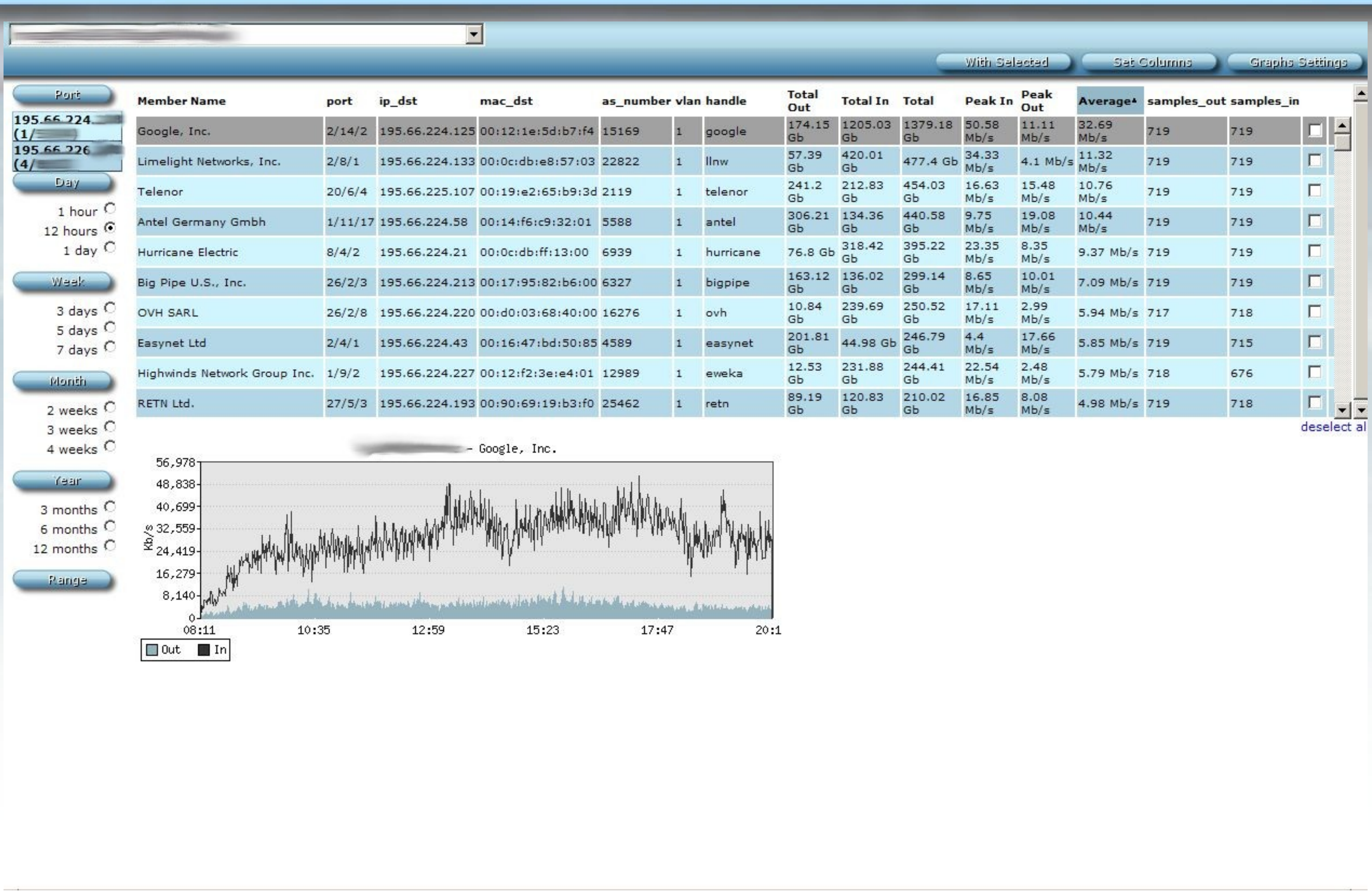
Member Name	port	ip_dst	mac_dst	as_number	vlan	handle	Total Out	Total In	Total	Peak In	Peak Out	samples_in	
Google, Inc.	2/14/2	195.66.224.125	00:12:1e:5d:b7:f4	15169	1	google	174.15 Gb	1205.03 Gb	1379.18 Gb	50.58 Mb/s	11.11 Mb/s	719	□
Limelight Networks, Inc.	2/8/1	195.66.224.133	00:0c:db:e8:57:03	22822	1	llnw	57.39 Gb	420.01 Gb	477.4 Gb	34.33 Mb/s	4.1 Mb/s	719	□
Telenor	20/6/4	195.66.225.107	00:19:e2:65:b9:3d	2119	1	telenor	241.2 Gb	212.83 Gb	454.03 Gb	16.63 Mb/s	15.48 Mb/s	719	□
Antel Germany Gmbh	1/11/17	195.66.224.58	00:14:f6:c9:32:01	5588	1	antel	306.21 Gb	134.36 Gb	440.58 Gb	9.75 Mb/s	19.08 Mb/s	719	□
Hurricane Electric	8/4/2	195.66.224.21	00:0c:db:ff:13:00	6939	1	hurricane	76.8 Gb	318.42 Gb	395.22 Gb	23.35 Mb/s	8.35 Mb/s	719	□
Big Pipe U.S., Inc.	26/2/3	195.66.224.213	00:17:95:82:b6:00	6327	1	bigpipe	163.12 Gb	136.02 Gb	299.14 Gb	8.65 Mb/s	10.01 Mb/s	719	□
OVH SARL	26/2/8	195.66.224.220	00:d0:03:68:40:00	16276	1	ovh	10.84 Gb	239.69 Gb	250.52 Gb	17.11 Mb/s	2.99 Mb/s	718	□
Easynet Ltd	2/4/1	195.66.224.43	00:16:47:bd:50:85	4589	1	easynet	201.81 Gb	44.98 Gb	246.79 Gb	4.4 Mb/s	17.66 Mb/s	715	□
Highwinds Network Group Inc.	1/9/2	195.66.224.227	00:12:f2:3e:e4:01	12989	1	eweka	12.53 Gb	231.88 Gb	244.41 Gb	22.54 Mb/s	2.48 Mb/s	676	□
RETN Ltd.	27/5/3	195.66.224.193	00:90:69:19:b3:f0	25462	1	retn	89.19 Gb	120.83 Gb	210.02 Gb	16.85 Mb/s	8.08 Mb/s	718	□

deselect al

Close

Member Name ☑
port ☑
ip_dst ☑
mac_dst ☑
as_number ☑
vlan ☑
handle ☑
Total Out ☑
Total In ☑
Total ☑
Peak In ☑
Peak Out ☑
Average ☑
samples_out ☑
samples_in ☑

Select Time Window, Show Graph



[redacted]

With Selected | Set Columns | Graphs Settings

Port

195.66.224.[redacted] (1/[redacted]

195.66.226.[redacted] (4/[redacted]

Day

1 hour ○
12 hours ●
1 day ○

Week

3 days ○
5 days ○
7 days ○

Month

2 weeks ○
3 weeks ○
4 weeks ○

Year

3 months ○
6 months ○
12 months ○

Range

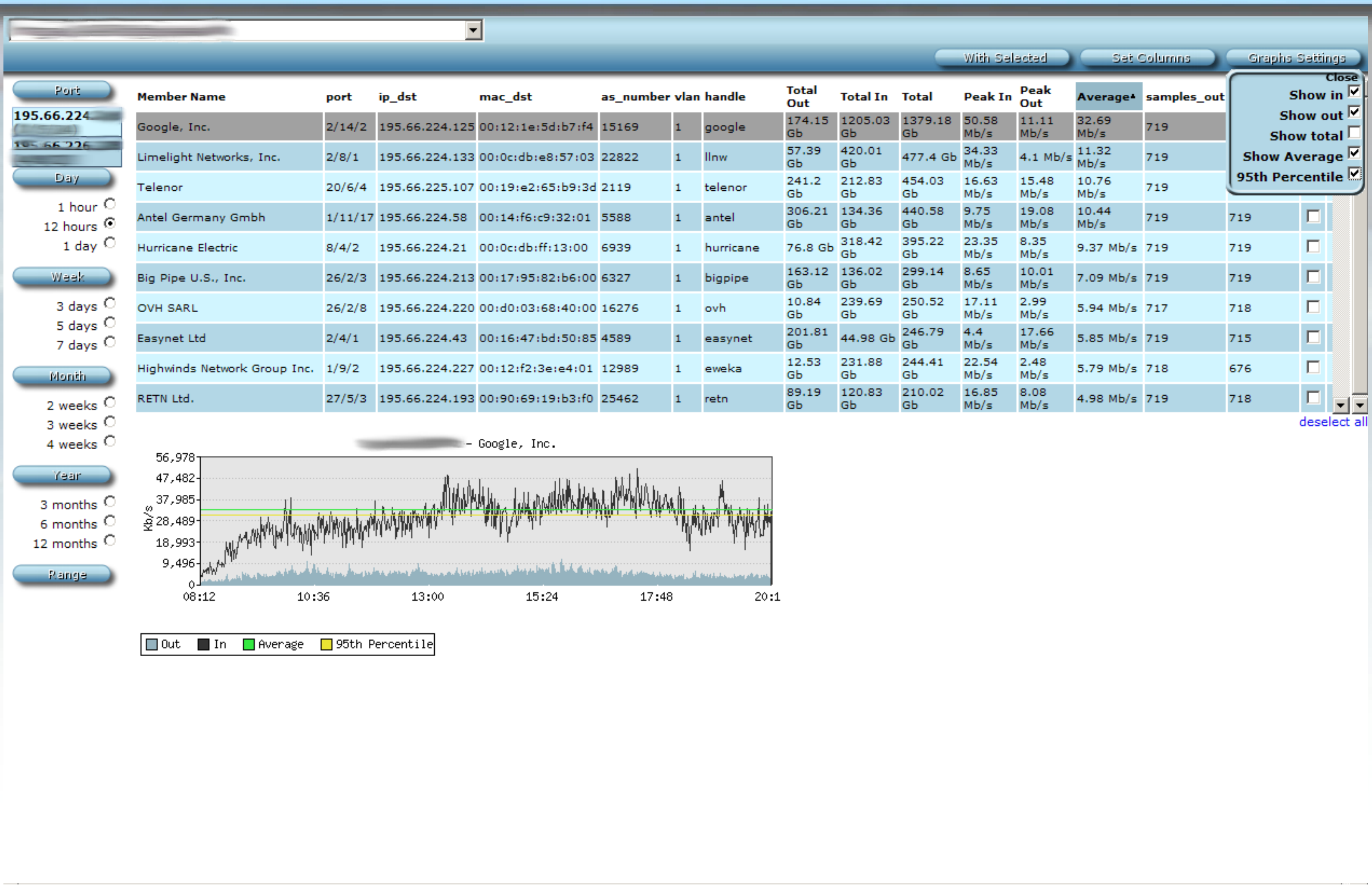
Member Name	port	ip_dst	mac_dst	as_number	vlan	handle	Total Out	Total In	Total	Peak In	Peak Out	Average▴	samples_out	samples_in	
Google, Inc.	2/14/2	195.66.224.125	00:12:1e:5d:b7:f4	15169	1	google	174.15 Gb	1205.03 Gb	1379.18 Gb	50.58 Mb/s	11.11 Mb/s	32.69 Mb/s	719	719	☐
Limelight Networks, Inc.	2/8/1	195.66.224.133	00:0c:db:e8:57:03	22822	1	llnw	57.39 Gb	420.01 Gb	477.4 Gb	34.33 Mb/s	4.1 Mb/s	11.32 Mb/s	719	719	☐
Telenor	20/6/4	195.66.225.107	00:19:e2:65:b9:3d	2119	1	telenor	241.2 Gb	212.83 Gb	454.03 Gb	16.63 Mb/s	15.48 Mb/s	10.76 Mb/s	719	719	☐
Antel Germany Gmbh	1/11/17	195.66.224.58	00:14:f6:c9:32:01	5588	1	antel	306.21 Gb	134.36 Gb	440.58 Gb	9.75 Mb/s	19.08 Mb/s	10.44 Mb/s	719	719	☐
Hurricane Electric	8/4/2	195.66.224.21	00:0c:db:ff:13:00	6939	1	hurricane	76.8 Gb	318.42 Gb	395.22 Gb	23.35 Mb/s	8.35 Mb/s	9.37 Mb/s	719	719	☐
Big Pipe U.S., Inc.	26/2/3	195.66.224.213	00:17:95:82:b6:00	6327	1	bigpipe	163.12 Gb	136.02 Gb	299.14 Gb	8.65 Mb/s	10.01 Mb/s	7.09 Mb/s	719	719	☐
OVH SARL	26/2/8	195.66.224.220	00:d0:03:68:40:00	16276	1	ovh	10.84 Gb	239.69 Gb	250.52 Gb	17.11 Mb/s	2.99 Mb/s	5.94 Mb/s	717	718	☐
Easynet Ltd	2/4/1	195.66.224.43	00:16:47:bd:50:85	4589	1	easynet	201.81 Gb	44.98 Gb	246.79 Gb	4.4 Mb/s	17.66 Mb/s	5.85 Mb/s	719	715	☐
Highwinds Network Group Inc.	1/9/2	195.66.224.227	00:12:f2:3e:e4:01	12989	1	eweka	12.53 Gb	231.88 Gb	244.41 Gb	22.54 Mb/s	2.48 Mb/s	5.79 Mb/s	718	676	☐
RETN Ltd.	27/5/3	195.66.224.193	00:90:69:19:b3:f0	25462	1	retn	89.19 Gb	120.83 Gb	210.02 Gb	16.85 Mb/s	8.08 Mb/s	4.98 Mb/s	719	718	☐

deselect al

[redacted] - Google, Inc.

Engineering

Show Average & 95th Percentile



195.66.224 [redacted]

[redacted]

1 hour ○
12 hours ●
1 day ○

3 days ○
5 days ○
7 days ○

2 weeks ○
3 weeks ○
4 weeks ○

3 months ○
6 months ○
12 months ○

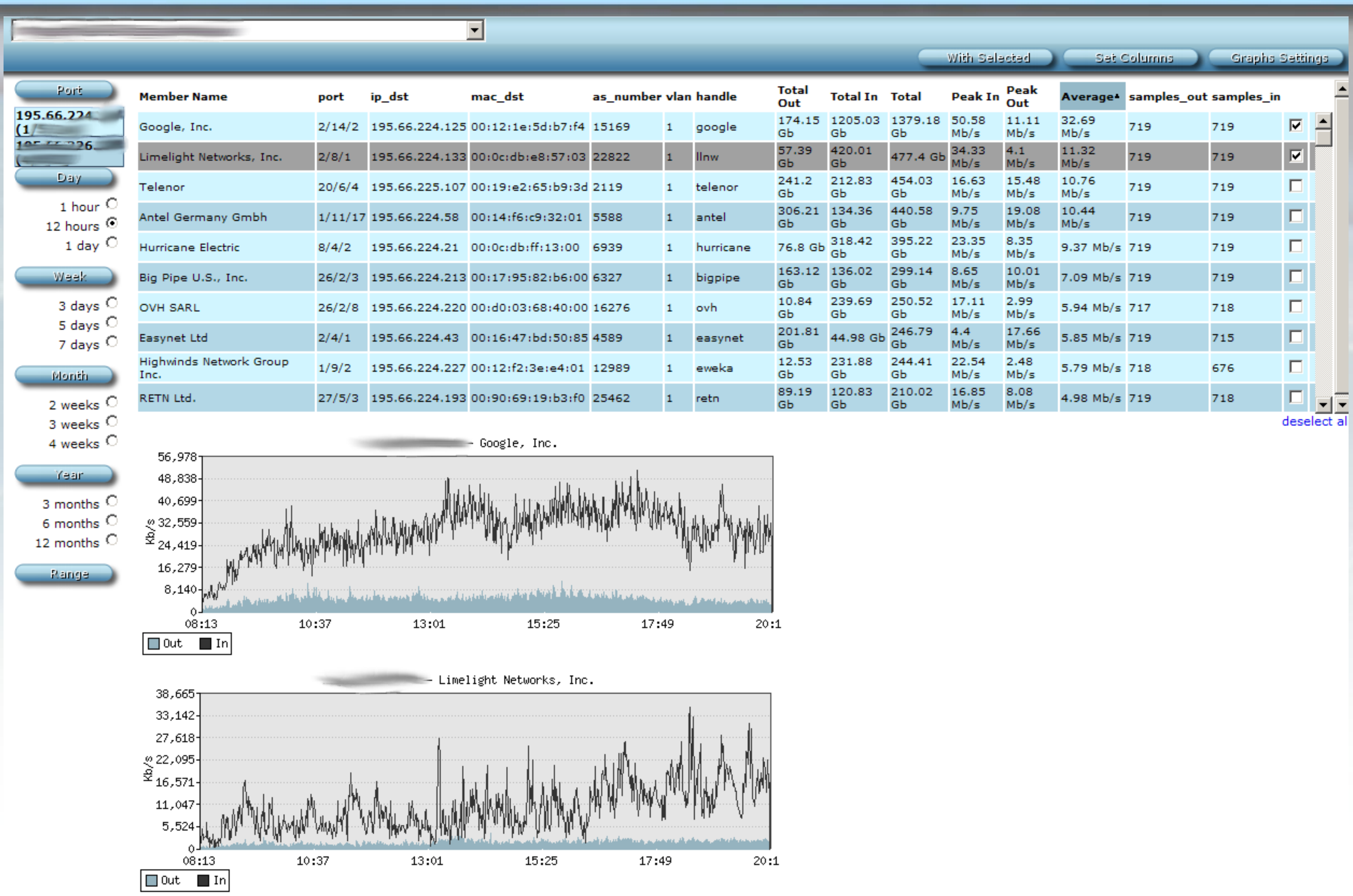
Close
Show in ☑
Show out ☑
Show total □
Show Average ☑
95th Percentile ☑

Member Name	port	ip_dst	mac_dst	as_number	vlan	handle	Total Out	Total In	Total	Peak In	Peak Out	Average▲	samples_out		
Google, Inc.	2/14/2	195.66.224.125	00:12:1e:5d:b7:f4	15169	1	google	174.15 Gb	1205.03 Gb	1379.18 Gb	50.58 Mb/s	11.11 Mb/s	32.69 Mb/s	719		
Limelight Networks, Inc.	2/8/1	195.66.224.133	00:0c:db:e8:57:03	22822	1	llnw	57.39 Gb	420.01 Gb	477.4 Gb	34.33 Mb/s	4.1 Mb/s	11.32 Mb/s	719		
Telenor	20/6/4	195.66.225.107	00:19:e2:65:b9:3d	2119	1	telenor	241.2 Gb	212.83 Gb	454.03 Gb	16.63 Mb/s	15.48 Mb/s	10.76 Mb/s	719		
Antel Germany Gmbh	1/11/17	195.66.224.58	00:14:f6:c9:32:01	5588	1	antel	306.21 Gb	134.36 Gb	440.58 Gb	9.75 Mb/s	19.08 Mb/s	10.44 Mb/s	719	719	□
Hurricane Electric	8/4/2	195.66.224.21	00:0c:db:ff:13:00	6939	1	hurricane	76.8 Gb	318.42 Gb	395.22 Gb	23.35 Mb/s	8.35 Mb/s	9.37 Mb/s	719	719	□
Big Pipe U.S., Inc.	26/2/3	195.66.224.213	00:17:95:82:b6:00	6327	1	bigpipe	163.12 Gb	136.02 Gb	299.14 Gb	8.65 Mb/s	10.01 Mb/s	7.09 Mb/s	719	719	□
OVH SARL	26/2/8	195.66.224.220	00:d0:03:68:40:00	16276	1	ovh	10.84 Gb	239.69 Gb	250.52 Gb	17.11 Mb/s	2.99 Mb/s	5.94 Mb/s	717	718	□
Easynet Ltd	2/4/1	195.66.224.43	00:16:47:bd:50:85	4589	1	easynet	201.81 Gb	44.98 Gb	246.79 Gb	4.4 Mb/s	17.66 Mb/s	5.85 Mb/s	719	715	□
Highwinds Network Group Inc.	1/9/2	195.66.224.227	00:12:f2:3e:e4:01	12989	1	eweka	12.53 Gb	231.88 Gb	244.41 Gb	22.54 Mb/s	2.48 Mb/s	5.79 Mb/s	718	676	□
RETN Ltd.	27/5/3	195.66.224.193	00:90:69:19:b3:f0	25462	1	retn	89.19 Gb	120.83 Gb	210.02 Gb	16.85 Mb/s	8.08 Mb/s	4.98 Mb/s	719	718	□

deselect all

[redacted] - Google, Inc.

Select multiple peers & compare



[redacted]

195.66.224 [redacted]
(1 [redacted]
[redacted] 226 [redacted]

1 hour ○
12 hours ●
1 day ○

3 days ○
5 days ○
7 days ○

2 weeks ○
3 weeks ○
4 weeks ○

3 months ○
6 months ○
12 months ○

Member Name	port	ip_dst	mac_dst	as_number	vlan	handle	Total Out	Total In	Total	Peak In	Peak Out	Average▴	samples_out	samples_in	
Google, Inc.	2/14/2	195.66.224.125	00:12:1e:5d:b7:f4	15169	1	google	174.15 Gb	1205.03 Gb	1379.18 Gb	50.58 Mb/s	11.11 Mb/s	32.69 Mb/s	719	719	☑
Limelight Networks, Inc.	2/8/1	195.66.224.133	00:0c:db:e8:57:03	22822	1	llnw	57.39 Gb	420.01 Gb	477.4 Gb	34.33 Mb/s	4.1 Mb/s	11.32 Mb/s	719	719	☑
Telenor	20/6/4	195.66.225.107	00:19:e2:65:b9:3d	2119	1	telenor	241.2 Gb	212.83 Gb	454.03 Gb	16.63 Mb/s	15.48 Mb/s	10.76 Mb/s	719	719	□
Antel Germany Gmbh	1/11/17	195.66.224.58	00:14:f6:c9:32:01	5588	1	antel	306.21 Gb	134.36 Gb	440.58 Gb	9.75 Mb/s	19.08 Mb/s	10.44 Mb/s	719	719	□
Hurricane Electric	8/4/2	195.66.224.21	00:0c:db:ff:13:00	6939	1	hurricane	76.8 Gb	318.42 Gb	395.22 Gb	23.35 Mb/s	8.35 Mb/s	9.37 Mb/s	719	719	□
Big Pipe U.S., Inc.	26/2/3	195.66.224.213	00:17:95:82:b6:00	6327	1	bigpipe	163.12 Gb	136.02 Gb	299.14 Gb	8.65 Mb/s	10.01 Mb/s	7.09 Mb/s	719	719	□
OVH SARL	26/2/8	195.66.224.220	00:d0:03:68:40:00	16276	1	ovh	10.84 Gb	239.69 Gb	250.52 Gb	17.11 Mb/s	2.99 Mb/s	5.94 Mb/s	717	718	□
Easynet Ltd	2/4/1	195.66.224.43	00:16:47:bd:50:85	4589	1	easynet	201.81 Gb	44.98 Gb	246.79 Gb	4.4 Mb/s	17.66 Mb/s	5.85 Mb/s	719	715	□
Highwinds Network Group Inc.	1/9/2	195.66.224.227	00:12:f2:3e:e4:01	12989	1	eweka	12.53 Gb	231.88 Gb	244.41 Gb	22.54 Mb/s	2.48 Mb/s	5.79 Mb/s	718	676	□
RETN Ltd.	27/5/3	195.66.224.193	00:90:69:19:b3:f0	25462	1	retn	89.19 Gb	120.83 Gb	210.02 Gb	16.85 Mb/s	8.08 Mb/s	4.98 Mb/s	719	718	□

deselect al

[redacted] - Google, Inc.

[redacted] - Limelight Networks, Inc.

Engineering

Future Wishlist

- Add-in processing of Extreme data
  - Held up due to implementation issues
- XML schema and authenticated direct XML interface for members
  - Integrate directly into their own systems
  - Perform direct queries of the db
- Peering Matrix
- Improve engineering tools
  - “Toptalkers”, interswitch traffic matrix

Other wacky ideas?

- Pro-active notification agent
  - Be able to configure various thresholds, receive alerts
- Weekly “overview report”

End Results

- Allow members to manage their peerings more intelligently
- Allow LINX to better understand flows inside the peering networks
- Identify traffic flows for optimisation
  - Switch platform relief
  - Through PNI or regrooming of member connections onto same switch

Questions?

-