

A primer on network flow visualization

Gregory Travis

Advanced Network Management Lab

Indiana University

greg@iu.edu

Problem: Seeing the Forest through the trees

“Too much information”

- Abilene generating 5-6,000 flows/second
- Typically about 270,000-350,000 “active” active flows during the day

“Raw” data analysis inadequate

- Forest through trees

SNORT raw log file example

cybersecuritylabs.iu.edu

```
[ (spp_portscan2) Portscan detected from 207.75.xxx.xxx: 4 targets 21 ports in 28 seconds [**]  
5.727011 207.75.xxx.xxx:80 -> 149.165.xxx.xxx:49194  
S:0x0 ID:0 IpLen:20 DgmLen:60 DF  
0xD756E195 Ack: 0xDDC23C59 Win: 0x16A0 TcpLen: 40  
6) => MSS: 1460 NOP NOP TS: 518109736 2681681736  
> NOP WS: 0
```

```
[ (snort_decoder): Short UDP packet, length field > payload length [**]  
8.526214 149.165.xxx.xxx:0 -> 149.165.xxx.xxx:0  
OS:0x0 ID:16642 IpLen:20 DgmLen:206
```

```
[ ICMP Destination Unreachable (Communication Administratively Prohibited) [**]  
on: Misc activity] [Priority: 3]  
1.494517 128.109.xxx.xxx -> 149.165.xxx.xxx  
TOS:0x0 ID:0 IpLen:20 DgmLen:56  
13 DESTINATION UNREACHABLE: ADMINISTRATIVELY PROHIBITED,  
ED  
ATAGRAM DUMP:  
xx -> 149.168.xxx.xxx  
TOS:0x0 ID:2394 IpLen:20 DgmLen:92  
P
```

```
[ (spp_rpc_decode) Incomplete RPC segment [**]  
2.345311 64.12.xxx.xxx:5190 -> 149.165.xxx.xxx:32771  
OS:0x0 ID:45414 IpLen:20 DgmLen:98 DF  
0xD9256CFA Ack: 0xC79F78B9 Win: 0x4000 TcpLen: 20
```

```
[ (spp_stream4) STEALTH ACTIVITY (FIN scan) detection [**]  
0.235714 66.250.xxx.xxx:25111 -> 149.165.xxx.xxx:13091  
S:0x0 ID:59791 IpLen:20 DgmLen:52 DF  
0x32BE0760 Ack: 0x0 Win: 0xFFFF TcpLen: 32  
3) => NOP NOP TS: 234082903 0
```

Problems with that

Visually unattractive

- “Angry Fruit Salad”

Information overload

False-positives

Forest through the trees

Evolution of visualization techniques

- Text-Based
- 2D visualization of old text information
 - I.e. ACID interface to SNORT

ACID display

Analysis Console for Intrusion Databases (ACID)

https://.../acid/acid_main.php

Apple .Mac eBay Yahoo! News comics geek ANML DDOS Project Page Metafilter Shortcuts weather OpenGL - Examples

ANML customized Analysis Console for Intrusion Databases

Added 0 alert(s) to the Alert cache

Queried on : Wed January 21, 2004 11:56:48
 Database: snort_db2@localhost (schema version: 106)
 Time window: [2003-10-14 09:50:20-05] - [2003-10-17 12:51:47-05]

<p>Sensors: 1</p> <p>Unique Alerts: 16 (3 categories)</p> <p>Total Number of Alerts: 26913</p> <ul style="list-style-type: none"> Source IP addresses: 3869 Dest. IP addresses: 296 Unique IP links 4431 Source Ports: 1251 <ul style="list-style-type: none"> TCP (532) UDP (726) Dest. Ports: 477 <ul style="list-style-type: none"> TCP (435) UDP (47) 	<p>Traffic Profile by Protocol</p> <p>TCP (7%)</p> <p>UDP (9%)</p> <p>ICMP (84%)</p> <p>Portscan Traffic (0%)</p>
--	--

- [Search](#)
- [Graph Alert data](#)
- **Snapshot**
 - Most recent Alerts: [any protocol](#), [TCP](#), [UDP](#), [ICMP](#)
 - Today's: alerts [unique](#), [listing](#); IP [src](#) / [dst](#)
 - Last 24 Hours: alerts [unique](#), [listing](#); IP [src](#) / [dst](#)
 - Last 72 Hours: alerts [unique](#), [listing](#); IP [src](#) / [dst](#)
 - Most [recent 15 Unique Alerts](#)
 - Most frequent 5 Alerts
 - Most Frequent Source Ports: [any](#), [TCP](#), [UDP](#)
 - Most Frequent Destination Ports: [any](#), [TCP](#), [UDP](#)
 - Most frequent 15 addresses: [source](#), [destination](#)
 - Last Source Ports: [any](#), [TCP](#), [UDP](#)
 - Last Destination Ports: [any](#), [TCP](#), [UDP](#)
- Graph alert [detection time](#)
- Alert Group (AG) [maintenance](#)
- Application [cache and status](#)

[Loaded in 5 seconds]

ACID display

Ok, getting better. System is doing some aggregating for us.

We have some visualization (traffic profile)

But still showing us the same alerts, the vast majority of which are not actually issues.

Emergence of statistical tools

Next step was emergence of so-called statistical tools

Idea of establishing a baseline of “normal” activity

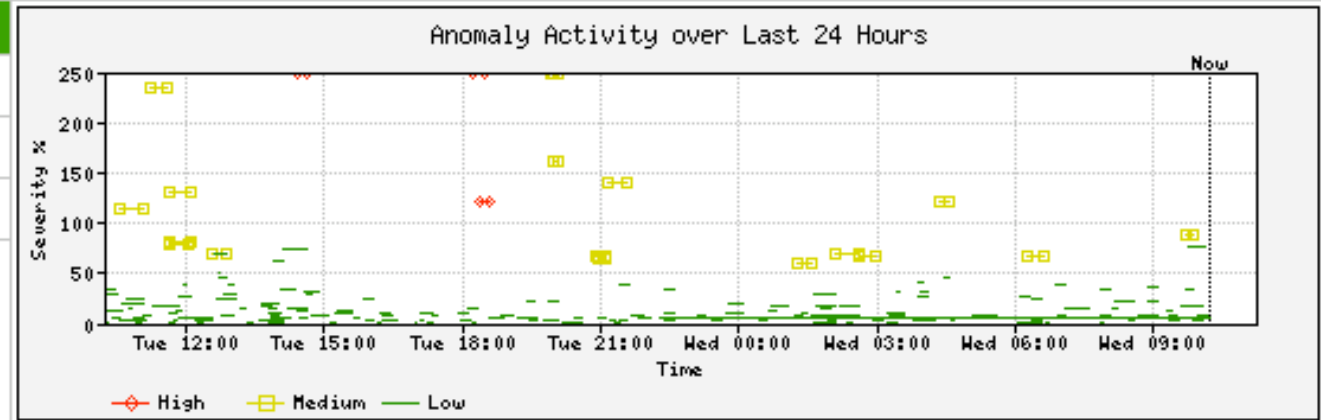
Detect deviations from “normal”

Throw a nice 2D front-end on it

ARBOR display

Snapshot

Totals	High	Medium	Low
[23]:	0	0	23
[2227]:	4	37	2186
[2225]:	4	18	2203



Statistical tools

But the bias is still there

- What's more damning, overreporting or underreporting

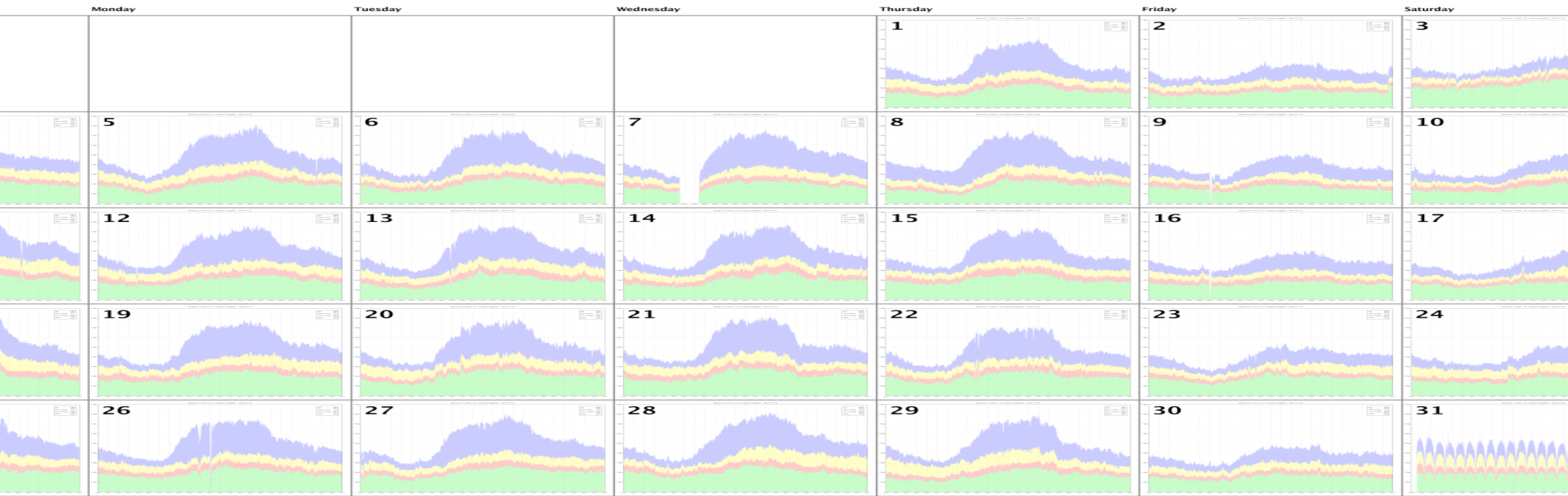
And you have to be able to establish a baseline of “normal” activity

- Not possible in dynamic environment
- Miss low-level “noise” activity

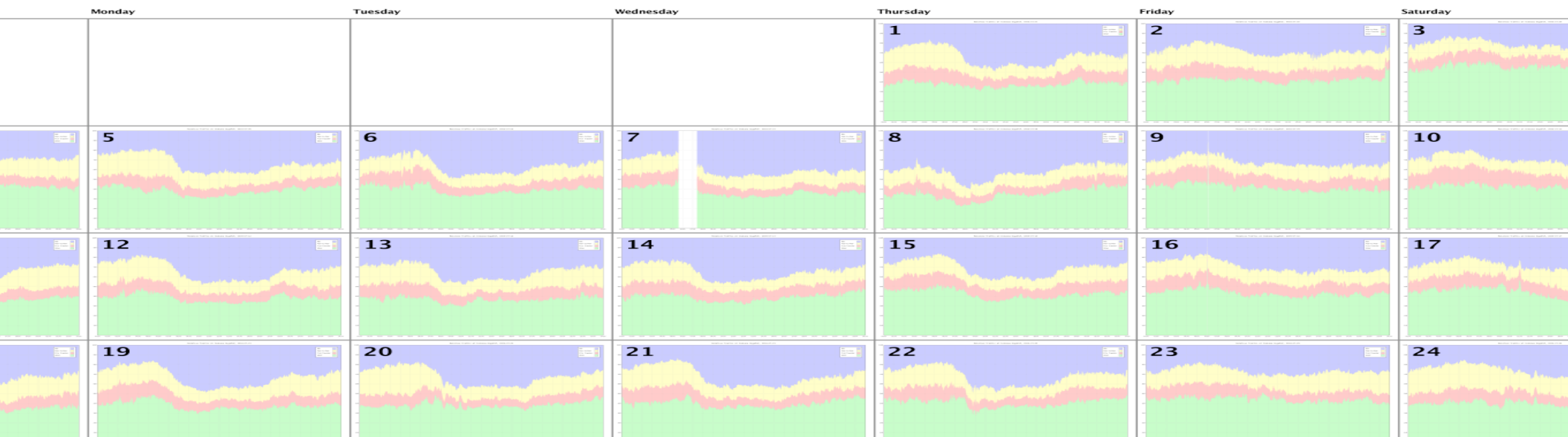
Some more examples

Pure Visualization Tools

Absolute Traffic at Indiana GigaPop, July 2004

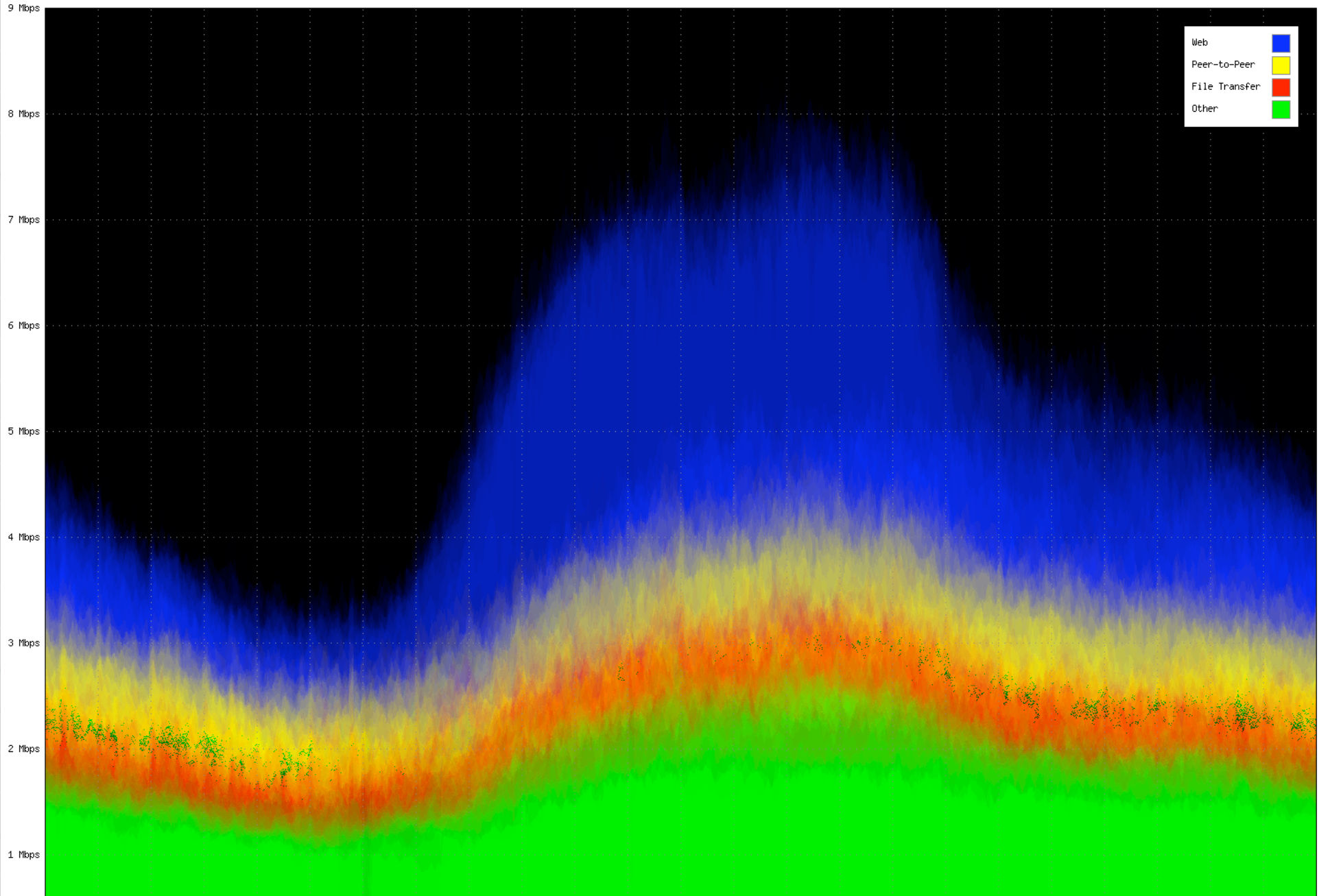


Relative Traffic at Indiana GigaPop, July 2004



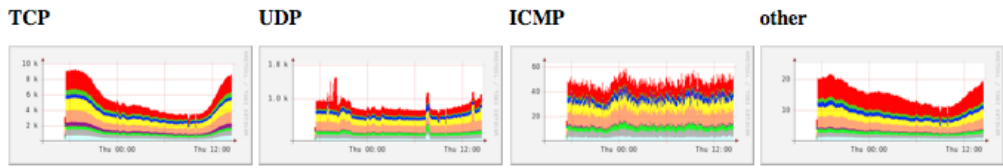
Same thing, only different

Absolute Traffic at Indiana GigaPop, 2004-07 (Cumulative View)

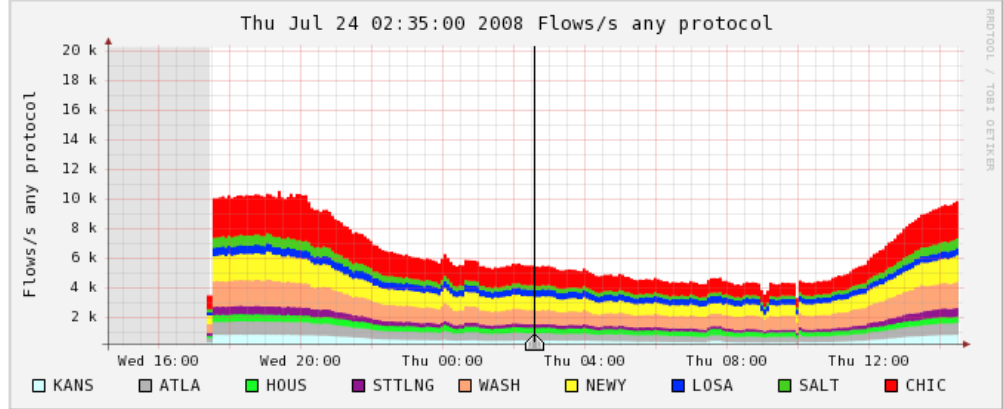


NFSEN

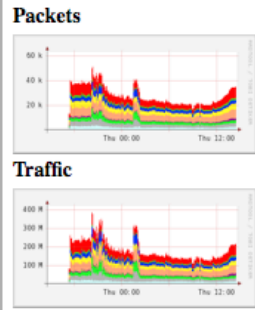
Profile: live



Profileinfo:
 Type: live
 Max: 976.6 GB
 Exp: never
 Start: Jul 23 2008 - 17:30 UTC
 End: Jul 24 2008 - 14:35 UTC



t_start 2008-07-24-02-35
 t_end 2008-07-24-02-35



Select | Display: 1 day | Lin Scale | Stacked Graph | Log Scale | Line Graph

▼ Statistics timeslot Jul 24 2008 - 02:35

Channel:	Flows:					Packets:					Traffic:				
	all:	tcp:	udp:	icmp:	other:	all:	tcp:	udp:	icmp:	other:	all:	tcp:	udp:	icmp:	other:
<input checked="" type="checkbox"/> CHIC	1.3 k/s	1.1 k/s	160.0 /s	11.0 /s	6.0 /s	5.5 k/s	4.9 k/s	521.1 /s	12.1 /s	92.5 /s	37.9 Mb/s	35.0 Mb/s	2.3 Mb/s	9.5 kb/s	584.2 kb/s
<input checked="" type="checkbox"/> SALT	293.1 /s	260.7 /s	30.2 /s	1.4 /s	0.8 /s	1.1 k/s	988.4 /s	114.2 /s	1.8 /s	29.0 /s	6.8 Mb/s	6.1 Mb/s	507.9 kb/s	1.3 kb/s	201.0 kb/s
<input checked="" type="checkbox"/> LOSA	451.2 /s	379.9 /s	66.5 /s	3.6 /s	1.0 /s	2.2 k/s	1.8 k/s	339.2 /s	10.3 /s	35.8 /s	14.4 Mb/s	12.4 Mb/s	1.7 Mb/s	6.0 kb/s	240.7 kb/s
<input checked="" type="checkbox"/> NEWY	930.3 /s	788.0 /s	133.5 /s	6.8 /s	2.0 /s	3.3 k/s	2.8 k/s	399.4 /s	7.3 /s	53.9 /s	19.5 Mb/s	17.6 Mb/s	1.6 Mb/s	5.8 kb/s	315.6 kb/s
<input checked="" type="checkbox"/> WASH	942.8 /s	764.3 /s	166.6 /s	9.6 /s	2.3 /s	4.9 k/s	4.2 k/s	608.8 /s	10.1 /s	77.2 /s	33.5 Mb/s	29.7 Mb/s	3.3 Mb/s	8.3 kb/s	512.5 kb/s
<input checked="" type="checkbox"/> STTLNG	241.9 /s	227.4 /s	13.6 /s	0.7 /s	0.3 /s	576.4 /s	510.1 /s	61.9 /s	0.8 /s	3.5 /s	3.1 Mb/s	2.8 Mb/s	355.0 kb/s	504.2 b/s	25.8 kb/s
<input checked="" type="checkbox"/> HOUS	370.1 /s	304.0 /s	61.6 /s	3.6 /s	0.9 /s	2.1 k/s	1.7 k/s	323.0 /s	9.7 /s	34.5 /s	14.4 Mb/s	12.5 Mb/s	1.6 Mb/s	5.4 kb/s	238.0 kb/s


REN-ISAC Threat Monitoring

Mozilla Firefox

http://www.ren-isac.net/cgi-bin/monitoring/Internet2TGa.cgi

netflow

Enter Salon Apple .Mac YMail comics geek RHN News ANML MeFi OED weather RSS Net development phone bt misc watches



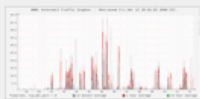
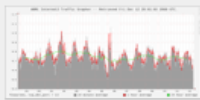




REN-ISAC
Research and Education Networking
Information Sharing and Analysis Center

Public Home

- [Member Log In](#)
- [About REN-ISAC](#)
- [Contact Us](#)
- [Membership](#)
- [Docs and Links](#)
- [Monitoring](#)
- [24x7 Watch Desk](#)
- [Advisory Groups](#)
- [Policies](#)
- [Contributors](#)

These port traffic graphs are generated from aggregate Internet2 netflow data, sampled at 1:100.

port	protocol	service	monitoring links	best practice	vulnerability / exploit / notes
1	tcp	.		.	
1	udp	.		.	
8	tcp	.		.	
13	tcp	Daytime Protocol (RFC-867)		.	
13	udp	Daytime Protocol (RFC-867)		.	
20	tcp	ftp-data		.	

PROBLEMS WITH THOSE approaches

Can only “see” ports you’ve decided to see.

Need to manually intervene to set up what to watch

Forest through the trees

Evolution of visualization techniques

- Text-Based
- 2D visualization of old text information
 - I.e. ACID interface to SNORT
- 3D visualization?

Other 3-D Visualizers

VIAssist (Commercial)

Nvision

DAVIX (Similar to gCube, but more extensive)

UniVis

www.vizsec.org (clearinghouse of network visualizers)

gCube

Nascent effort to develop a *useful & lightweight* 3D modeling capability.

Not an original idea (Shakespeare had it first)

- Saw a similar tool at SC2003
 - Steve Lau (LBNL) Cube of potential doom
 - BRO project (<http://www.icir.org/vern/bro.html>)

Nor the end of the line (see DAVIX, VIAssist, etc.)

What is it?

Simple & Basic version is 3D view of “flow” activity

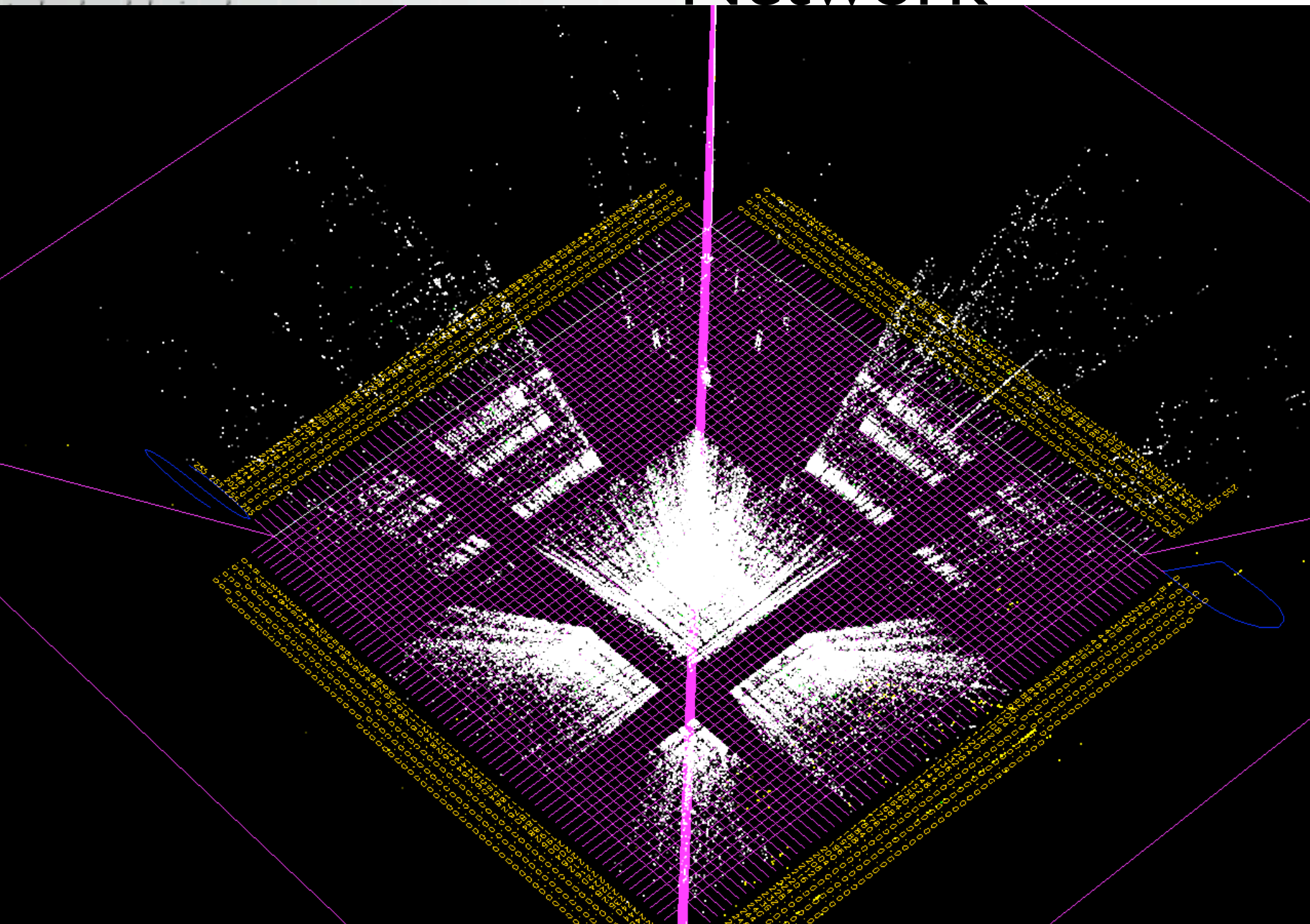
- X/Z axis determined by source/destination IP
- Y axis determined by port number
 - Usually destination port number

Where does it get its input?

Three possible inputs:

- Direct NETFLOW feed
- Archived NETFLOW (files)
- PCAP view of local network

Network

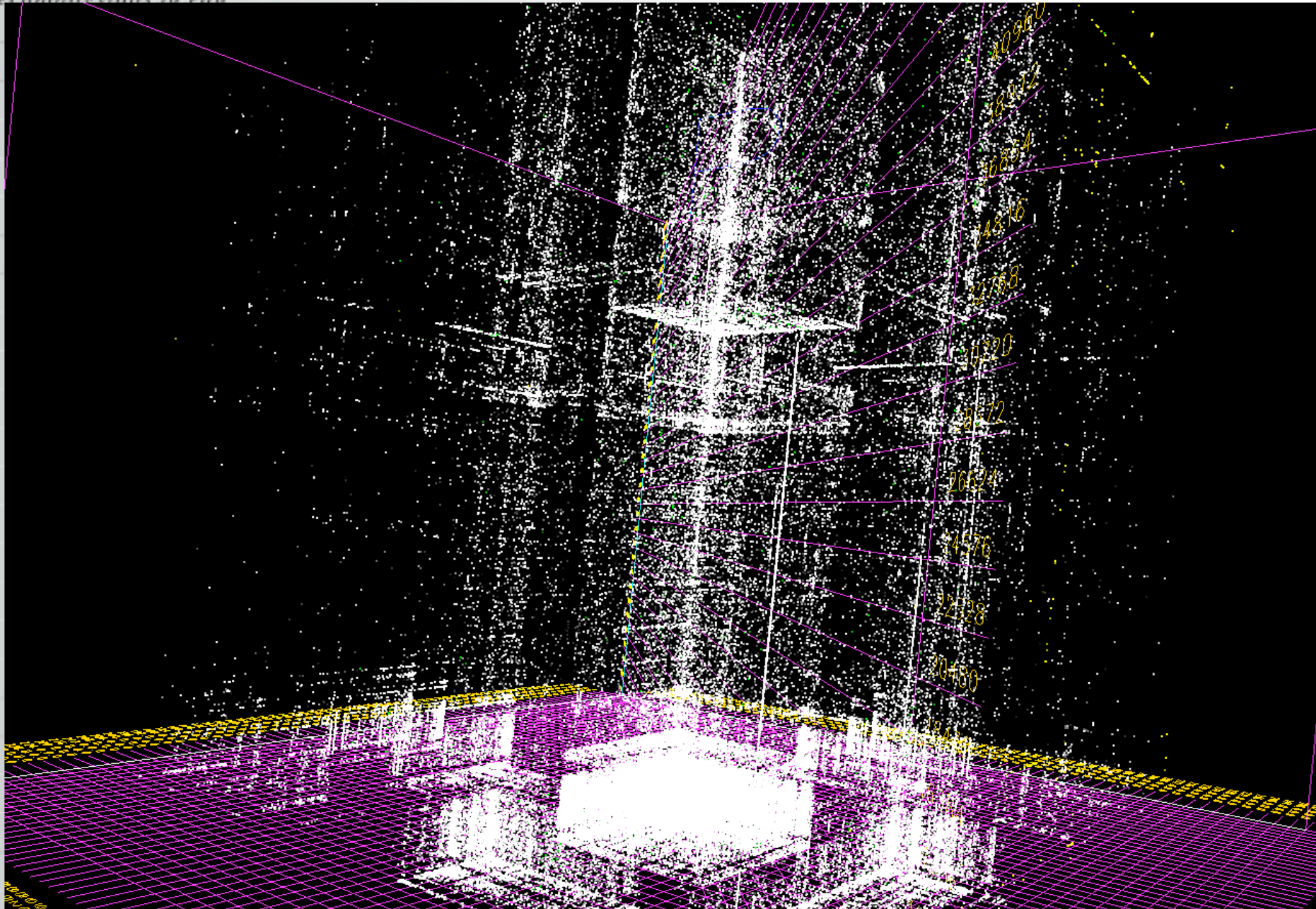


What are we seeing?

Entire IPv4 address space (all 4 billion possible source and destination addresses)

- Blank areas represent portions of IP space not allocated to Abilene-connected institutions
- Allocation pattern is interesting
 - 4 “towers”
 - Early remnants of class-A allocations
 - MIT, .gov, etc.

Side view of I2



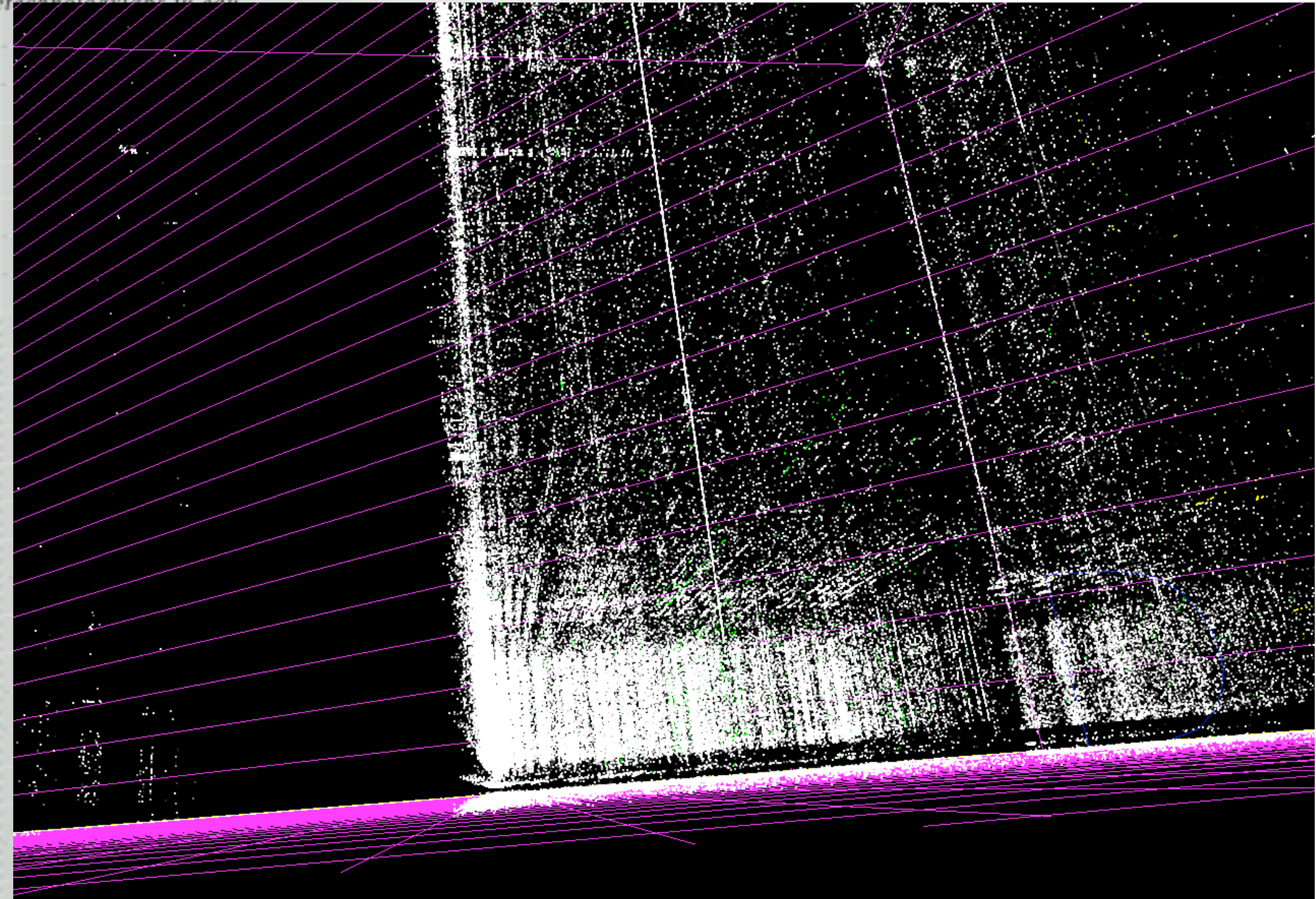
What structures are visible?

Special “floors”

- 32K port allocation floor
- 40K port allocation floor
- Density of port allocations at lower levels

An apparent port scan!

The low level

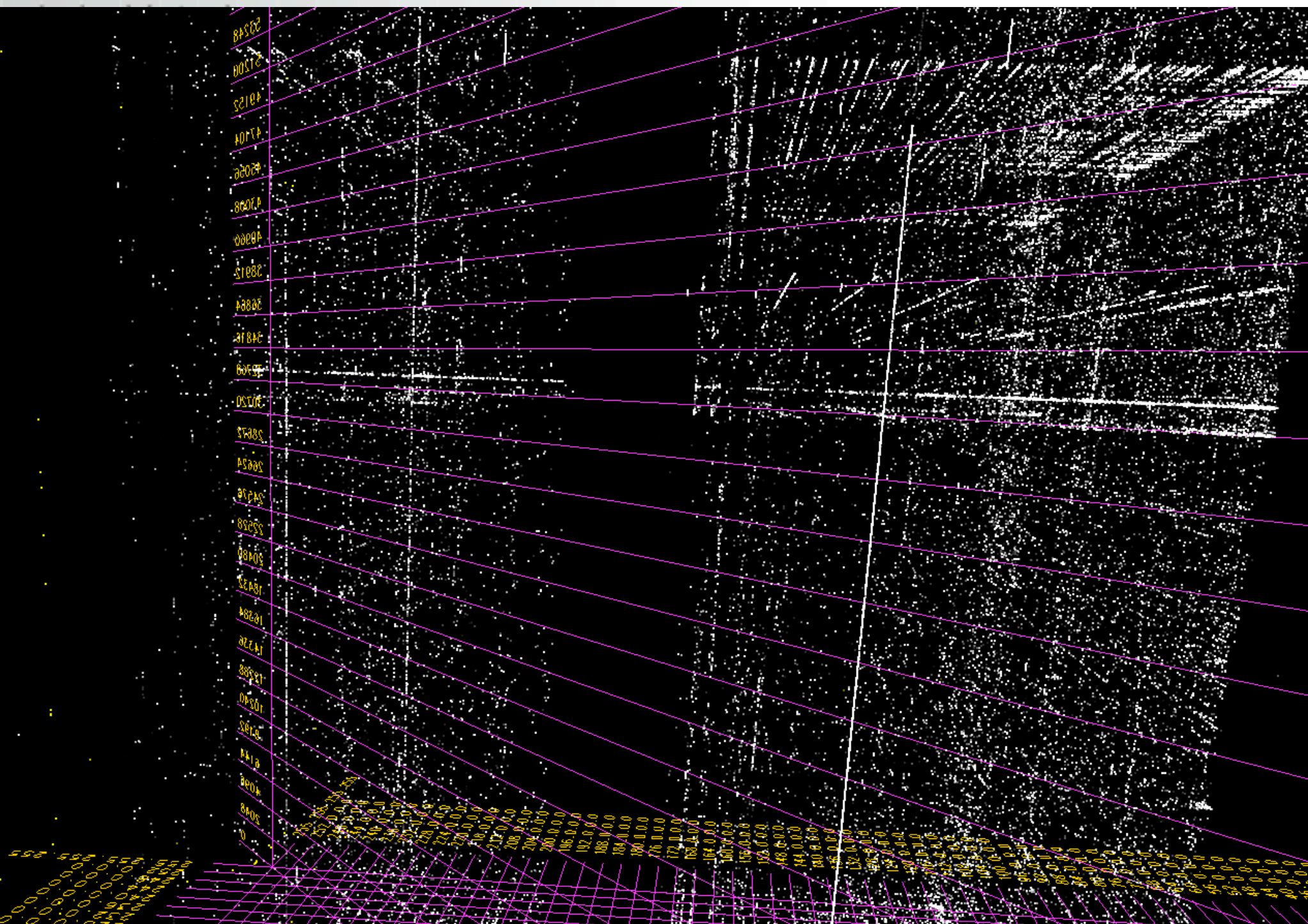


Visualizing DDoS with gCube

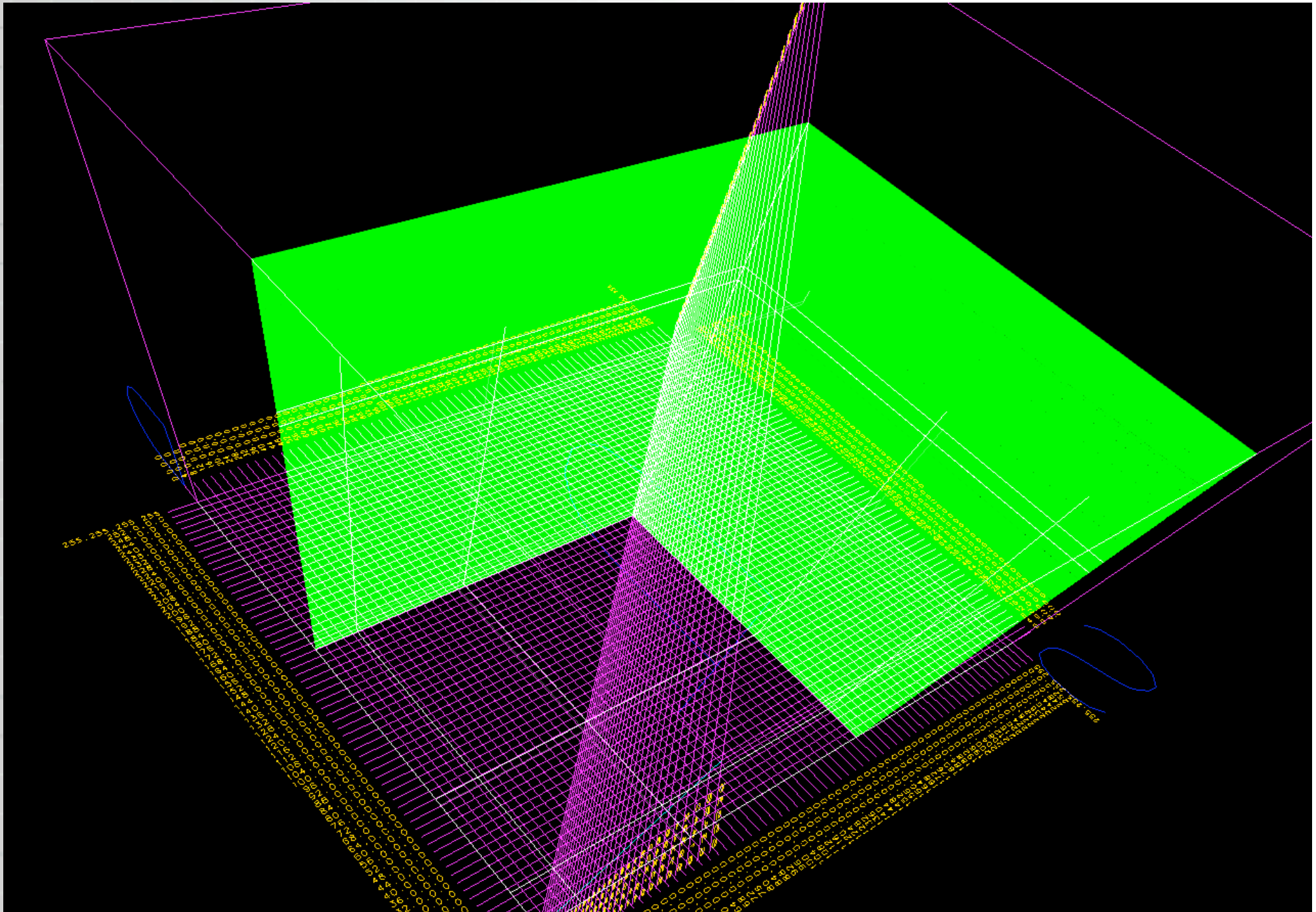
Eventual hope is to develop gCube into a DDoS visualization tool

- Particularly good at detecting
 - Port Scans
 - Host Scans
 - Scans into “abnormal” IP space
 - I.e. Slammer type stuff
 - Rate/bandwidth anomalies

Simple case, portscan

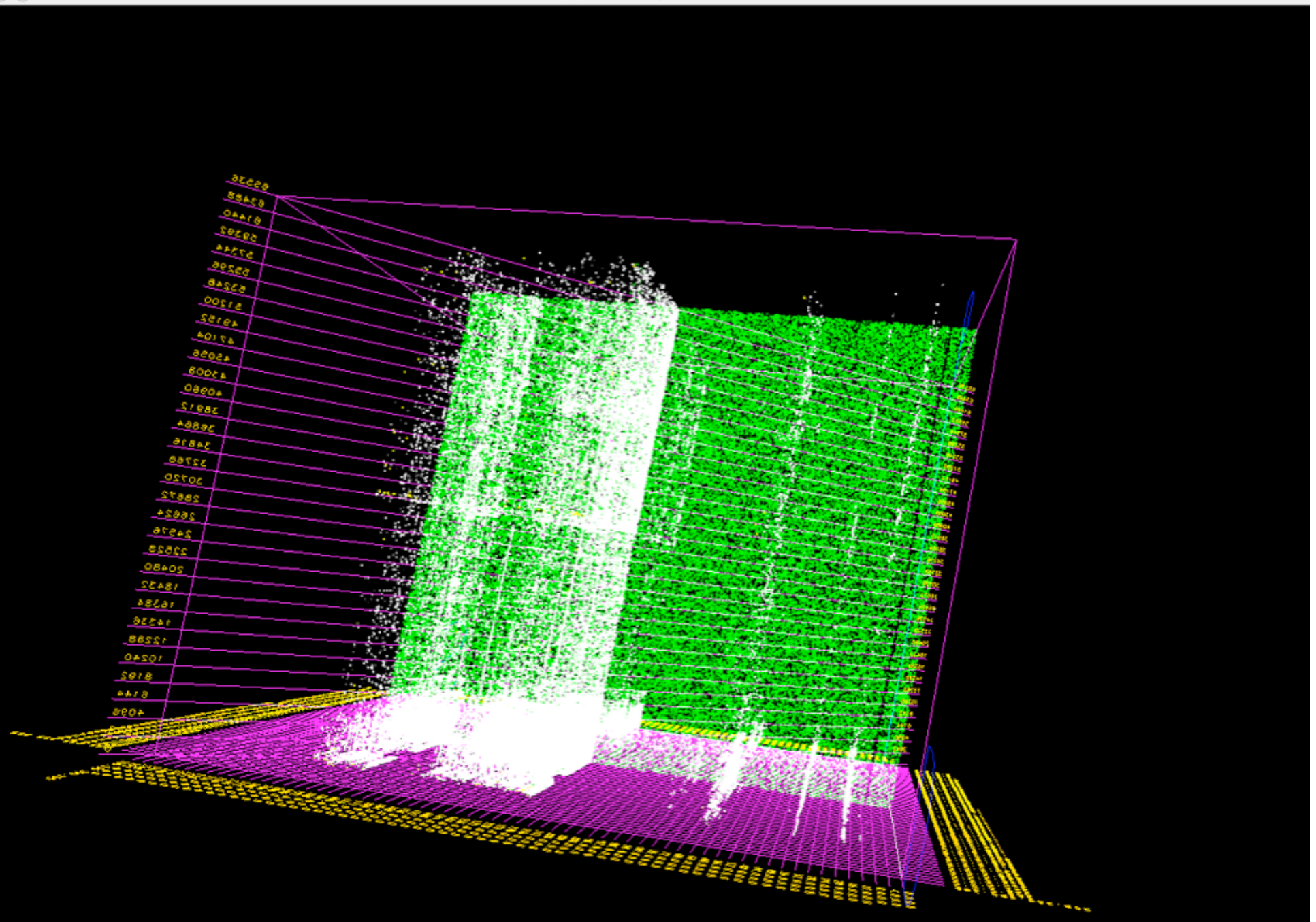


Simulated Portscan



DDoS in the real world

ANML : Internet Cube V1.1



What is that?

January 14th, 2003, ~2-3PM EST

Port scan of a destination address

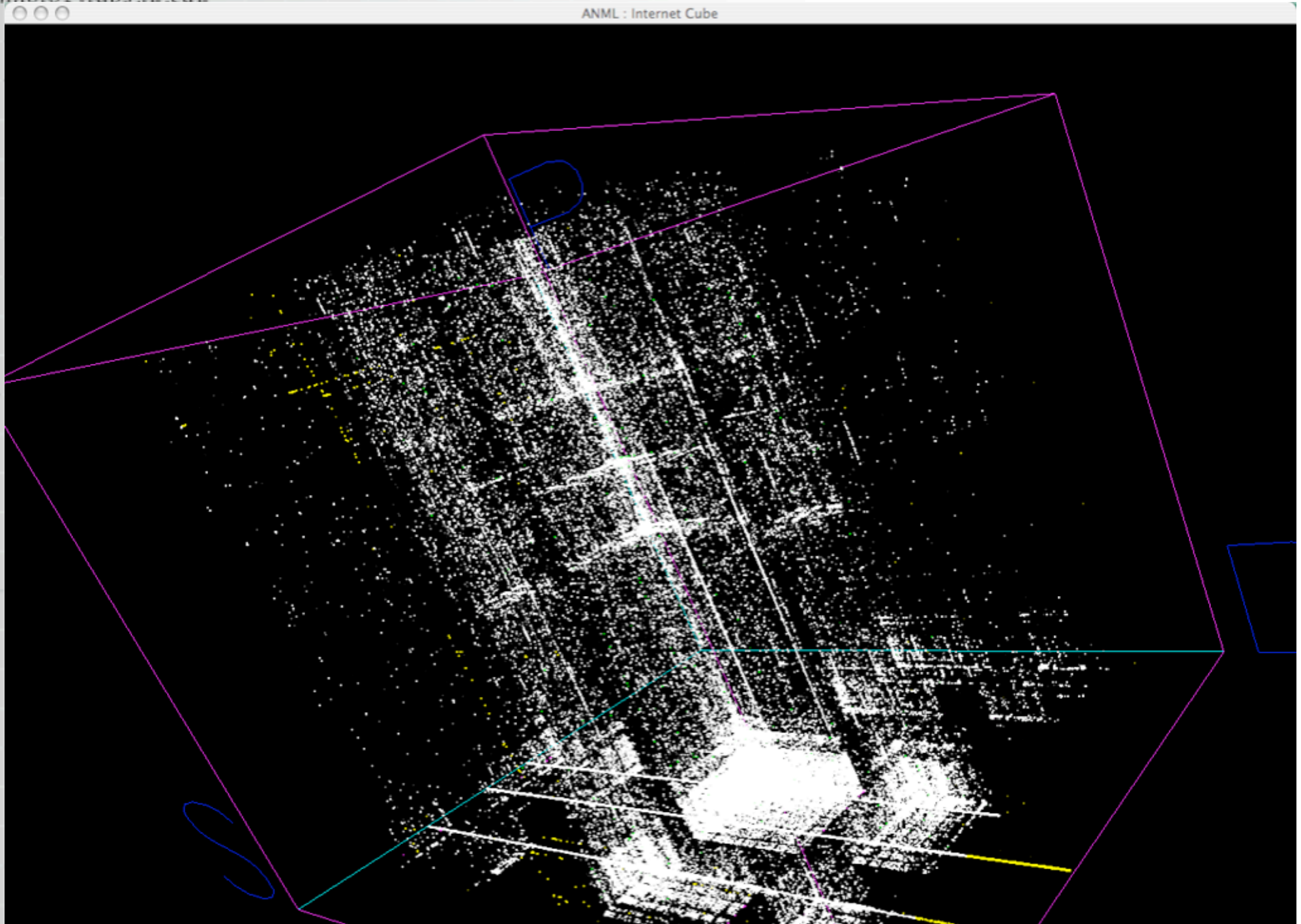
Spoofed source IP addresses

- Distributed equally through IP space

Had been preceded by apparent “experiments” earlier in the day and earlier in the week (Jan 5th)

- Experiments used only a single or few test ports

Experiments



Note

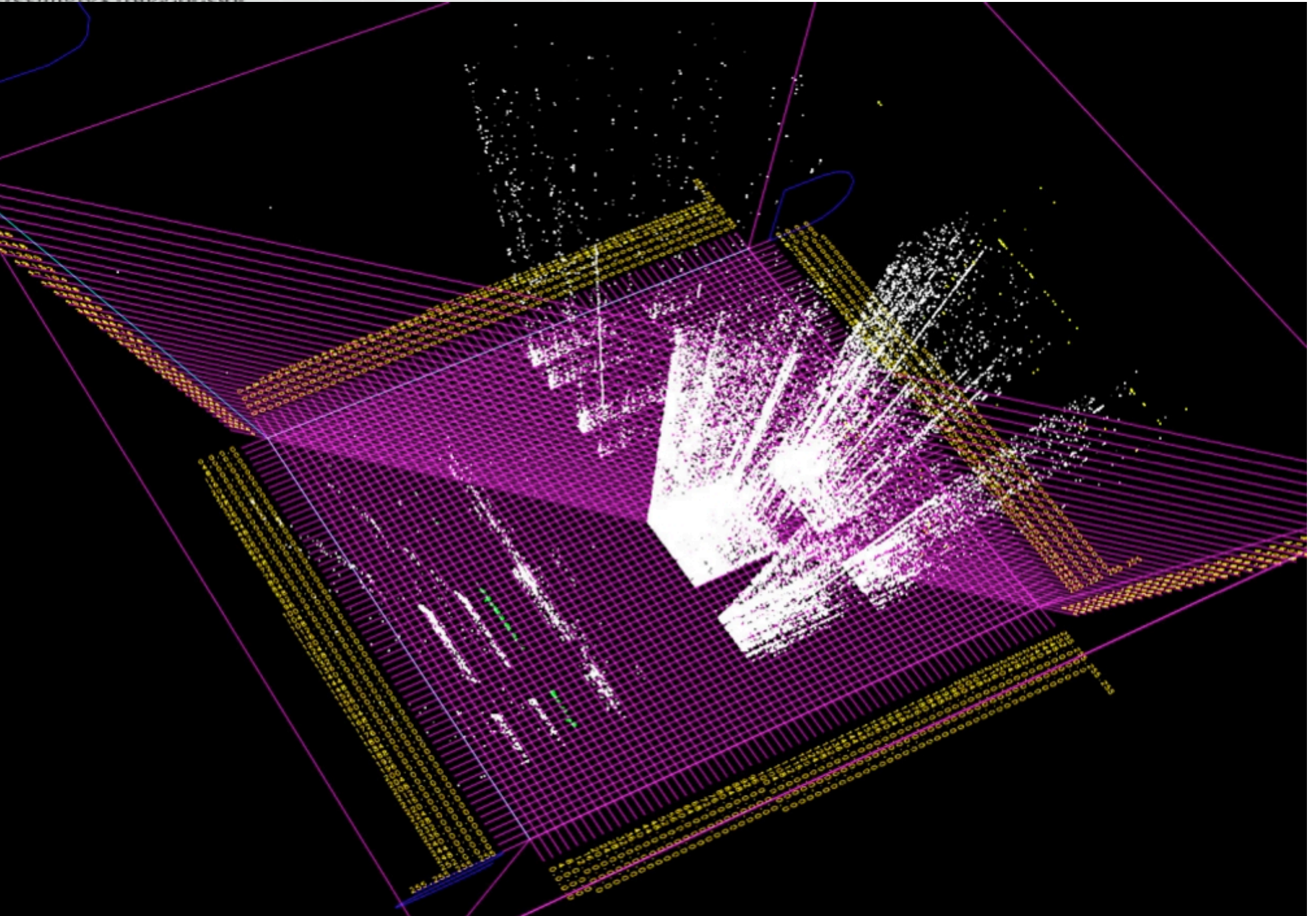
Attacks to three separate IPs/closely clustered groups of IPs

Spoofed source IPs

- But possibly from as many as three different organizations

At least one real source appeared to be suppressing sources from the multicast space

Backscatter



Backscatter

