



# Flow Visualization Using MS-Excel

## Visualization for the Common Man

Presented by Lee Rock and Jay Brown  
US-CERT Analysts  
Einstein Program



# Background

- US-CERT Mission
- Einstein Program
  - > Large volumes of traffic
  - > Architecture limitations
- Proactive vs. Reactive analysis
- Slow application certification process



# Pro's and Con's

- Pro's:
  - Visualization allows for rapid analysis
  - Patterns are easy to identify
  - Flexibility in analysis
  - Most enterprises have MS Office (Excel)
- Con's:
  - Excel plotting engine is limited
  - Max of 65K records (recommend  $\leq 50K$ )
  - Data must be imported and formatted
  - Memory management is an issue



# Data Preparation Steps

- Data Pull
- Data Reduction
- Importing Data
- Data Formatting
- Sample analysis slides





# Data Pull

Analysts have several options when trying to pull interesting datasets. Several methods we find useful are:

- Collecting data during non-business hours
  - Reduces traffic from users; helps expose automated sessions
- Search for outbound traffic only
  - Reduces noise from scanning, etc.
- Filtering for packets with the PSH/ACK flags set in the initial flags field
  - Focuses the traffic on sessions where data is actually transferred
- Filtering for packets with the SYN flag set in the initial flags field
  - Focuses on sessions initiated by your organization
- Limit traffic to records under 5K bytes
  - Most cyclical sessions (beaconing) happen in this range

Traffic should be refined to provide the best possible dataset for analysts to work with.



# Data Reduction

To further enhance the concentration of suspicious data, analysts should:

- Remove replies from servers (responses to inbound server requests)
  - Looking for genuine outbound traffic
- Remove loud, common talkers (instant messenger, web crawlers, etc)
  - Reduces the noise, especially in web traffic
- “Whitelists” and “blacklists” are helpful for filtering

This is an iterative approach – Analyze, Research, Remove.



# Importing Data

Data is imported from a pipe delimited text file

sIP	dIP	sPort	dPort	pro	packets	bytes	flags	sTime	dur	eT
10.147.82.96	10.130.166.158	80	5516	6	11	8258	FS PA	2007/10/29T15:07:38.807	0.290	2007/10/29T15:07:39.1
10.140.165.218	10.54.98.176	80	5705	6	86	120321	FS PA	2007/10/29T15:07:40.875	0.552	2007/10/29T15:07:41.1
10.95.46.146	10.34.134.191	80	5705	6	1	40	A	2007/10/29T15:07:42.473	0.000	2007/10/29T15:07:42.1
10.94.132.147	10.168.141.231	80	49094	6	10	7348	FS PA	2007/10/29T15:16:19.666	0.825	2007/10/29T15:16:20.1
10.226.143.219	10.162.254.83	80	49297	6	33	43588	FS PA	2007/10/29T15:16:23.020	0.498	2007/10/29T15:16:23.1
172.25.4.165	10.161.142.75	80	47356	6	10	7047	FS PA	2007/10/29T15:28:21.421	0.859	2007/10/29T15:28:22.1
10.120.9.241	10.36.140.83	80	47489	6	7	1839	FS PA	2007/10/29T15:28:23.285	0.326	2007/10/29T15:28:23.1
10.192.192.130	10.124.26.9	5516	80	6	9	945	FS PA	2007/10/29T15:07:38.771	0.292	2007/10/29T15:07:39.1
10.8.58.141	10.254.147.27	5705	80	6	55	3589	FS PA	2007/10/29T15:07:40.843	1.591	2007/10/29T15:07:42.1
10.215.49.170	10.30.5.168	49094	80	6	8	969	FS PA	2007/10/29T15:16:19.638	0.825	2007/10/29T15:16:20.1
10.207.158.173	10.15.150.60	49297	80	6	22	1666	FS PA	2007/10/29T15:16:23.004	0.488	2007/10/29T15:16:23.1
192.168.45.69	10.88.159.210	47356	80	6	9	1796	FS PA	2007/10/29T15:28:21.381	0.874	2007/10/29T15:28:22.1
10.227.193.146	10.237.117.172	47489	80	6	7	888	FS PA	2007/10/29T15:28:23.261	0.313	2007/10/29T15:28:23.1
10.115.234.230	10.144.241.122	80	24503	6	10	6735	FS PA	2007/10/29T16:01:28.698	0.220	2007/10/29T16:01:28.1
10.52.224.171	10.232.170.176	80	24601	6	7	1475	FS PA	2007/10/29T16:01:29.421	0.236	2007/10/29T16:01:29.1
10.144.199.78	10.208.138.229	80	64021	6	10	6437	FS PA	2007/10/29T16:09:08.791	0.161	2007/10/29T16:09:08.1
10.9.152.19	10.233.152.178	80	64124	6	7	1310	FS PA	2007/10/29T16:09:09.883	0.247	2007/10/29T16:09:10.1
10.116.235.116	10.27.192.234	80	64021	6	1	40	A	2007/10/29T16:09:08.951	0.000	2007/10/29T16:09:08.1
10.0.158.212	10.131.10.198	80	40079	6	11	6378	FS PA	2007/10/29T16:16:40.586	0.247	2007/10/29T16:16:40.1
10.40.145.167	10.229.195.82	80	40167	6	15	15095	FS PA	2007/10/29T16:16:41.718	0.317	2007/10/29T16:16:42.1
10.40.157.25	10.12.36.164	80	18275	6	10	6242	FS PA	2007/10/29T16:24:32.546	0.235	2007/10/29T16:24:32.1
10.33.232.60	10.224.241.212	80	18385	6	21	21877	FS PA	2007/10/29T16:24:34.100	0.409	2007/10/29T16:24:34.1
10.75.204.191	10.52.57.127	24503	80	6	8	1439	FS PA	2007/10/29T16:01:28.654	0.223	2007/10/29T16:01:28.1
10.6.83.30	10.218.84.41	24601	80	6	7	888	FS PA	2007/10/29T16:01:29.393	0.225	2007/10/29T16:01:29.1
10.123.207.187	10.211.245.126	64021	80	6	9	2544	FS PA	2007/10/29T16:09:08.762	0.153	2007/10/29T16:09:08.1
10.239.204.27	10.140.242.63	64021	80	6	1	40	R	2007/10/29T16:09:08.951	0.000	2007/10/29T16:09:08.1





# Data Formatting

Columns within the spreadsheet should be aligned to each field of the flows, Einstein data is formatted to encompass:

- Source IP
- Destination IP
- Source Port
- Destination Port
- Protocol
- Packets
- Bytes
- Flags
- Start Time
- Duration
- End Time
- Sensor
- Type
- Initial Flags

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	slIP	dlIP	sPort	dPort	pro	packets	bytes	flags	sTime	dur	eTime	sensor	type	initialFlag
2	x.x.x.x	y.y.y.y	52611	22	6	2569	3385741	FS PA	2007/10/17T00:10:40.722	9.5	2007/10/17T00:10:50.222	X	out	S
3	x.x.x.x	y.y.y.y	7774	22	6	136	10750	PA	2007/10/17T00:08:28.293	1795.691	2007/10/17T00:38:23.984	X	out	A
4	x.x.x.x	y.y.y.y	7774	22	6	106	9046	PA	2007/10/17T00:38:36.714	1800.05	2007/10/17T01:08:36.764	X	out	PA
5	x.x.x.x	y.y.y.y	1688	22	6	1	41	A	2007/10/17T00:01:18.787	0	2007/10/17T00:01:18.787	X	out	A
6	x.x.x.x	y.y.y.y	1688	22	6	1	41	A	2007/10/17T00:06:18.690	0	2007/10/17T00:06:18.690	X	out	A
7	x.x.x.x	y.y.y.y	1688	22	6	1	41	A	2007/10/17T00:11:18.598	0	2007/10/17T00:11:18.598	X	out	A
8	x.x.x.x	y.y.y.y	1688	22	6	1	41	A	2007/10/17T00:16:18.514	0	2007/10/17T00:16:18.514	X	out	A
9	x.x.x.x	y.y.y.y	1688	22	6	1	41	A	2007/10/17T00:21:18.433	0	2007/10/17T00:21:18.433	X	out	A
10	x.x.x.x	y.y.y.y	1688	22	6	1	41	A	2007/10/17T00:26:18.349	0	2007/10/17T00:26:18.349	X	out	A
11	x.x.x.x	y.y.y.y	1688	22	6	1	41	A	2007/10/17T00:31:18.257	0	2007/10/17T00:31:18.257	X	out	A
12	x.x.x.x	y.y.y.y	1688	22	6	1	41	A	2007/10/17T00:36:18.164	0	2007/10/17T00:36:18.164	X	out	A
13	x.x.x.x	y.y.y.y	1281	22	6	956	40630	PA	2007/10/17T00:09:32.281	1798.994	2007/10/17T00:39:31.275	X	out	A
14	x.x.x.x	y.y.y.y	1688	22	6	1	41	A	2007/10/17T00:41:18.068	0	2007/10/17T00:41:18.068	X	out	A
15	x.x.x.x	y.y.y.y	1688	22	6	1	41	A	2007/10/17T00:46:17.971	0	2007/10/17T00:46:17.971	X	out	A





# Data Formatting Cont.

US-CERT analysts use two methods to format the Einstein time fields into a format that is able to be plotted:

A: Use the - - legacy-timestamps switch to place the time in a MM/DD/YYYY HH:MM:SS format from the default MM/DD/YYYYTHH:MM:SS.MMM

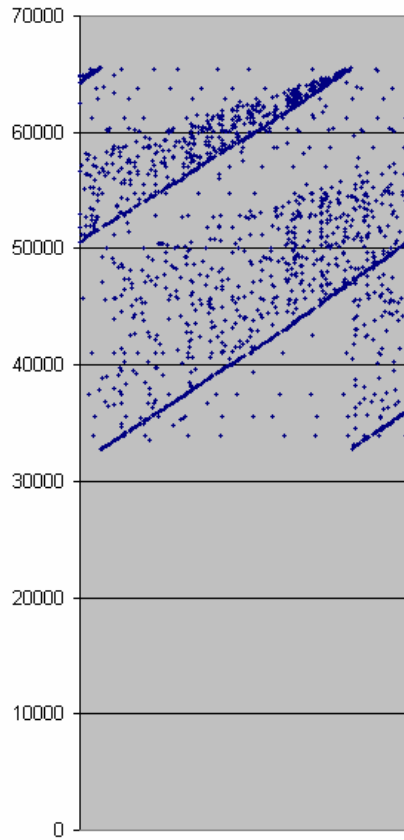
B: Utilize the replace function in excel to remove the milliseconds from the time and replace the T placeholder with a space:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	sIP	dIP	sPort	dPort	pro	packets	bytes	flags		x.x.x.x	y.y.y.y	52611	22	6	b	2569	3385/41	FS PA
2	x.x.x.x	y.y.y.y	52611		22	6	2569	3385741	FS PA	x.x.x.x								
3	x.x.x.x								PA	x.x.x.x								
4	x.x.x.x								PA	x.x.x.x								
5	x.x.x.x								PA	x.x.x.x								
6	x.x.x.x								A	x.x.x.x								
7	x.x.x.x								A	x.x.x.x								
8	x.x.x.x								A	x.x.x.x								
9	x.x.x.x								A	x.x.x.x								
10	x.x.x.x								A	x.x.x.x								
11	x.x.x.x								A	x.x.x.x								
12	x.x.x.x								A	x.x.x.x								
13	x.x.x.x								PA	x.x.x.x								
14	x.x.x.x								PA	x.x.x.x								
15	x.x.x.x								A	x.x.x.x	y.y.y.y	1688	22	6		1	41	A
16	x.x.x.x	y.y.y.y	1688	22	6	1	41	A	x.x.x.x	y.y.y.y	1688	22	6		1	41	A	A
17	x.x.x.x								A	x.x.x.x	y.y.y.y	1688	22	6		1	41	A

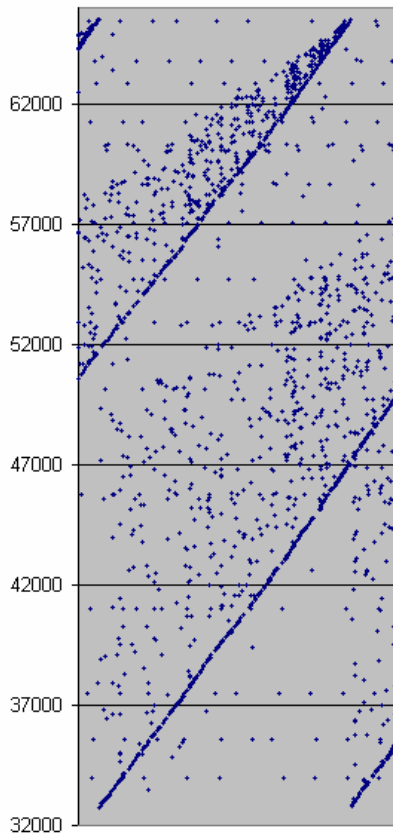


# Analysis Workflow

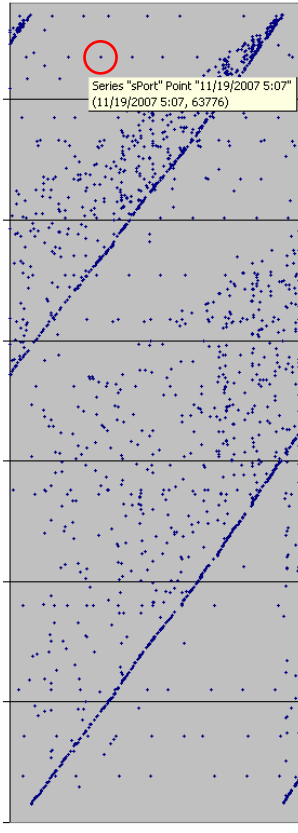
Plot



Zoom



Highlight



AutoFilter

C	D	E	F	G	H	I	J
sTime	sPort	dp	r	packe	byt	flags	sTime
11/19/2007 5:00	63776	443	6	3	164	PA	11/19/2007
11/19/2007 5:03	63776	443	6	10	851	PA	11/19/2007
11/19/2007 5:05	63776	443	6	2	102	PA	11/19/2007
11/19/2007 5:07	63776	443	6	1	62	PA	11/19/2007

**Custom AutoFilter**

Show rows where:

sPort  
equals

And  Or

Use ? to represent any single character  
Use \* to represent any series of characters

OK Cancel

11/19/2007 5:42	63776	443	6	1	62	PA	11/19/2007
11/19/2007 5:44	63776	443	6	1	62	PA	11/19/2007
11/19/2007 5:46	63776	443	6	2	102	PA	11/19/2007
11/19/2007 5:47	63776	443	6	1	62	PA	11/19/2007
11/19/2007 5:49	63776	443	6	1	62	PA	11/19/2007
11/19/2007 5:51	63776	443	6	2	102	PA	11/19/2007
11/19/2007 5:53	63776	443	6	1	62	PA	11/19/2007
11/19/2007 5:55	63776	443	6	3	164	PA	11/19/2007
11/19/2007 5:58	63776	443	6	1	62	PA	11/19/2007
11/19/2007 6:00	63776	443	6	2	102	PA	11/19/2007
11/19/2007 6:02	63776	443	6	3	322	PA	11/19/2007
11/19/2007 6:04	63776	443	6	6	262	PA	11/19/2007
11/19/2007 6:06	63776	443	6	6	284	PA	11/19/2007
11/19/2007 6:09	63776	443	6	1	62	PA	11/19/2007
11/19/2007 6:11	63776	443	6	3	142	PA	11/19/2007
11/19/2007 6:13	63776	443	6	1	62	PA	11/19/2007
11/19/2007 6:14	63776	443	6	4	204	PA	11/19/2007
11/19/2007 6:19	63776	443	6	1	62	PA	11/19/2007
11/19/2007 6:20	63776	443	6	2	102	PA	11/19/2007
11/19/2007 6:22	63776	443	6	3	142	PA	11/19/2007
11/19/2007 6:24	63776	443	6	1	62	PA	11/19/2007
11/19/2007 6:26	63776	443	6	2	102	PA	11/19/2007
11/19/2007 6:28	63776	443	6	1	62	PA	11/19/2007
11/19/2007 6:30	63776	443	6	2	102	PA	11/19/2007



# Plot

Creating charts from the selected data, allows for quick pattern identification

**Chart Wizard - Step 1 of 4 - Chart Type**

Standard Types Custom Types

Chart type:

- Column
- Bar
- Line
- XY (Scatter)
- Area
- Doughnut
- Radar
- Surface
- Bubble

Chart sub-type:

Scatter, Compare

Press and H

**Chart Wizard - Step 2 of 4 - Chart Source Data**

Data Range Series

Data range: `=SSL!$C$1:$D$57298`

Series in:

- Rows
- Columns

**Chart Wizard - Step 3 of 4 - Chart Options**

Titles Axes Gridlines Legend Data Labels

Chart title: intIP

Value (X) axis:

Value (Y) axis:

Second category (X) axis:

Second value (Y) axis:

intIP

4000000000  
3500000000  
3000000000  
2500000000  
2000000000  
1500000000  
1000000000  
500000000  
0

11/16/20 11/17/20 11/17/20 11/18/20 11/18/20  
07 07 0:00 07 07 0:00 07  
12:00 12:00 12:00

Cancel < Back Next > Finish

**Chart Wizard - Step 4 of 4 - Chart Location**

Place chart:

- As new sheet: intIP
- As object in: SSL

Cancel < Back Next > Finish

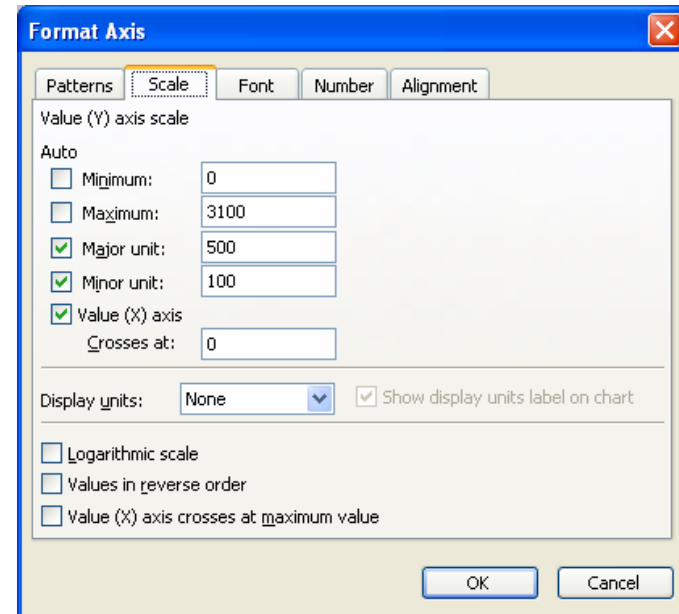
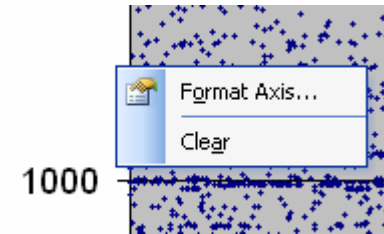




# Zoom

You can “zoom” in to specific data points, by changing the scale of the axis

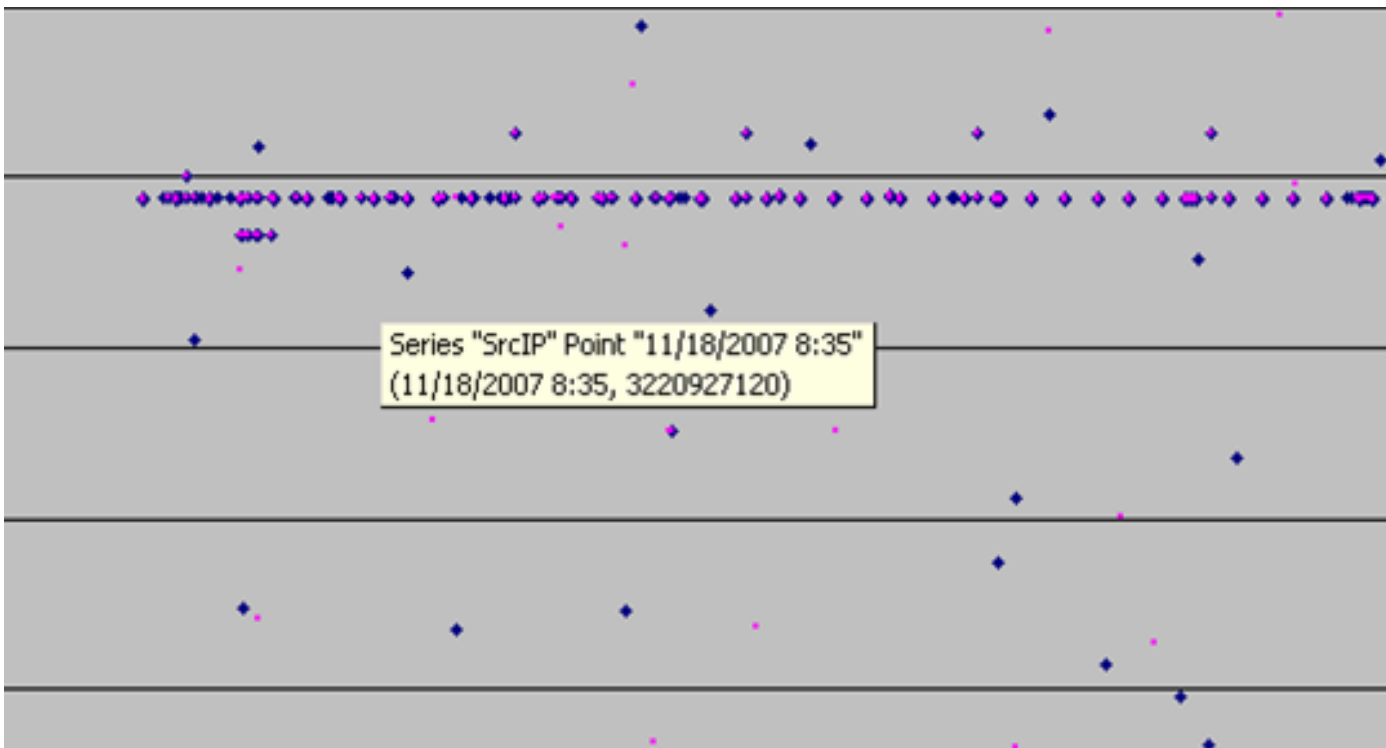
- Right click on the axis
- Select “Format Axis”
- Click on the “Scale” tab
- Adjust scale as desired
- Works for both axis
- Remember to remove





# Highlight

By hovering over a data point in the series an analyst can locate the point in the rest of the records by filtering for the displayed information





# AutoFilter

Method A – Drop down list:

Select the desired value from the drop down list

C	D	E	F	G
	sPort	dPort	pro	packe
	Sort Ascending	80	6	1
	Sort Descending	80	6	1
	(All)	80	6	1
	(Top 10...)	80	6	1
	(Custom...)	80	6	1
	49578	80	6	1
	50338	80	6	1
	52161	80	6	1
	53023	80	6	1
	54590	80	6	1
	54726	80	6	1
	55337	80	6	1
	56549	80	6	1
	56989	80	6	1
	59674	80	6	1
	60551	80	6	1
	62425	80	6	1
	64602	80	6	1
	65233	80	6	1

Method B – Custom Filter:

Select data by using Excel's built in boolean logic search functions

D	E	F	G	H	I
sPort	dPort	pro	packe	flags	
54726	80	6	1	PA	11/
50338	80	6	1	PA	11/
53023	80	6	1	PA	11/
56989	80	6	1	PA	11/
59674	80	6	1	PA	11/
60551	80	6	1	PA	11/

**Custom AutoFilter**

Show rows where:

dPort

equals 56989

does not equal

is greater than

is greater than or equal to

is less than

is less than or equal to

Use ? to represent any single character  
Use \* to represent any series of characters

OK Cancel



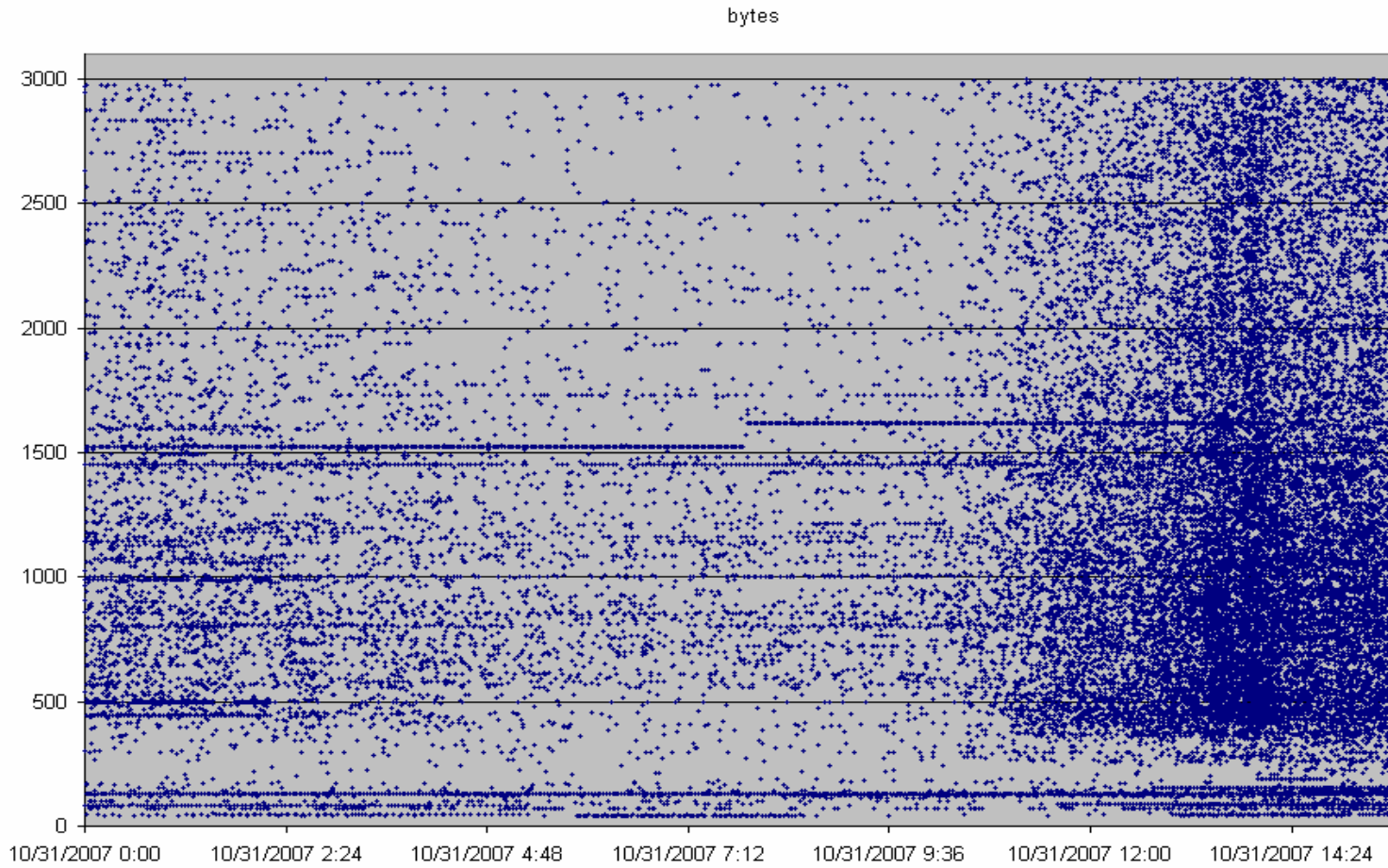


# Sample Analysis Slides

- Scatter Plot Analysis
  - Byte Based Patterns
  - Duration Based Patterns
  - sPort vs. dPort Patterns
  - IP Based Patterns
  - Application Pattern



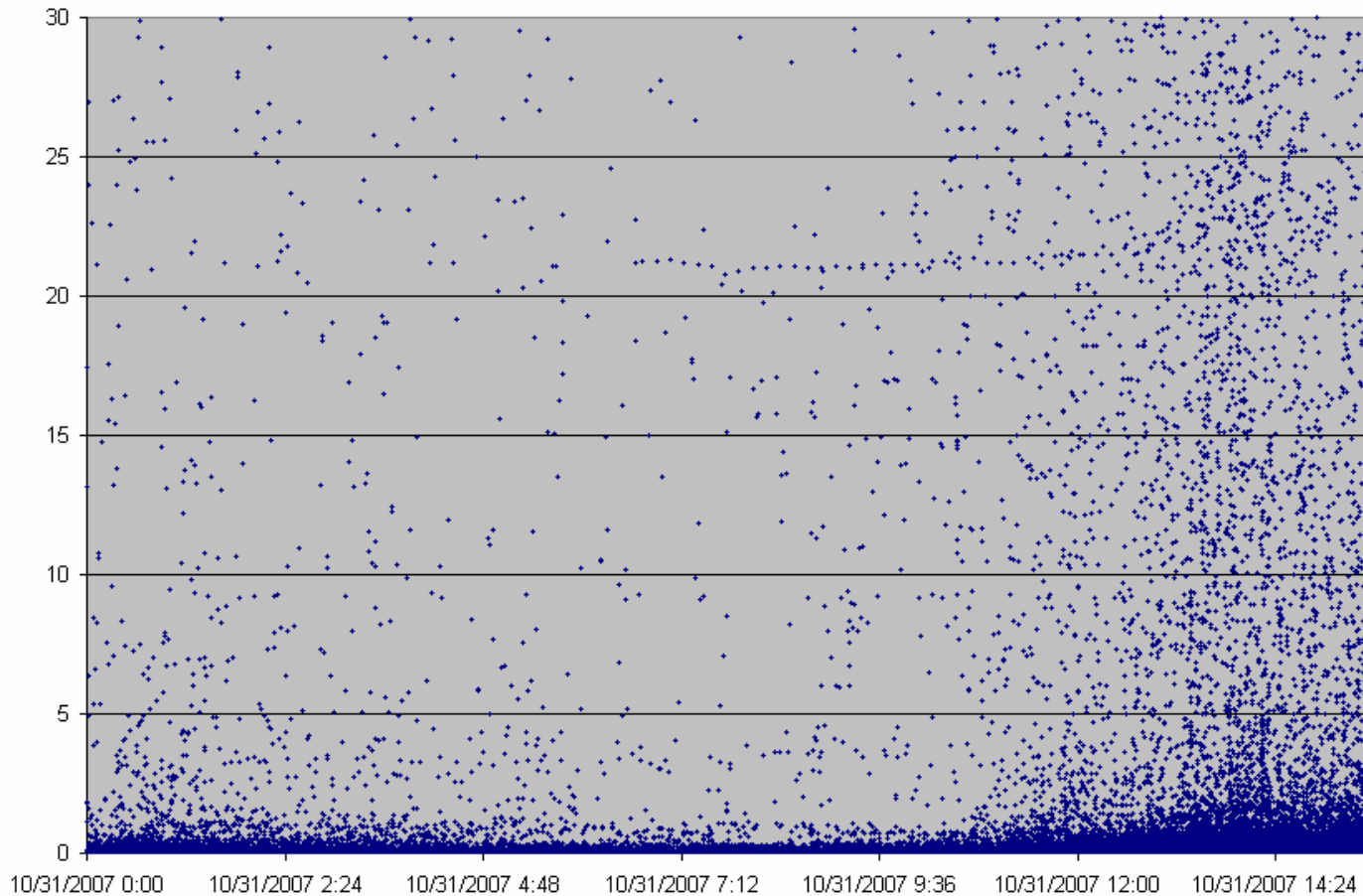
# Byte Based Patterns





# Duration Based Patterns

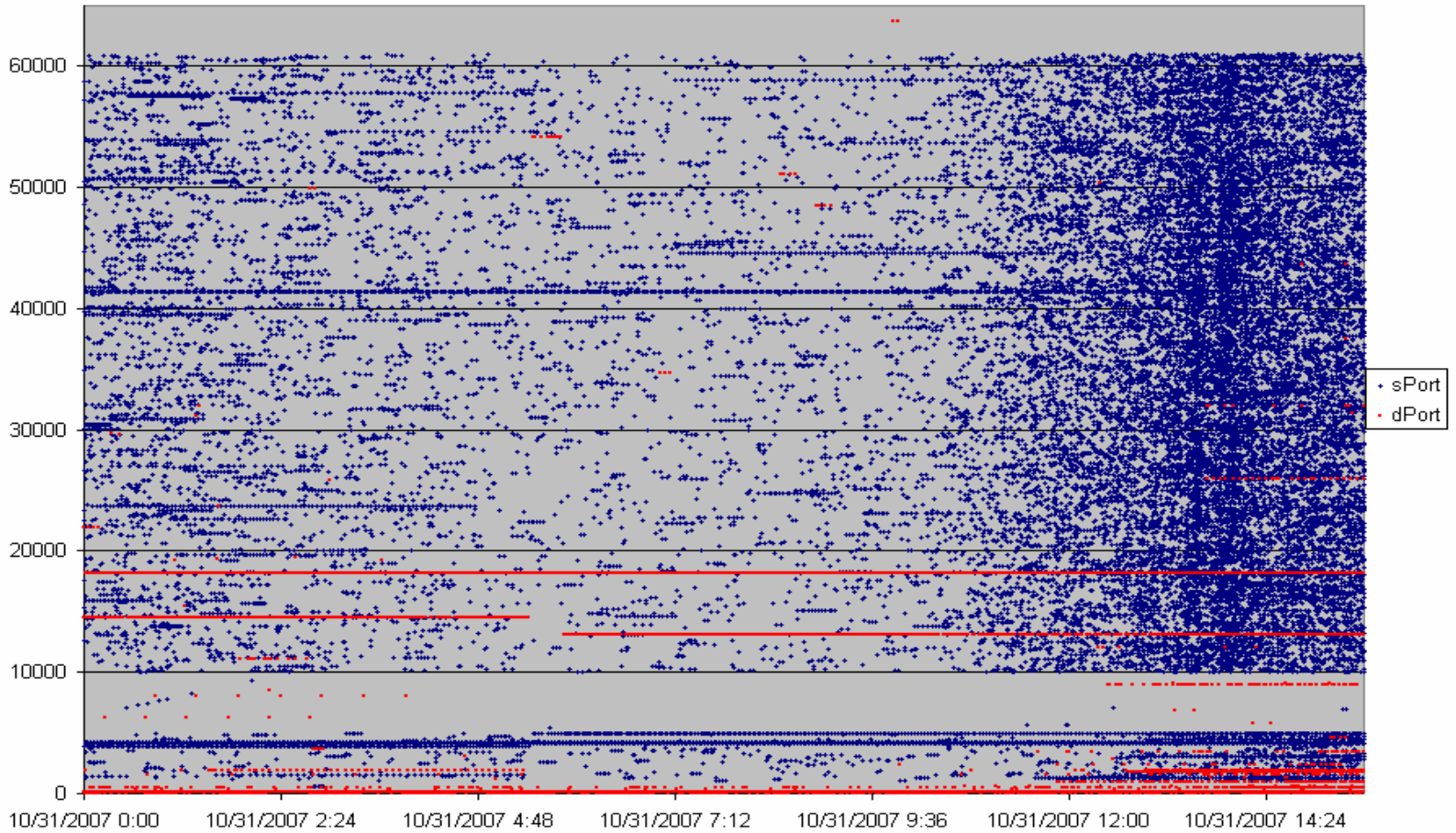
dur





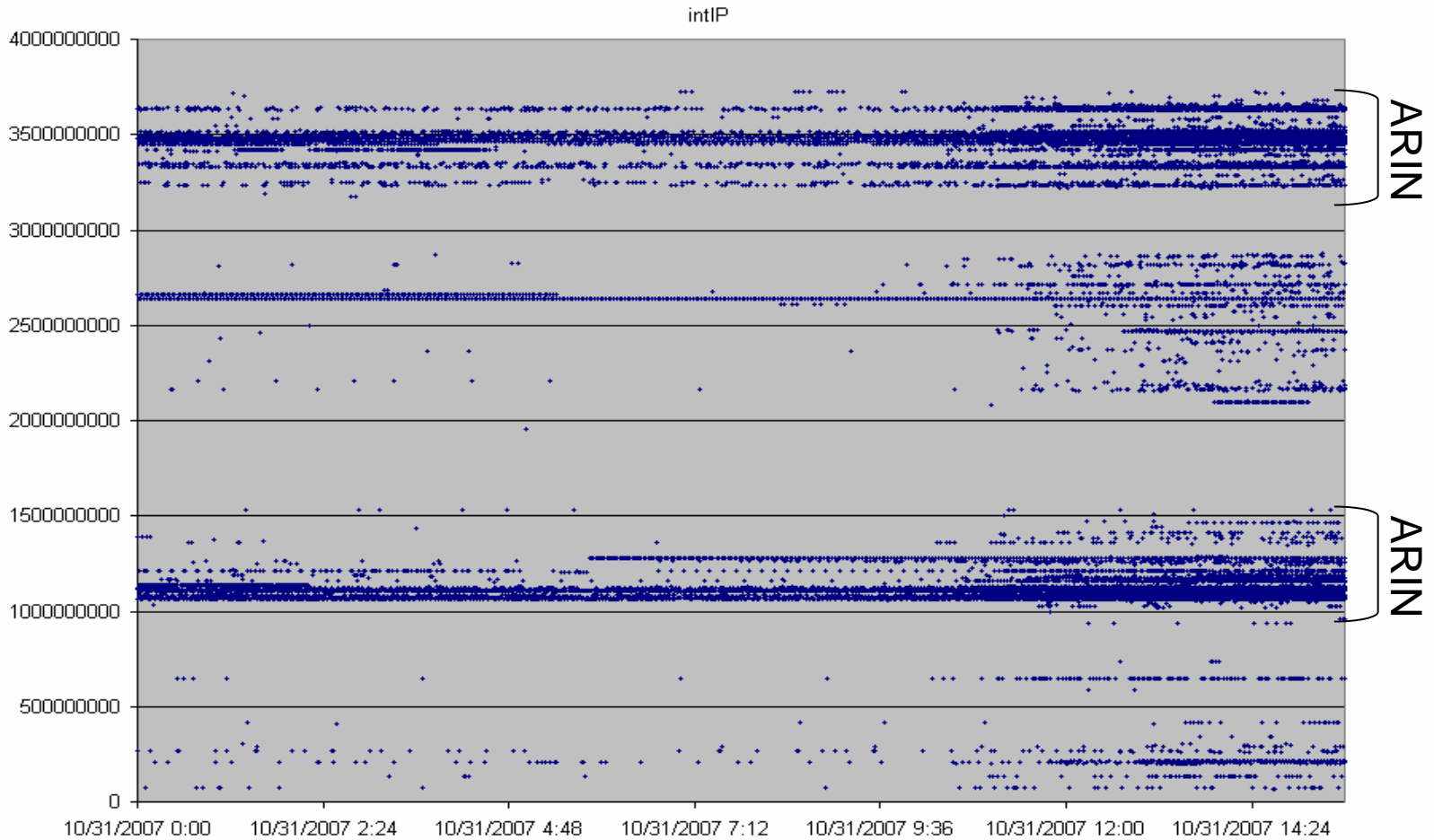


# sPort vs. dPort

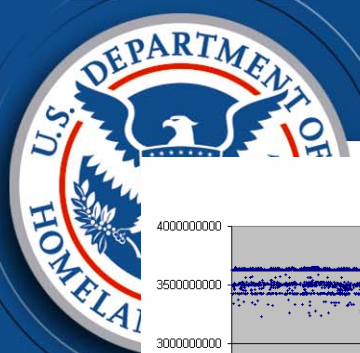




# IP Integer Patterns

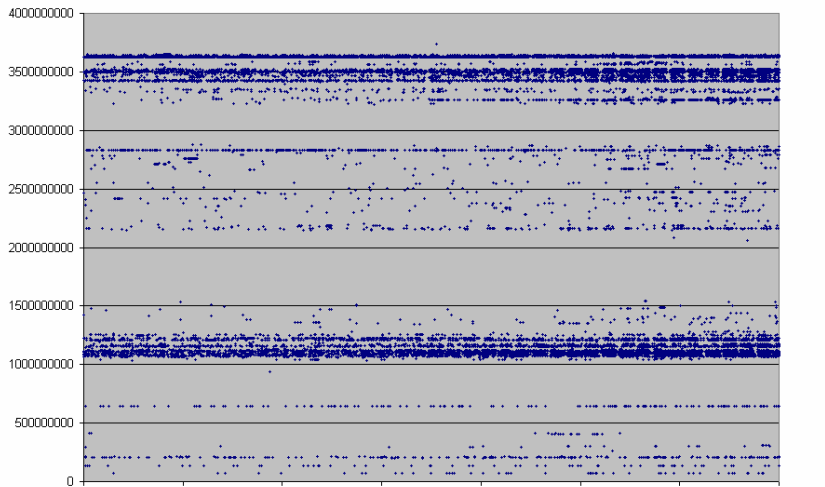






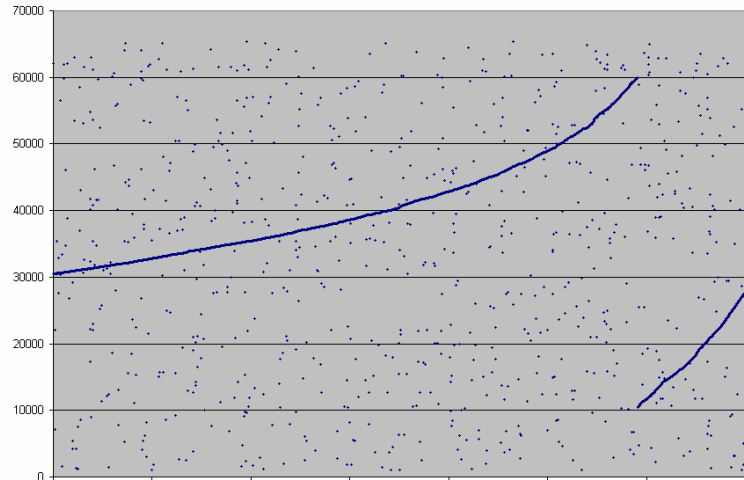
# Comprehensive View

intIP



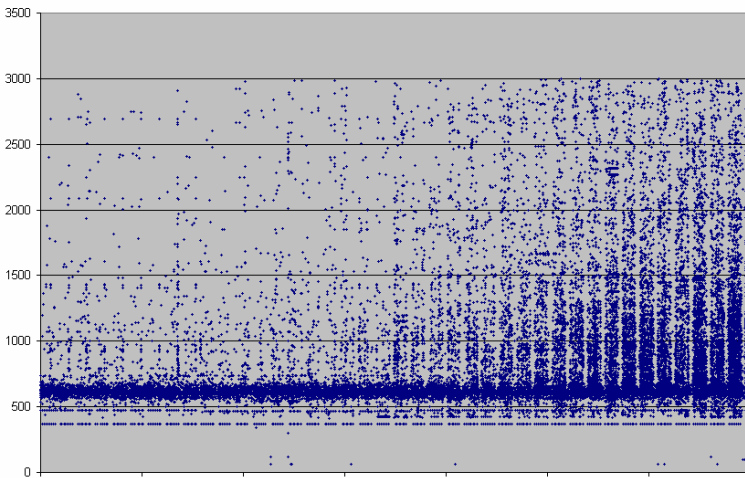
11/20/2007 10:04 11/20/2007 10:33 11/20/2007 11:02 11/20/2007 11:31 11/20/2007 12:00 11/20/2007 12:28 11/20/2007 12:57 11/20/2007 13:26

sPort



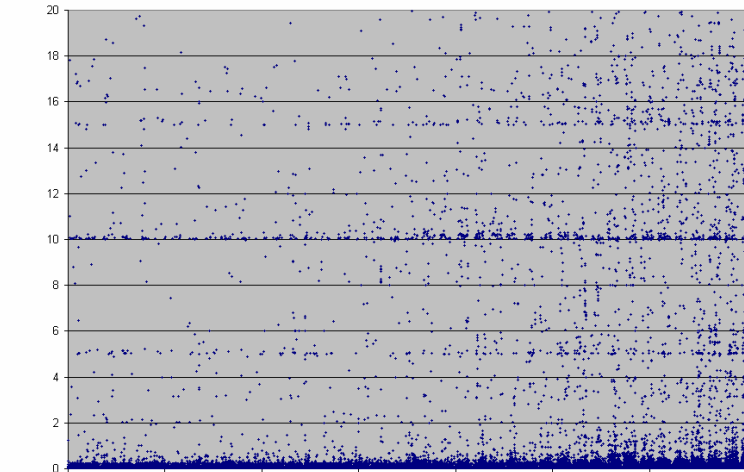
11/20/2007 10:04 11/20/2007 10:33 11/20/2007 11:02 11/20/2007 11:31 11/20/2007 12:00 11/20/2007 12:28 11/20/2007 12:57 11/20/2007 13:26

bytes



11/20/2007 10:04 11/20/2007 10:33 11/20/2007 11:02 11/20/2007 11:31 11/20/2007 12:00 11/20/2007 12:28 11/20/2007 12:57 11/20/2007 13:26

dur

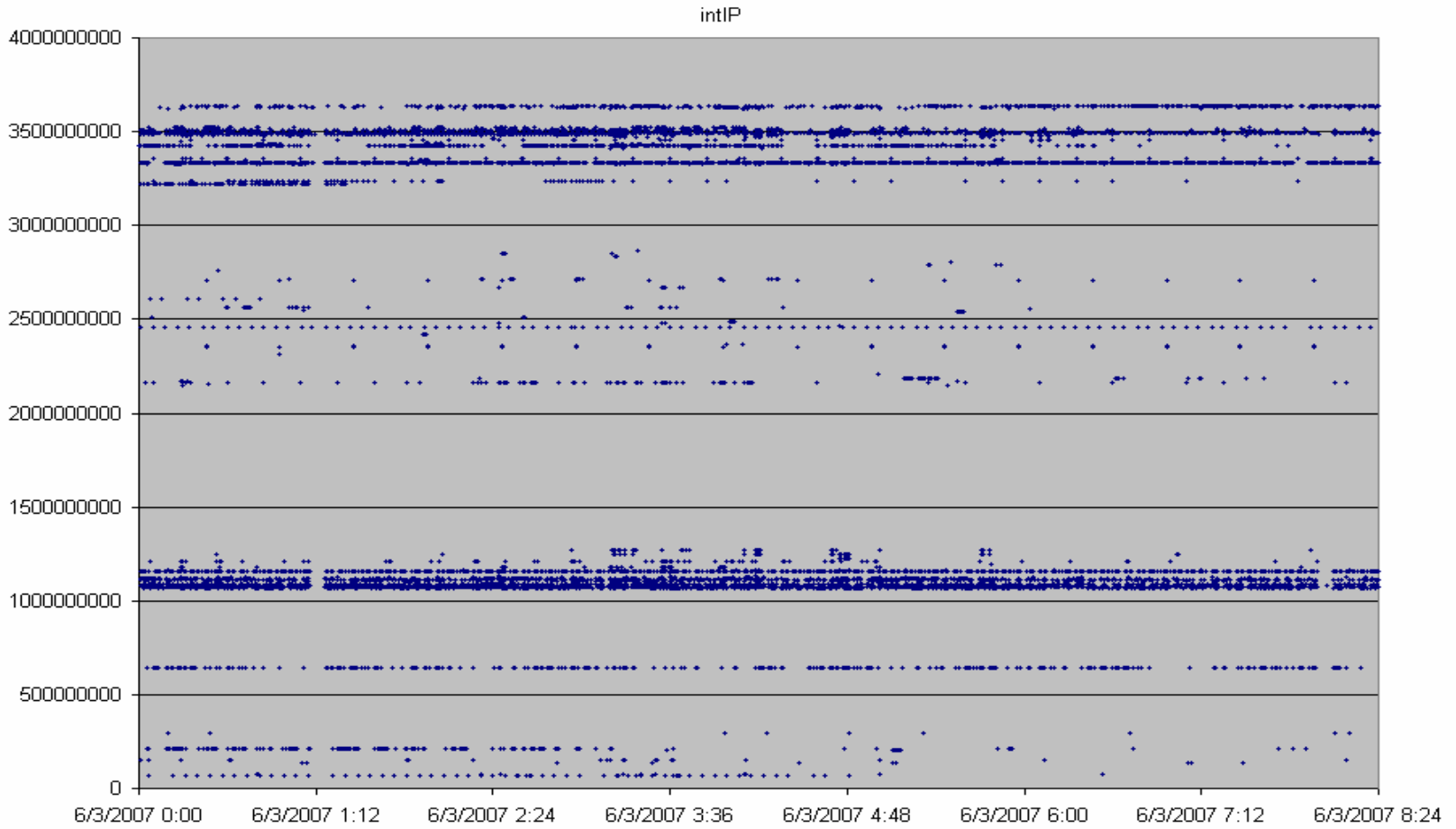


11/20/2007 10:04 11/20/2007 10:33 11/20/2007 11:02 11/20/2007 11:31 11/20/2007 12:00 11/20/2007 12:28 11/20/2007 12:57 11/20/2007 13:26



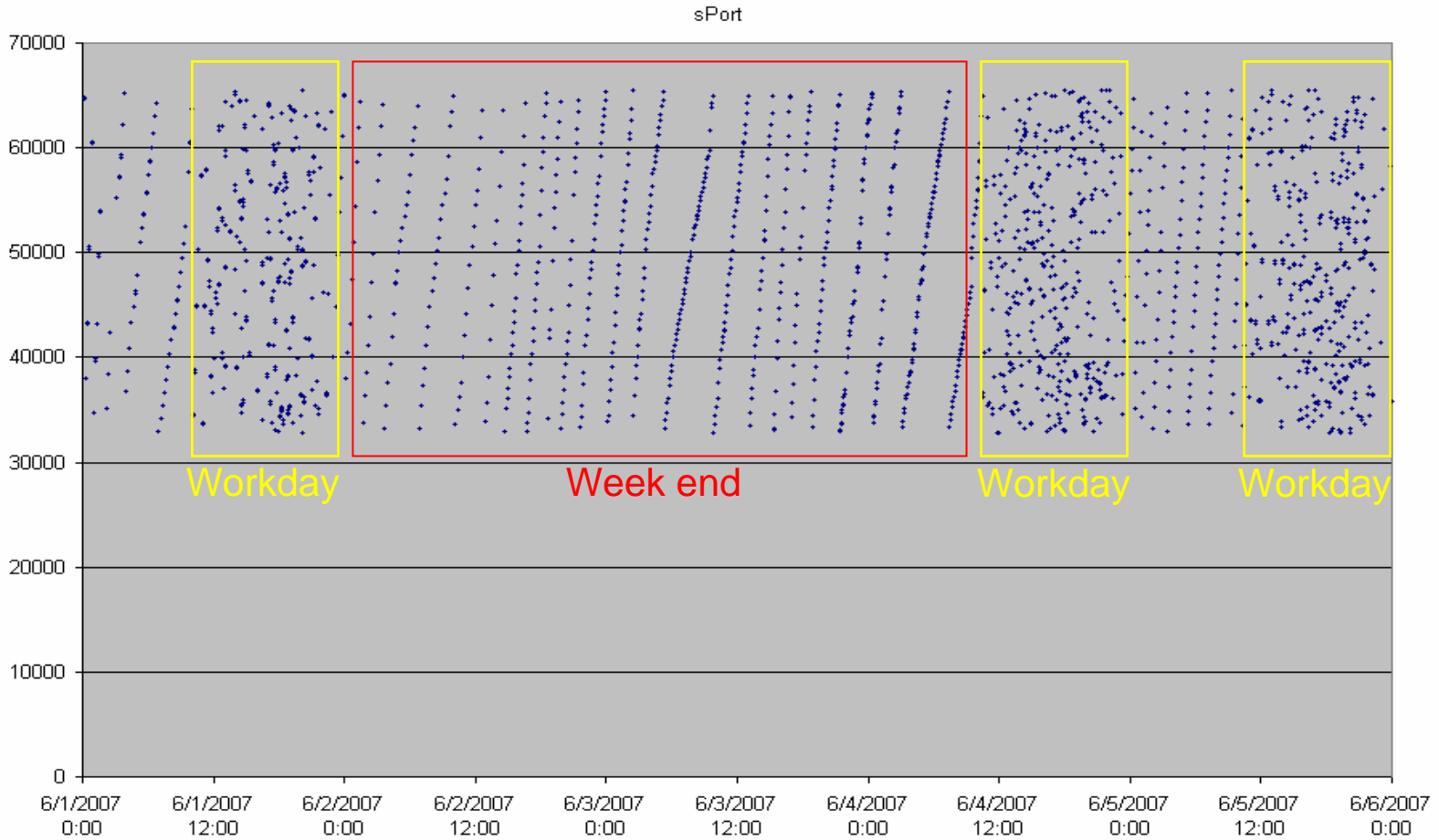


# Case Study





# Multi-day View





## Case Study Conclusion

After notifying the agency in question, the machines that were generating this traffic were found and forensically examined. The malware turned out to be a keystroke logger that posted data to a specific website and retrieved commands embedded on the same site. Prior to this incident, there was no malware associated with this site.





# Additional Analysis

Determining application patterns

- Identifying specific applications

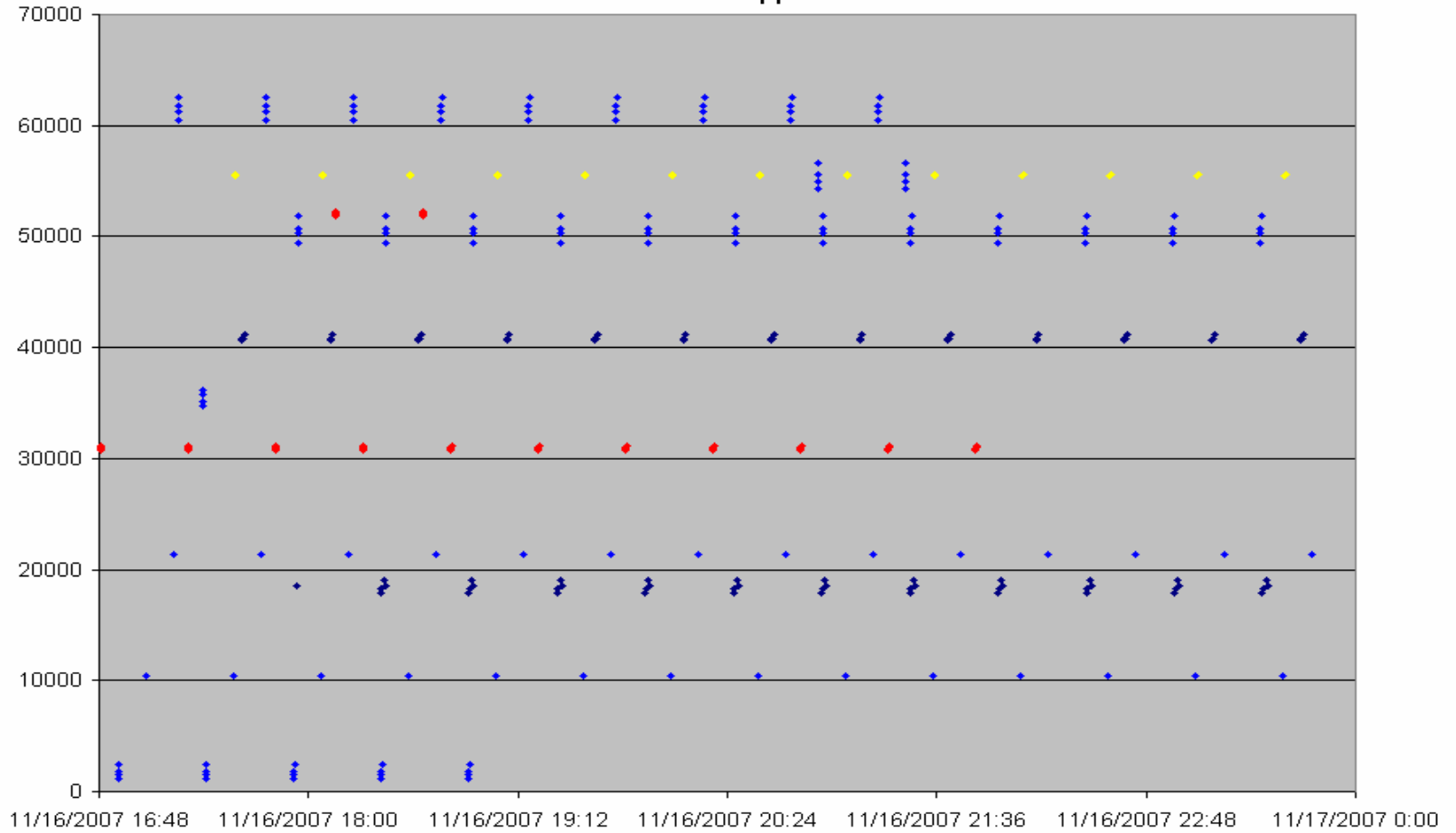
Working with gateway traffic

- Structured gateway
- Proxy gateway
- Gateway mannerisms



# Application Patterns

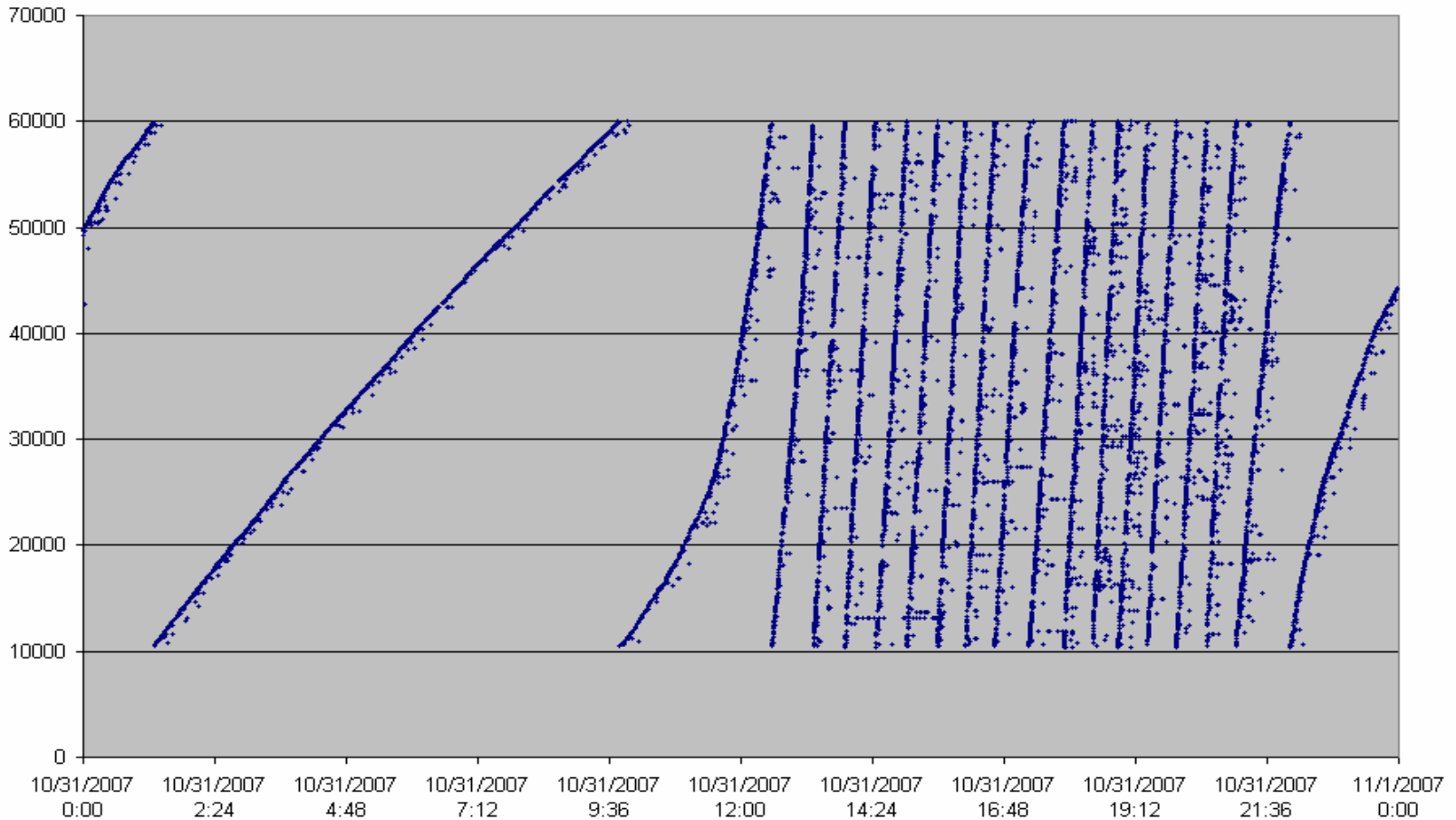
Solutions Pro Application





# Structured Gateway

sPort

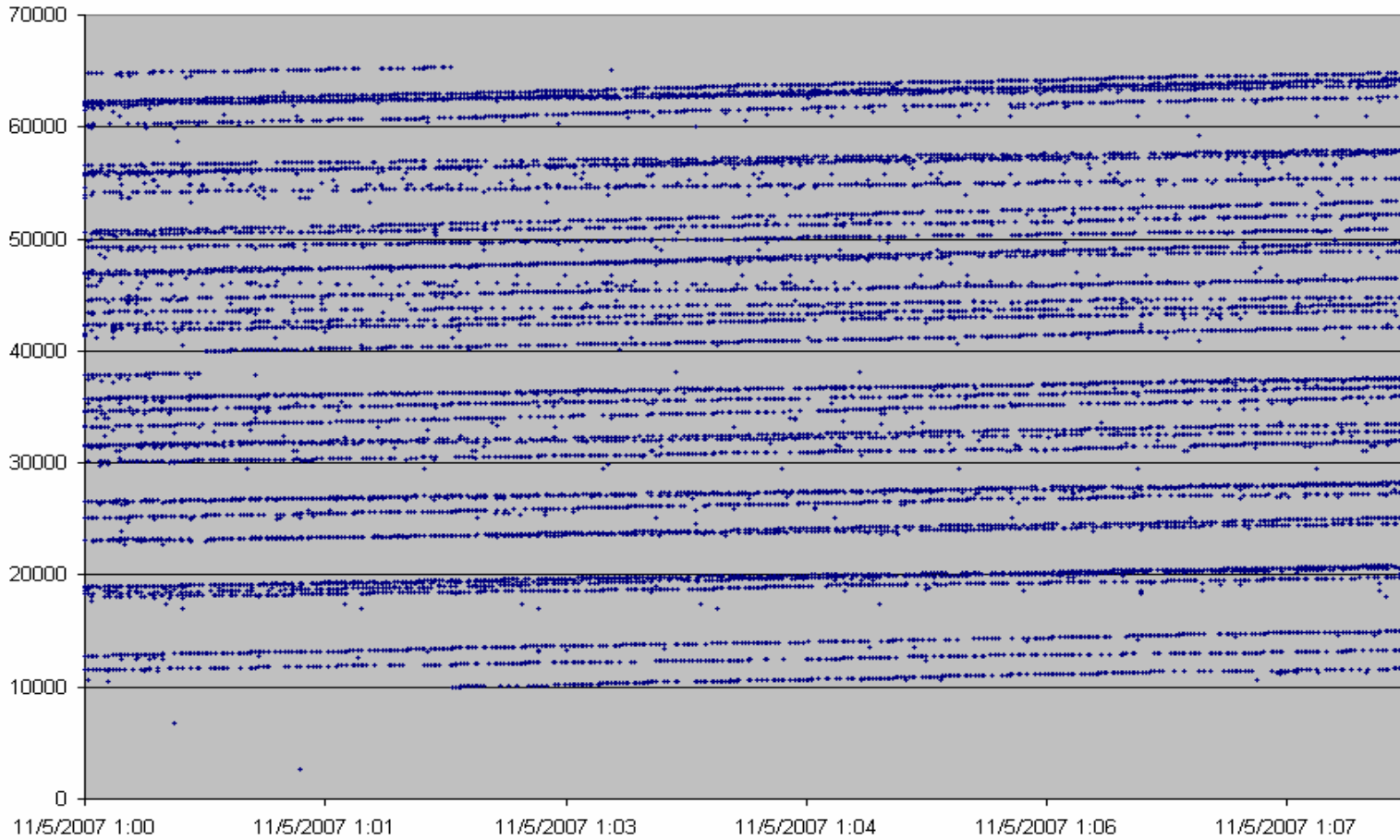






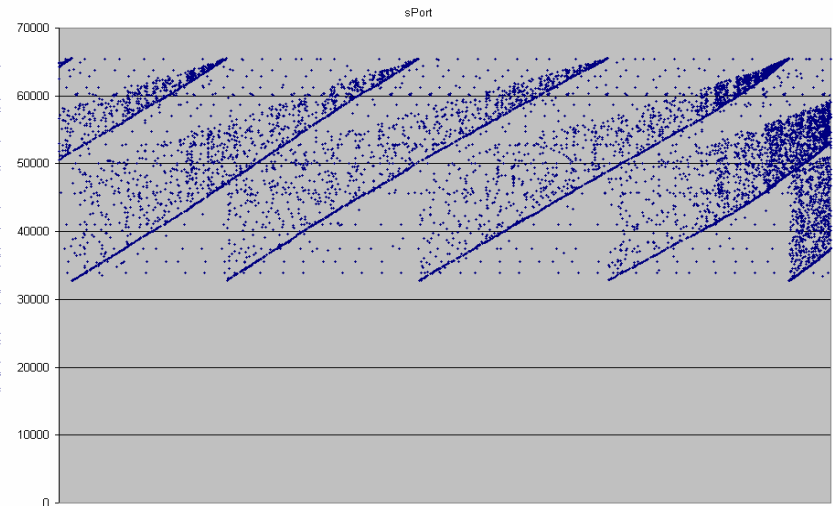
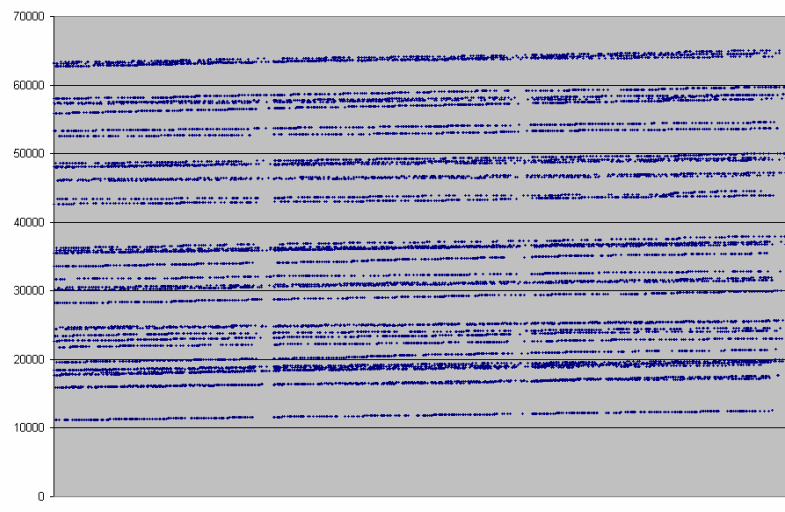
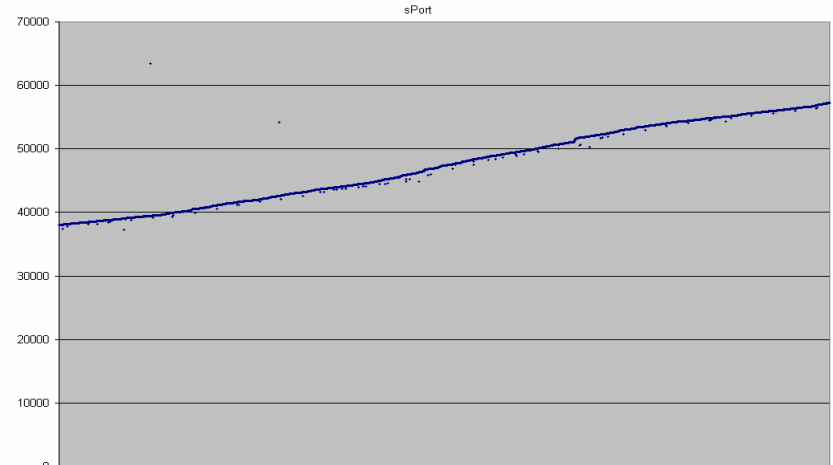
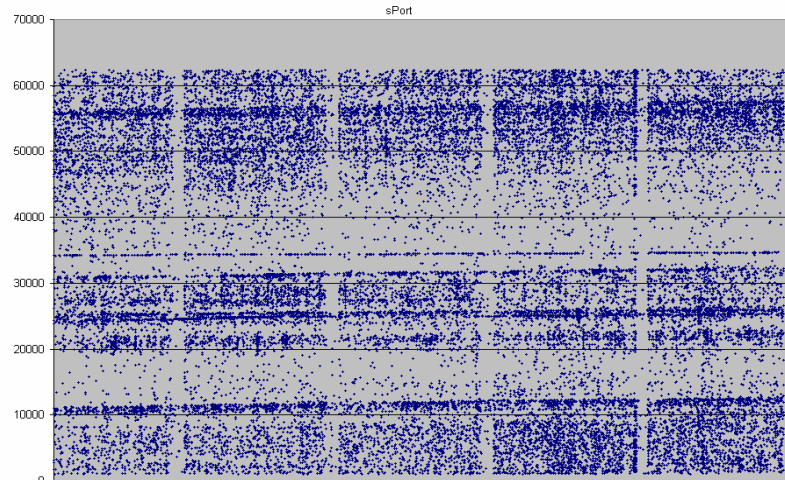
# Proxy Gateway

sPort





# Gateway Mannerisms





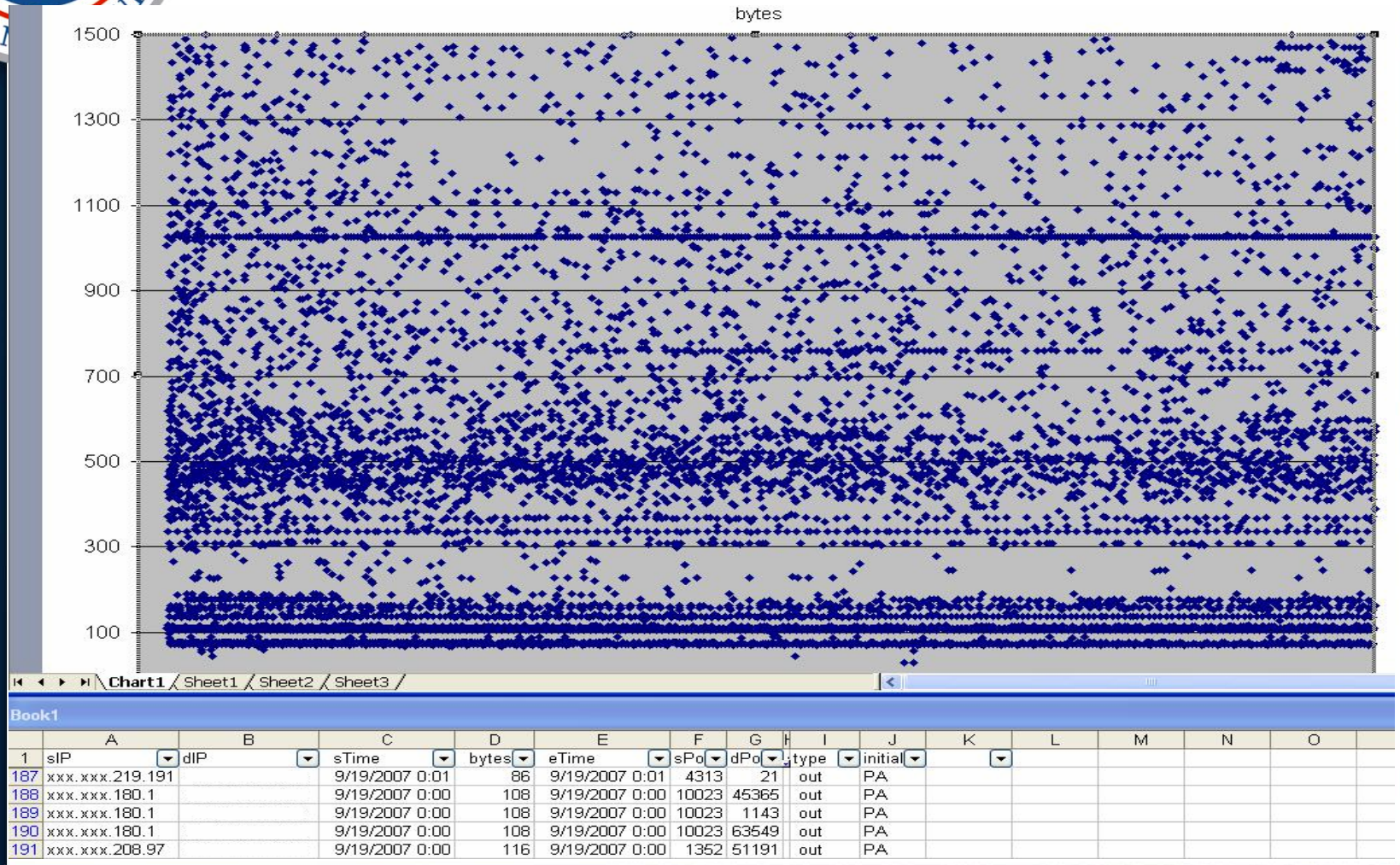
# Future Directions

- Split view analysis
- Coloring data
- Application coloring
- sPort colored by app
- Gateway coloring to IP





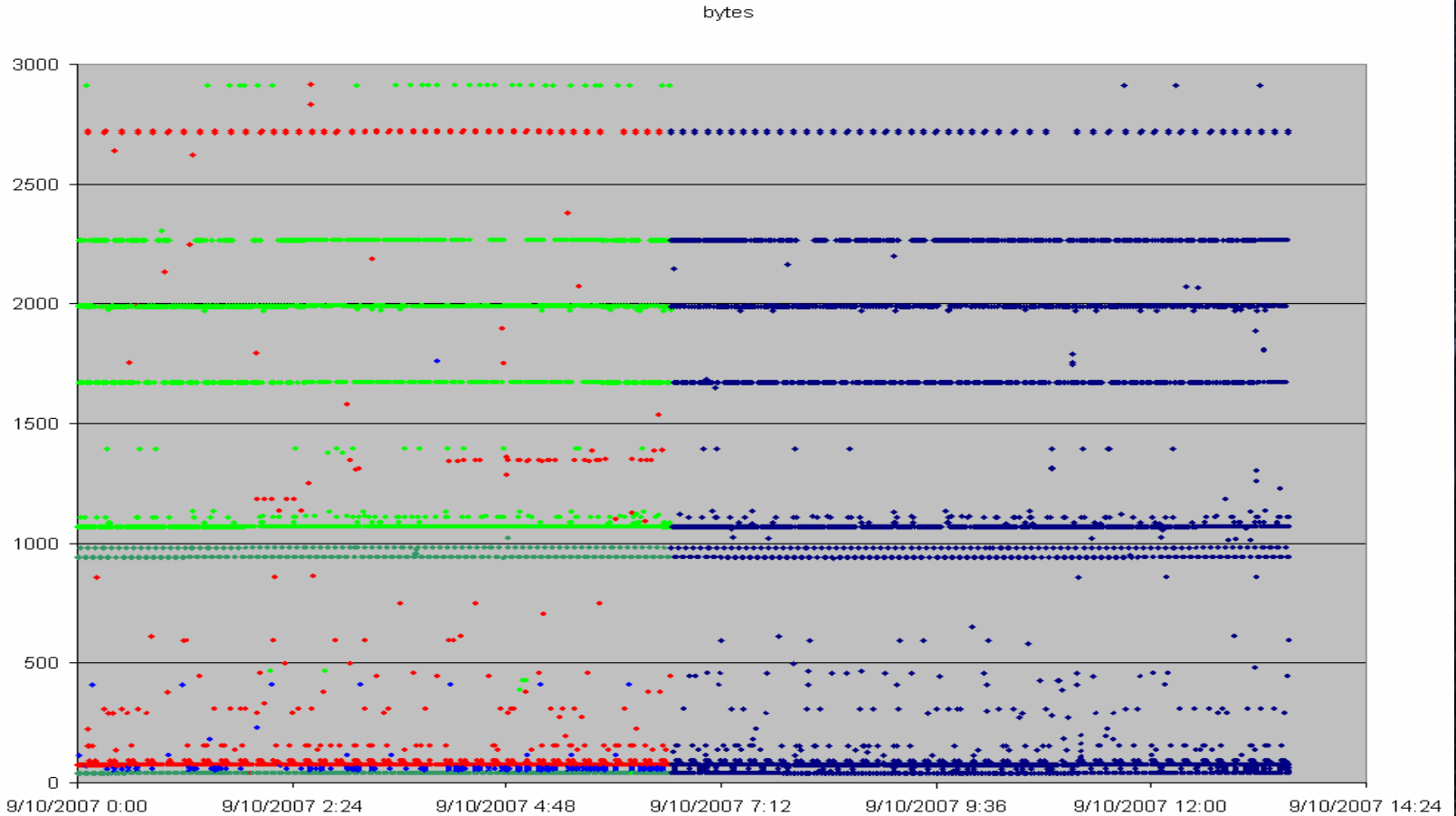
# Split View





# Coloring Example

Green = HTTP, Dark Green = HTTPS, Blue = DNS, Red = Other



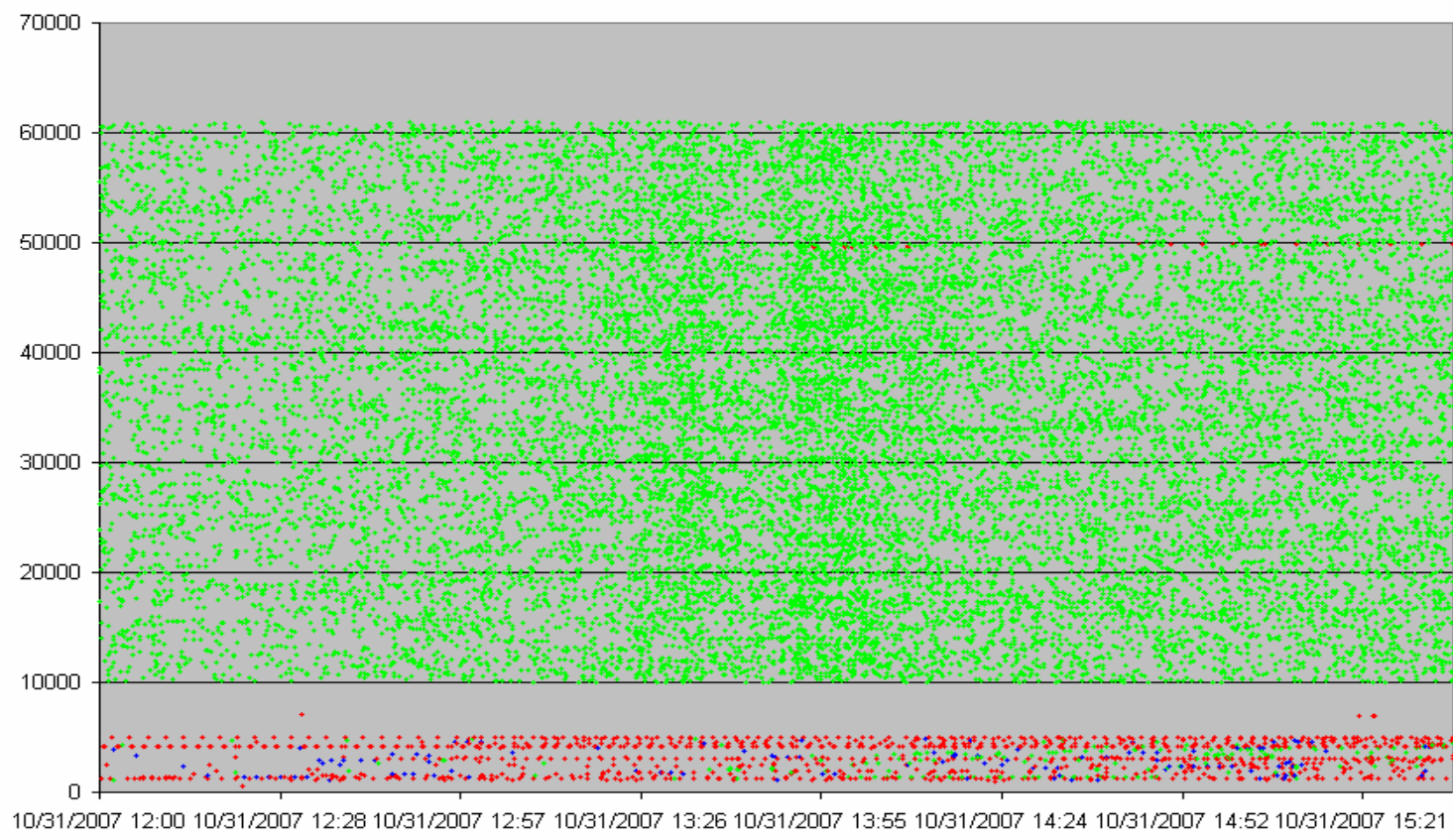




# Application Coloring

Green = HTTP, Blue = DNS, Red = Other

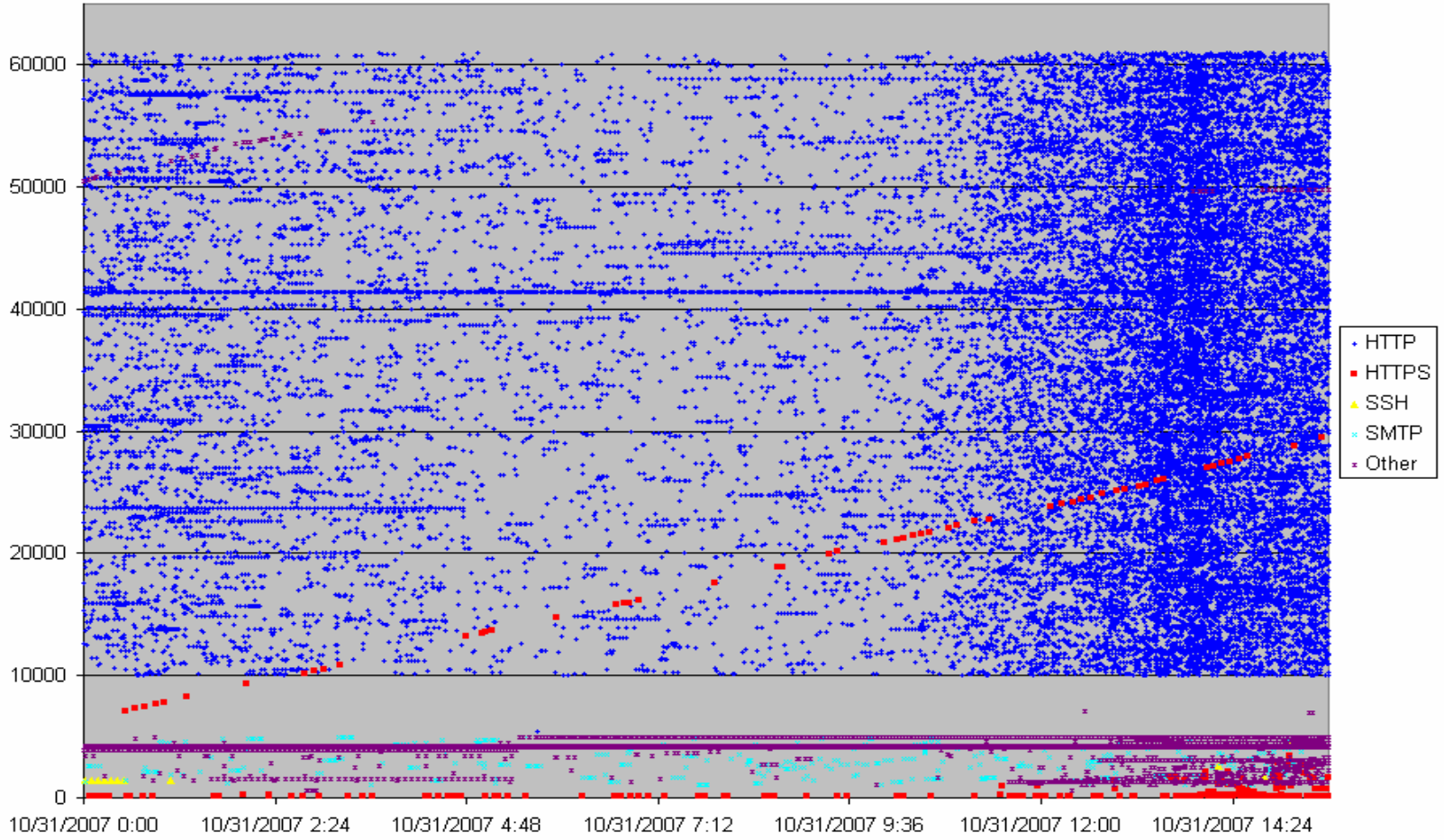
sPort





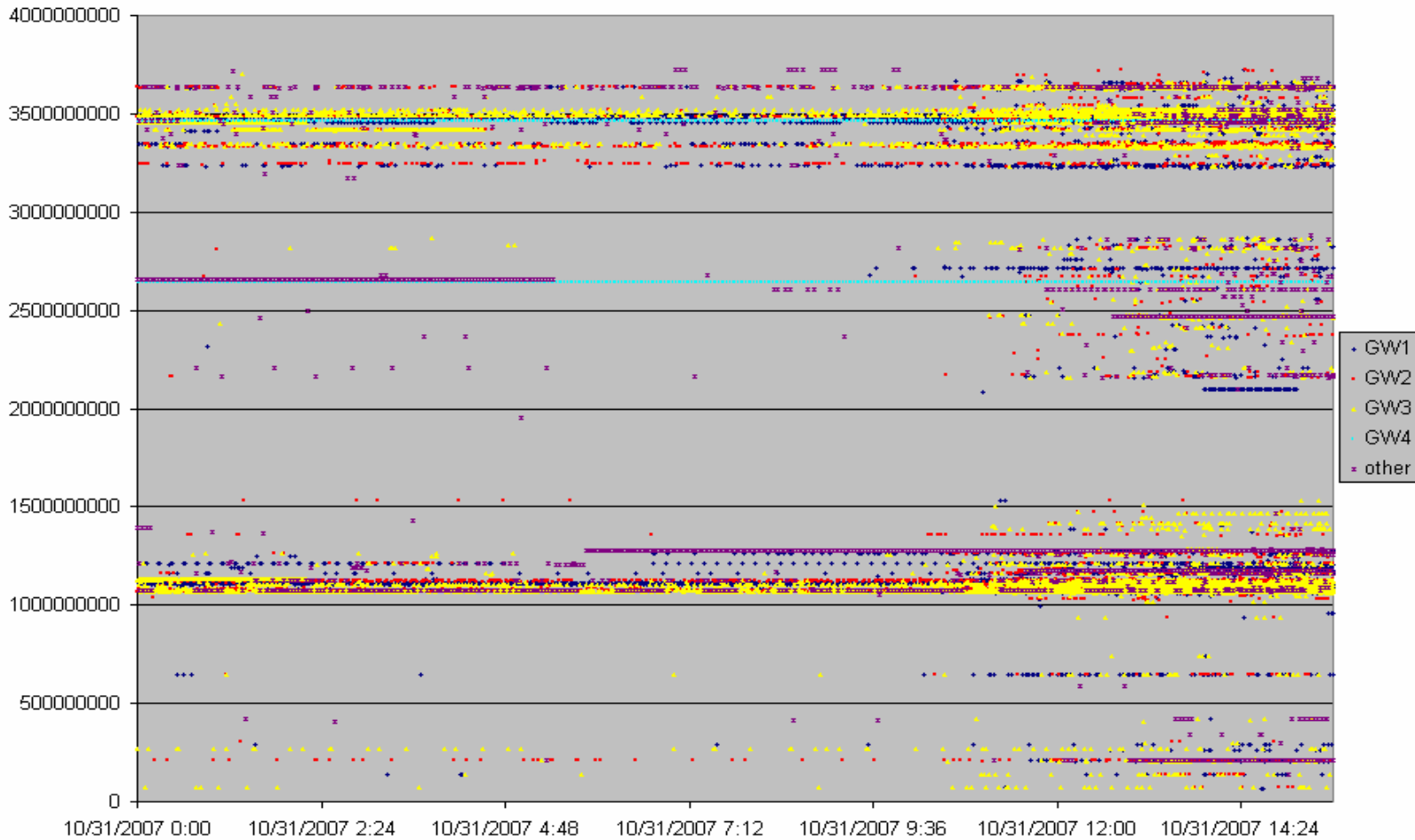


# Color sPort vs Application





# Colorization Example – GW2IP





# Contact Info

- **Technical comments or questions**
  - US-CERT Security Operations Center
  - Email: [soc@us-cert.gov](mailto:soc@us-cert.gov)
  - Phone: +1 888-282-0870
- **Media inquiries**
  - US-CERT Public Affairs
  - Email: [media@us-cert.gov](mailto:media@us-cert.gov)
  - Phone: +1 202-282-8010
- **General questions or suggestions**
  - US-CERT Information Request
  - Email: [info@us-cert.gov](mailto:info@us-cert.gov)
  - Phone: +1 703-235-5111
- **For more information, visit <http://www.us-cert.gov>**





Questions?