Isis:Visual Analysis of Network Flow Data with Timelines and Event Plots

Doantam Phan, John Gerth, Marcia Lee
Andreas Paepcke, Terry Winograd

Stanford University

The Analyst's Mindset

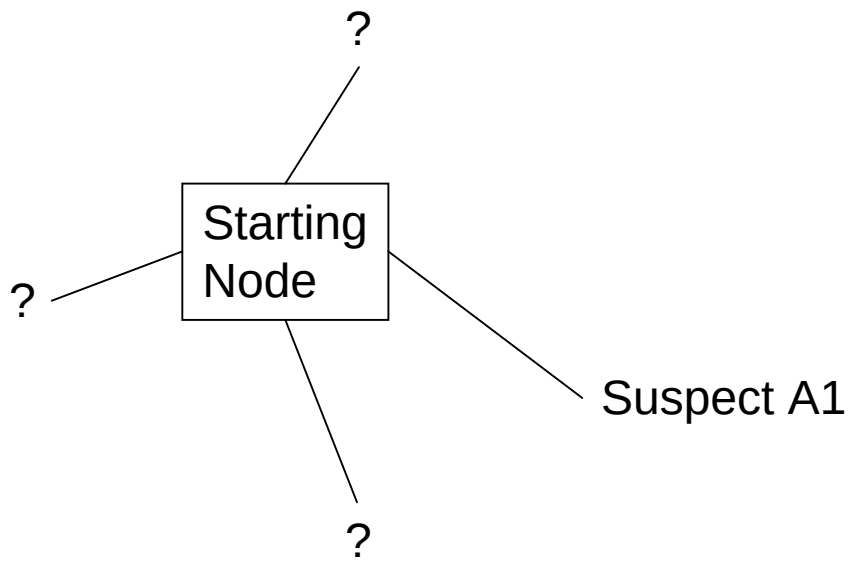
- A set of classic questions
  - Who
  - What
  - When
  - Where
  - Why
  - How
- Overlapping Tasks
  - *Perception*: Monitor
  - *Comprehension*: Explore and Investigate
  - *Projection*: Forecast and Present

Escalation Task Analysis

1. Inspect events of focus node
2. Compare different nodes to find correlations
3. Refine query to filter data
4. Pivot on an interesting node to refocus

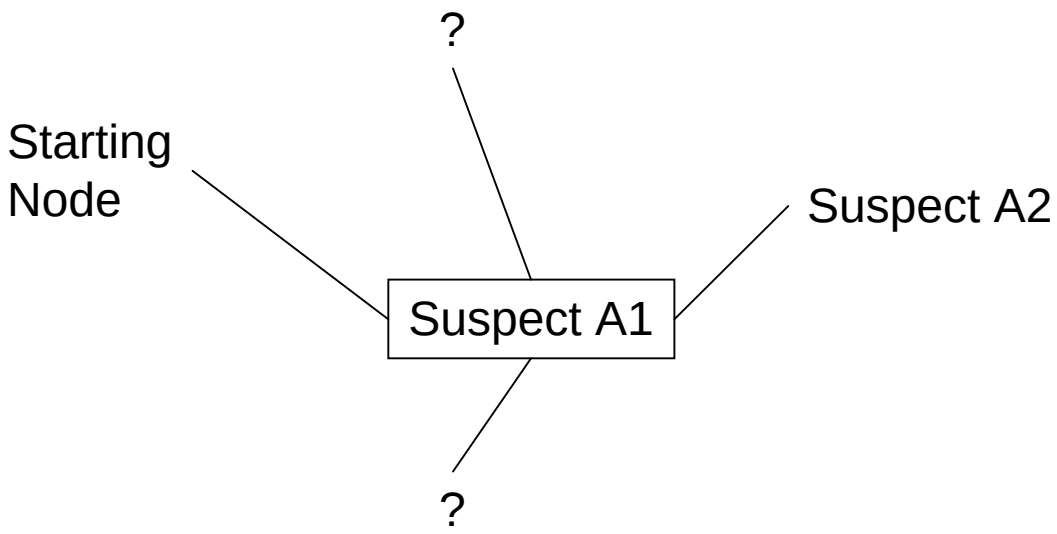
Investigation by Pivoting

- Who talked?
- When did they talk?
- How much did they say?



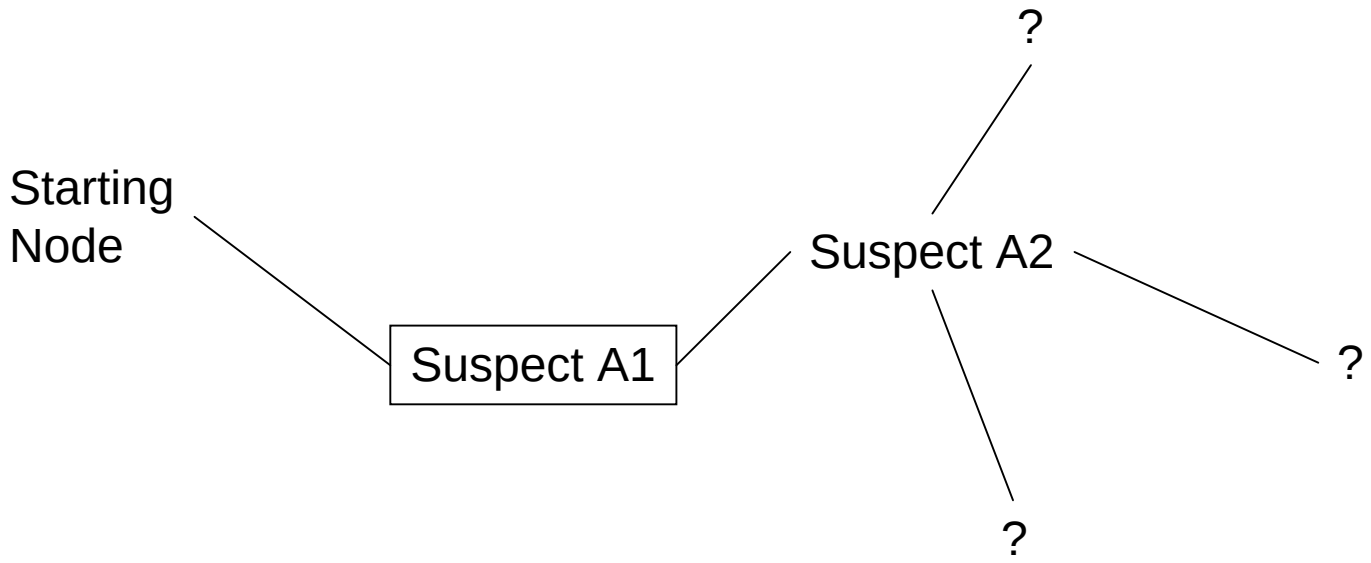
Investigation by Pivoting

- Who talked?
- When did they talk?
- How much did they say?



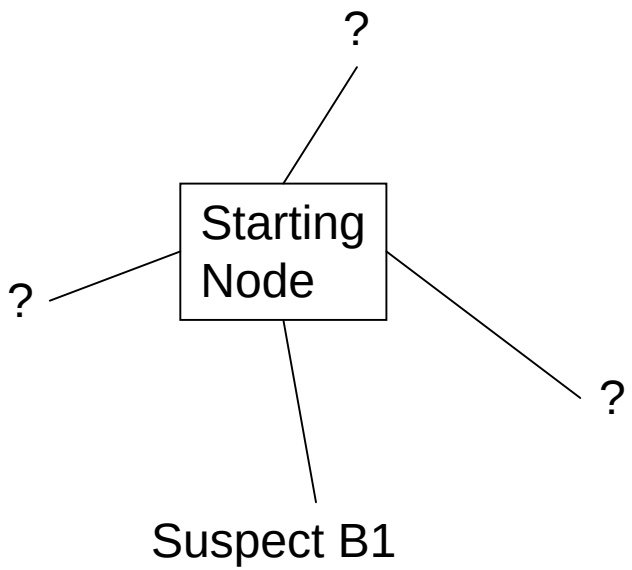
Investigation by Pivoting

- Who talked?
- When did they talk?
- How much did they say?



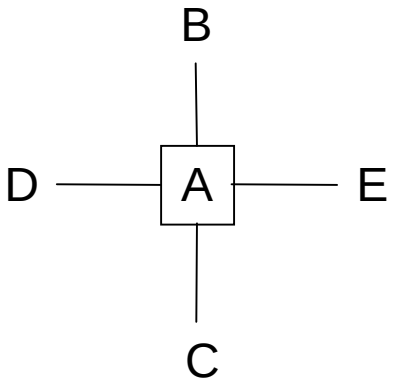
Investigation by Pivoting

- Who talked?
- When did they talk?
- How much did they say?

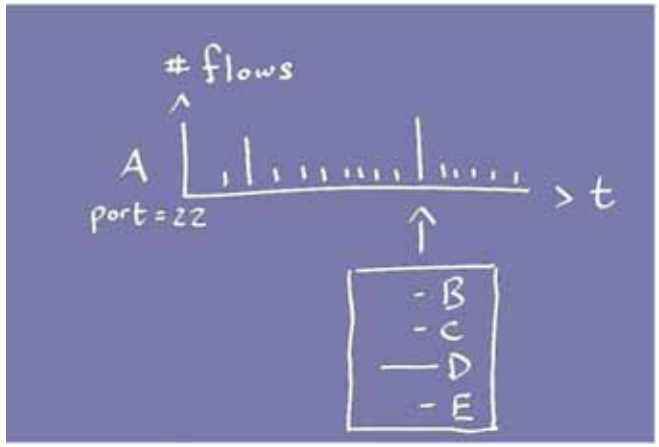


Progressive Multiples

Compare events of different nodes using row-based time-oriented displays



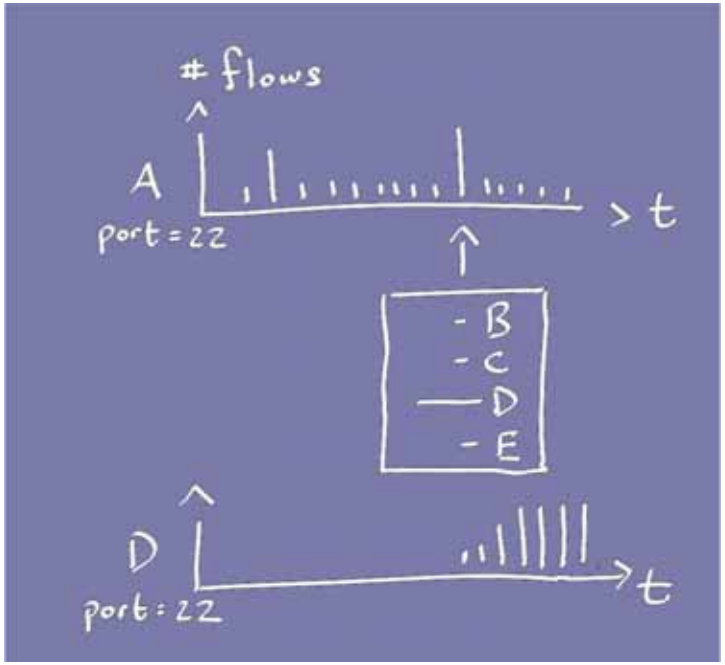
Traffic involving node A as node-link diagram



Traffic involving A as a timeline

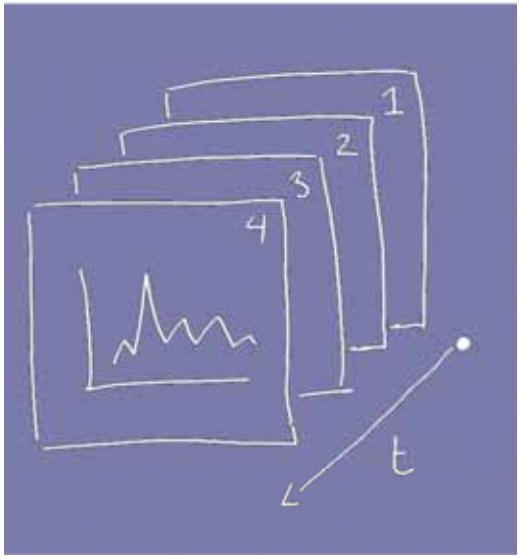
Progressive Multiples

Refocus with pivoting

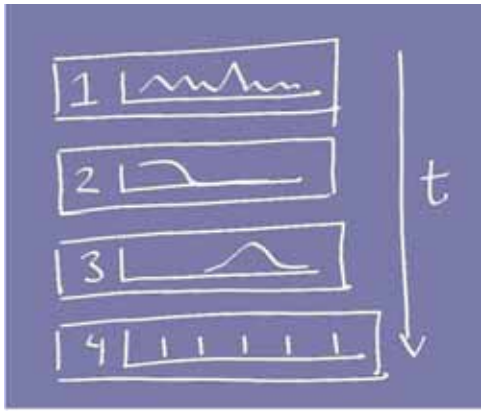


Progressive Multiples

Allow backtracking and comparisons with a structured layout that provides history



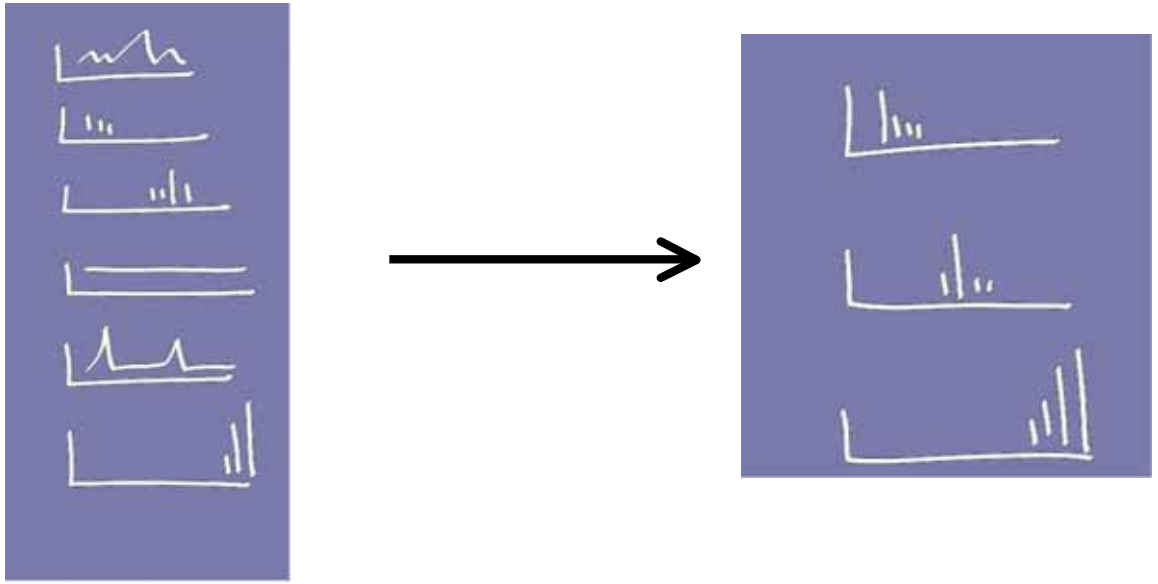
Previous visualization states



Making exploration history visible

Progressive Multiples

Organize events using row reordering to reveal underlying structure and sequence

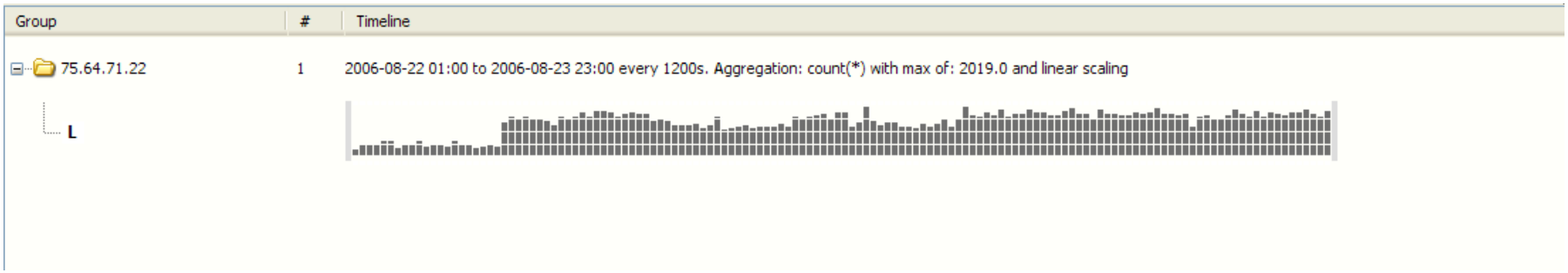


A Case of Suspicious Traffic

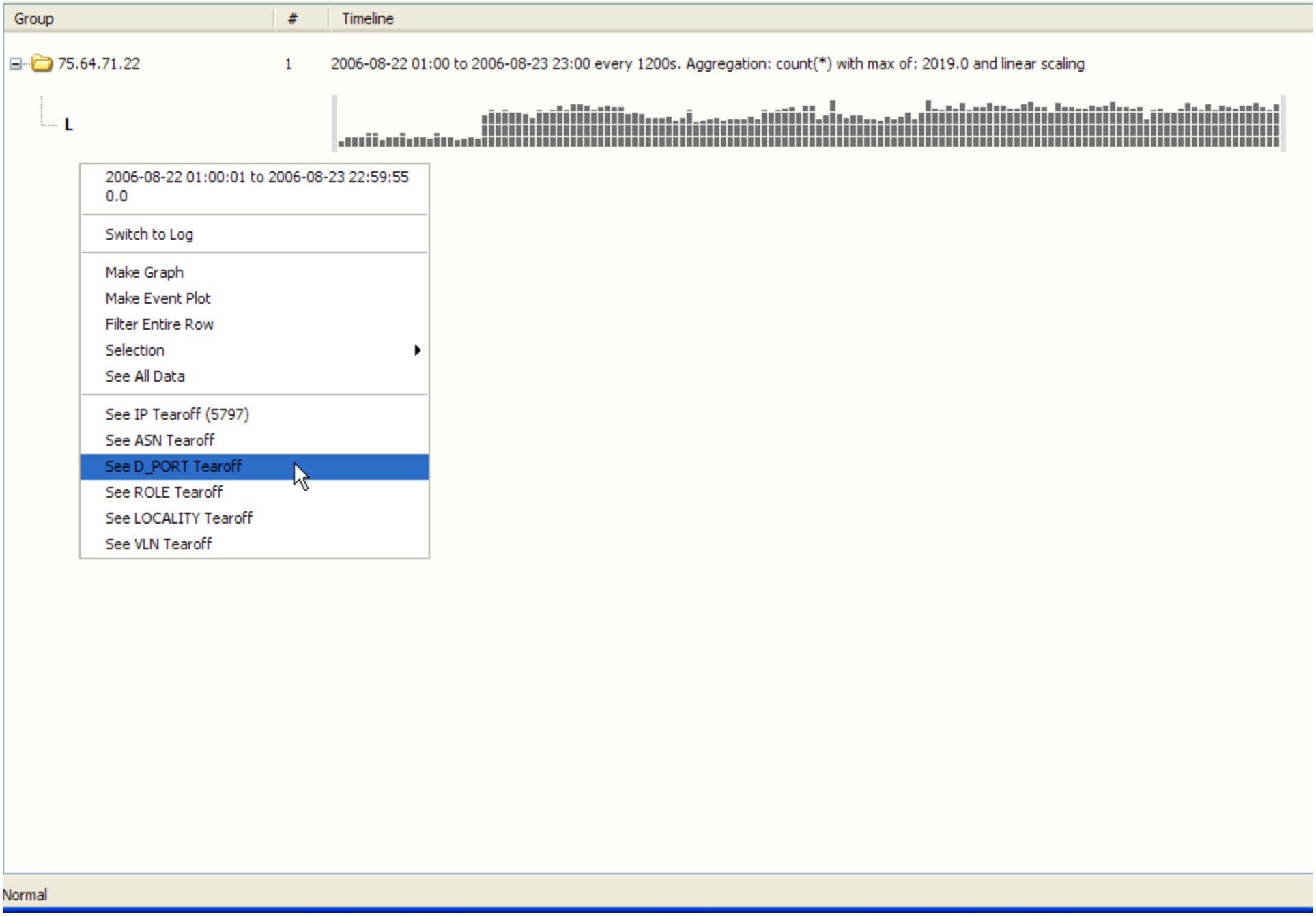
Network Monitoring contacts Security Analyst after observing extended sequence of short connections from local computer 75.64.71.22 to an Internet Relay Chat (IRC) server in Europe

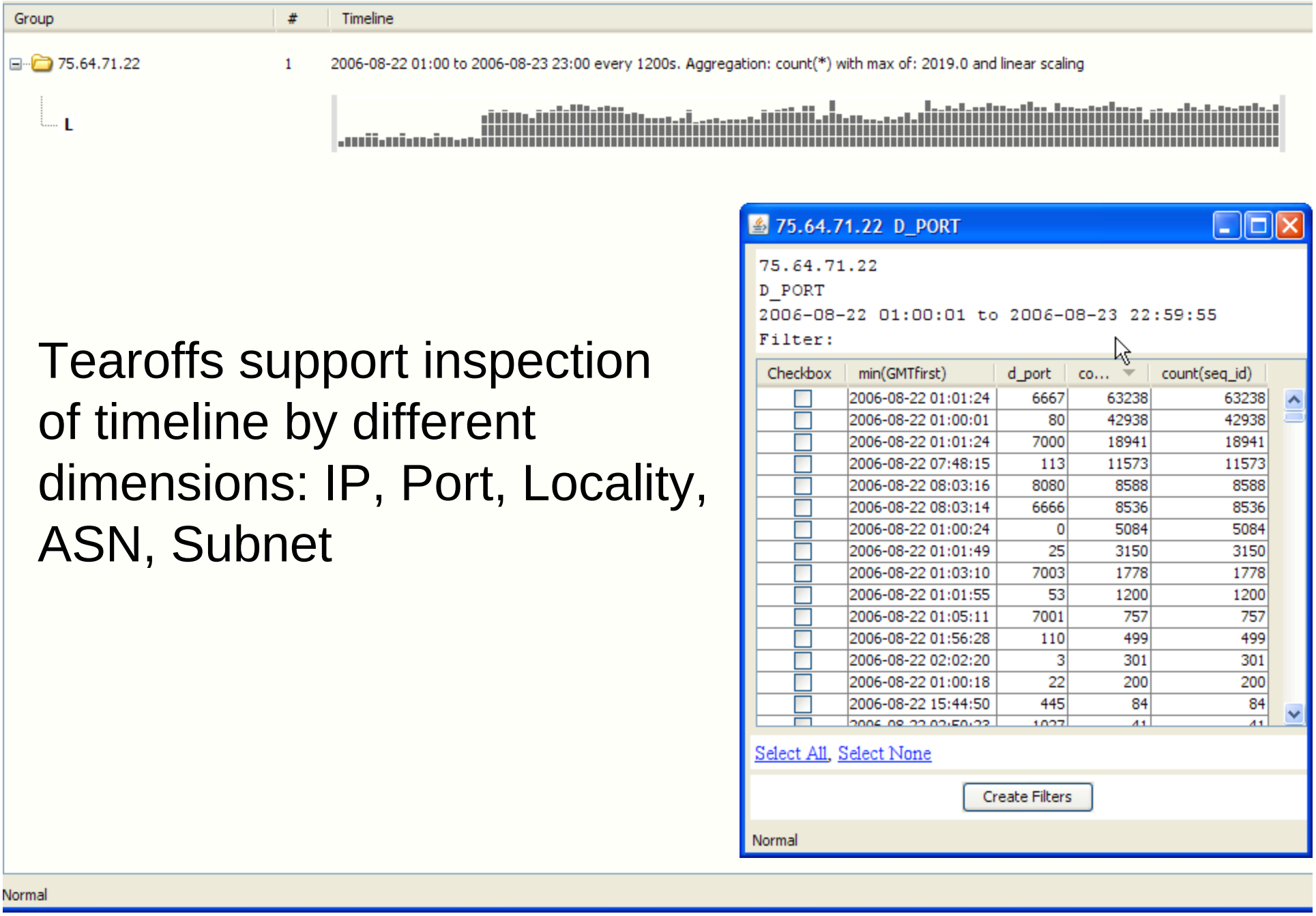
Analyst specifies initial query values

- Focus IP Address – 75.64.71.22
- Start and End Time – 1 day
- Aggregation Expression – count(*)
- Filter – start with all traffic

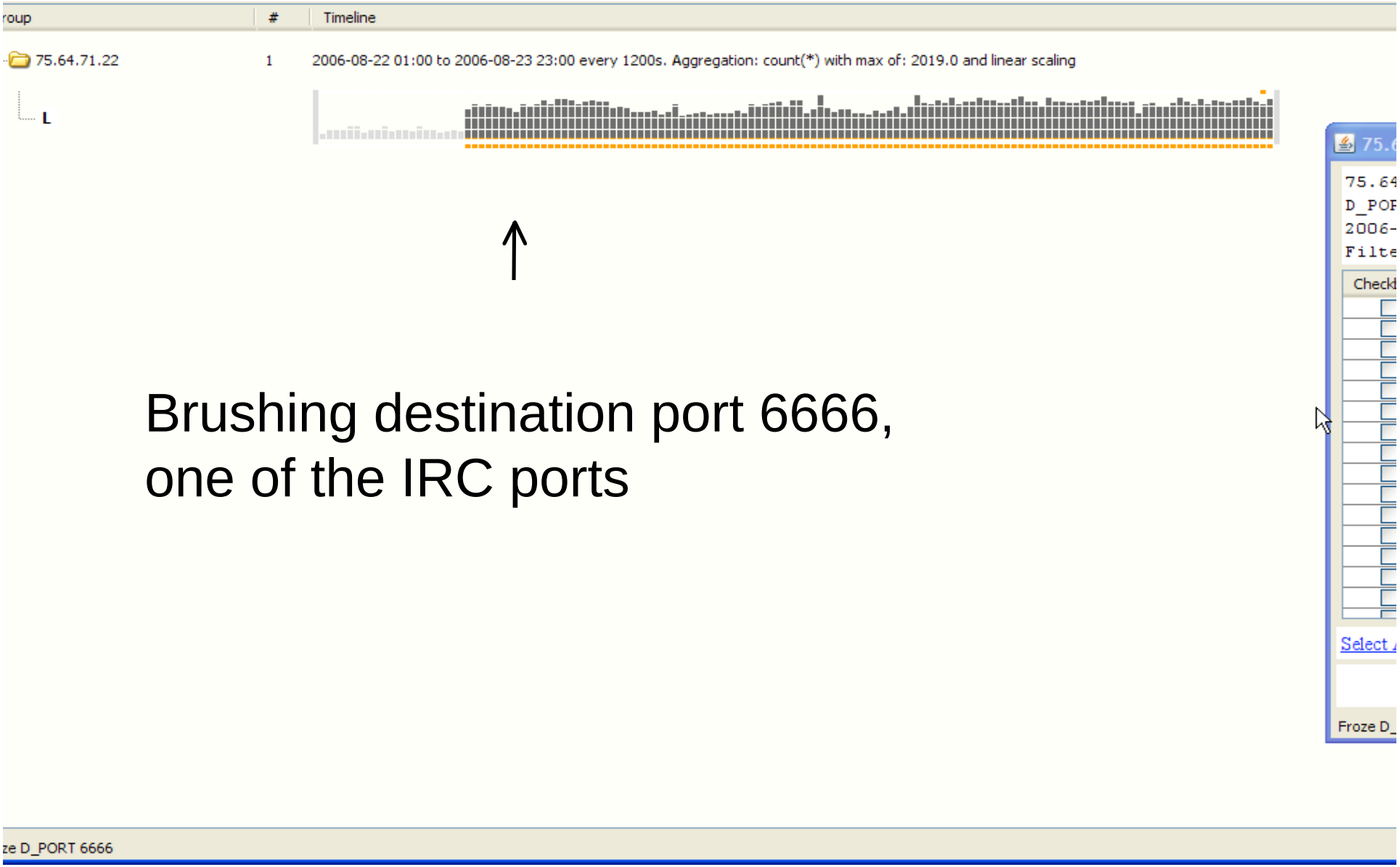


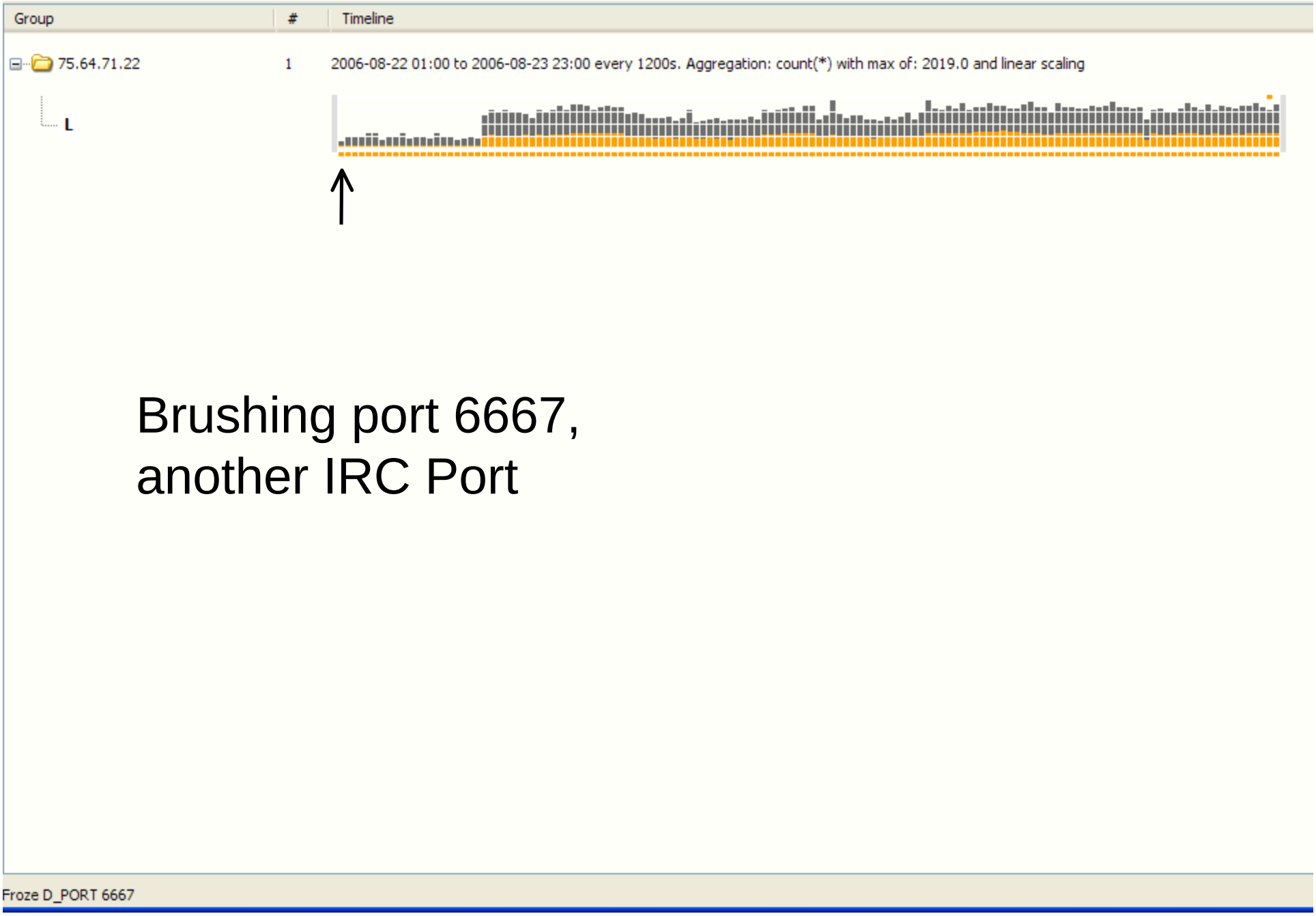
When did the IRC traffic start?


Group
#
Timeline
75.64.71.22
1
2006-08-22 01:00 to 2006-08-23 23:00 every 1200s. Aggregation: count(*) with max of: 2019.0 and linear scaling
L
2006-08-22 01:00:01 to 2006-08-23 22:59:55
0.0
Switch to Log
Make Graph
Make Event Plot
Filter Entire Row
Selection
See All Data
See IP Tearoff (5797)
See ASN Tearoff
See D_PORT Tearoff
See ROLE Tearoff
See LOCALITY Tearoff
See VLN Tearoff
Normal

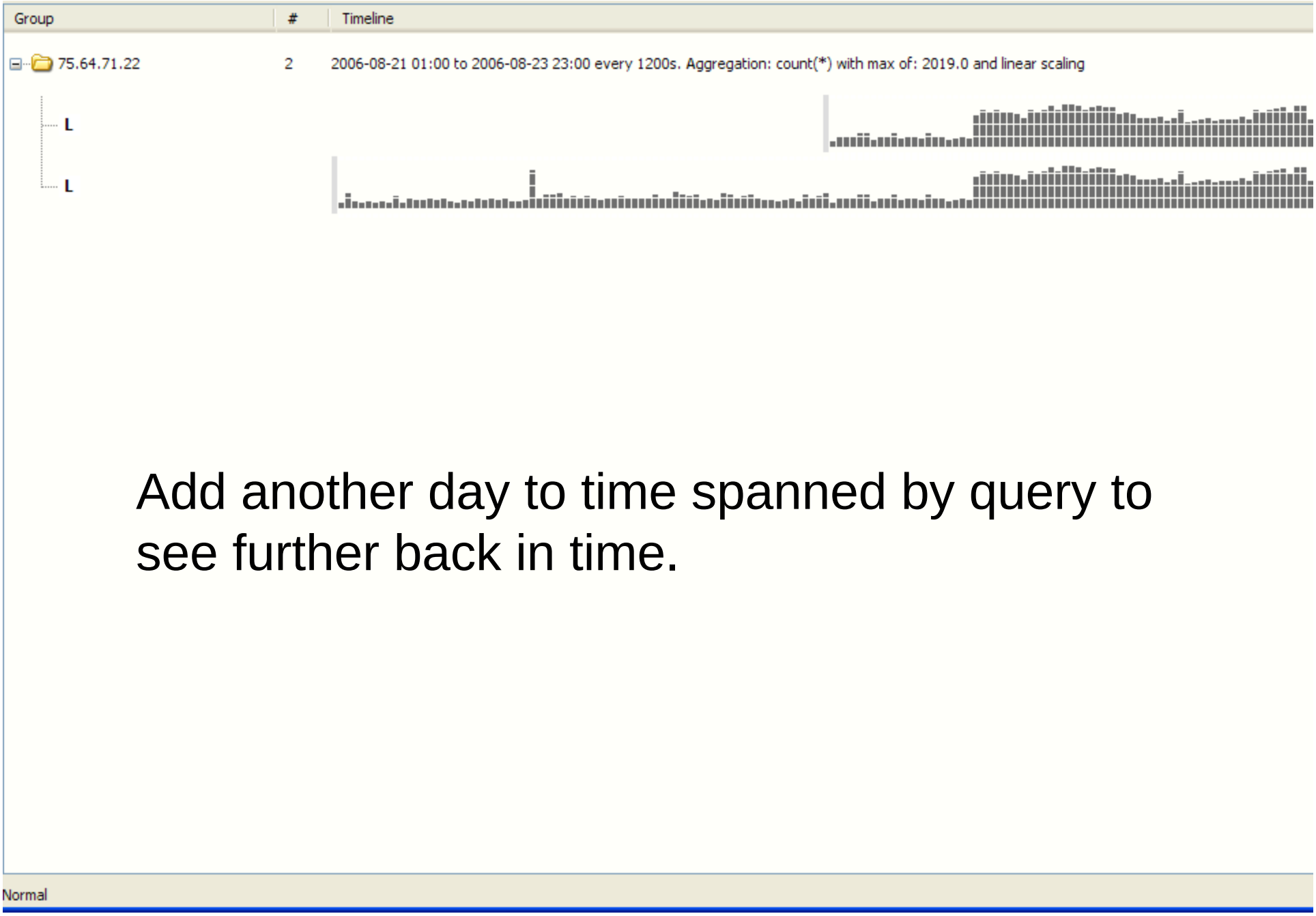


Tearoffs support inspection of timeline by different dimensions: IP, Port, Locality, ASN, Subnet

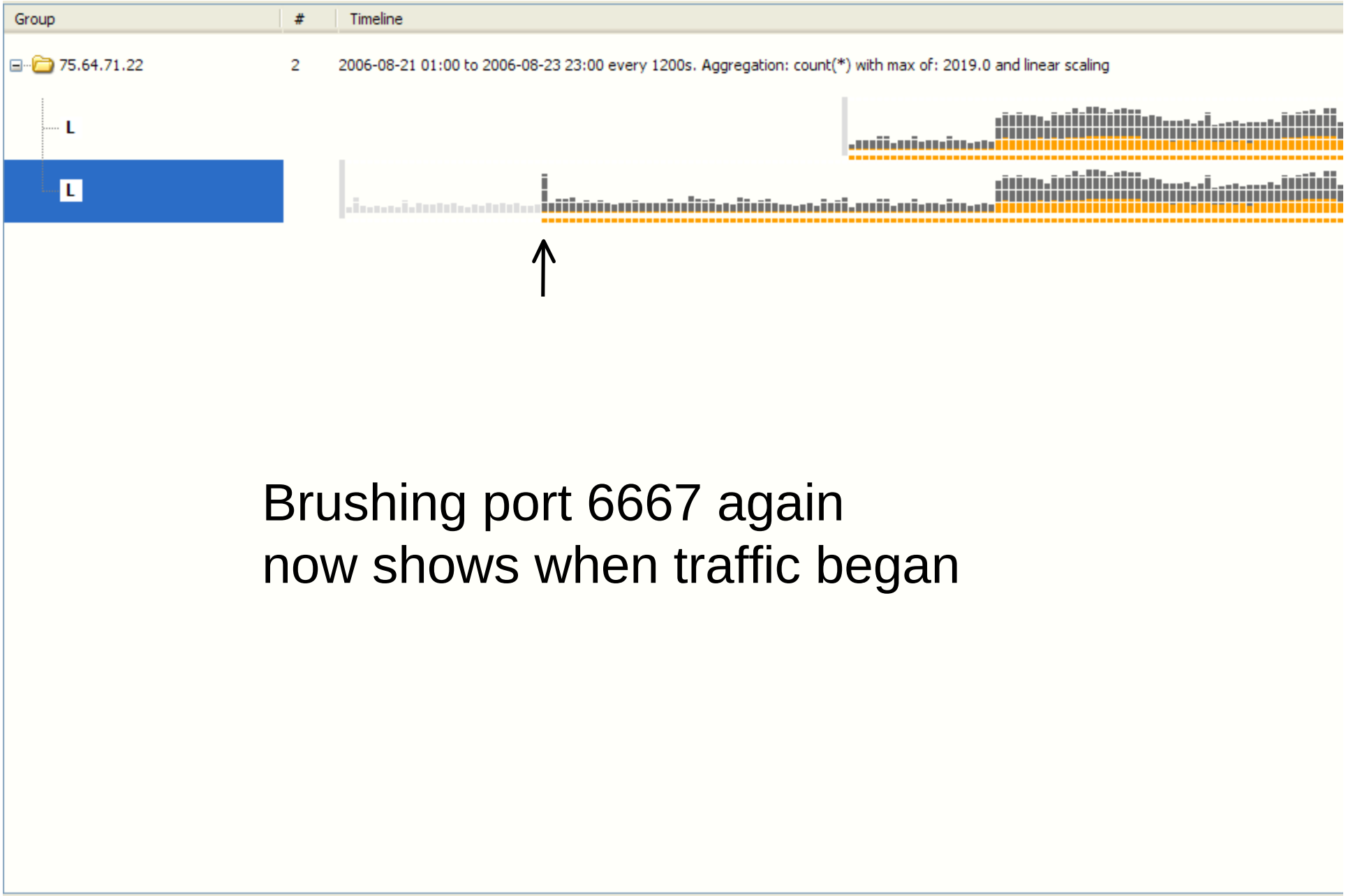

roup
#
Timeline
75.64.71.22
1
2006-08-22 01:00 to 2006-08-23 23:00 every 1200s. Aggregation: count(*) with max of: 2019.0 and linear scaling
L
Brushing destination port 6666,
one of the IRC ports
ze D_PORT 6666



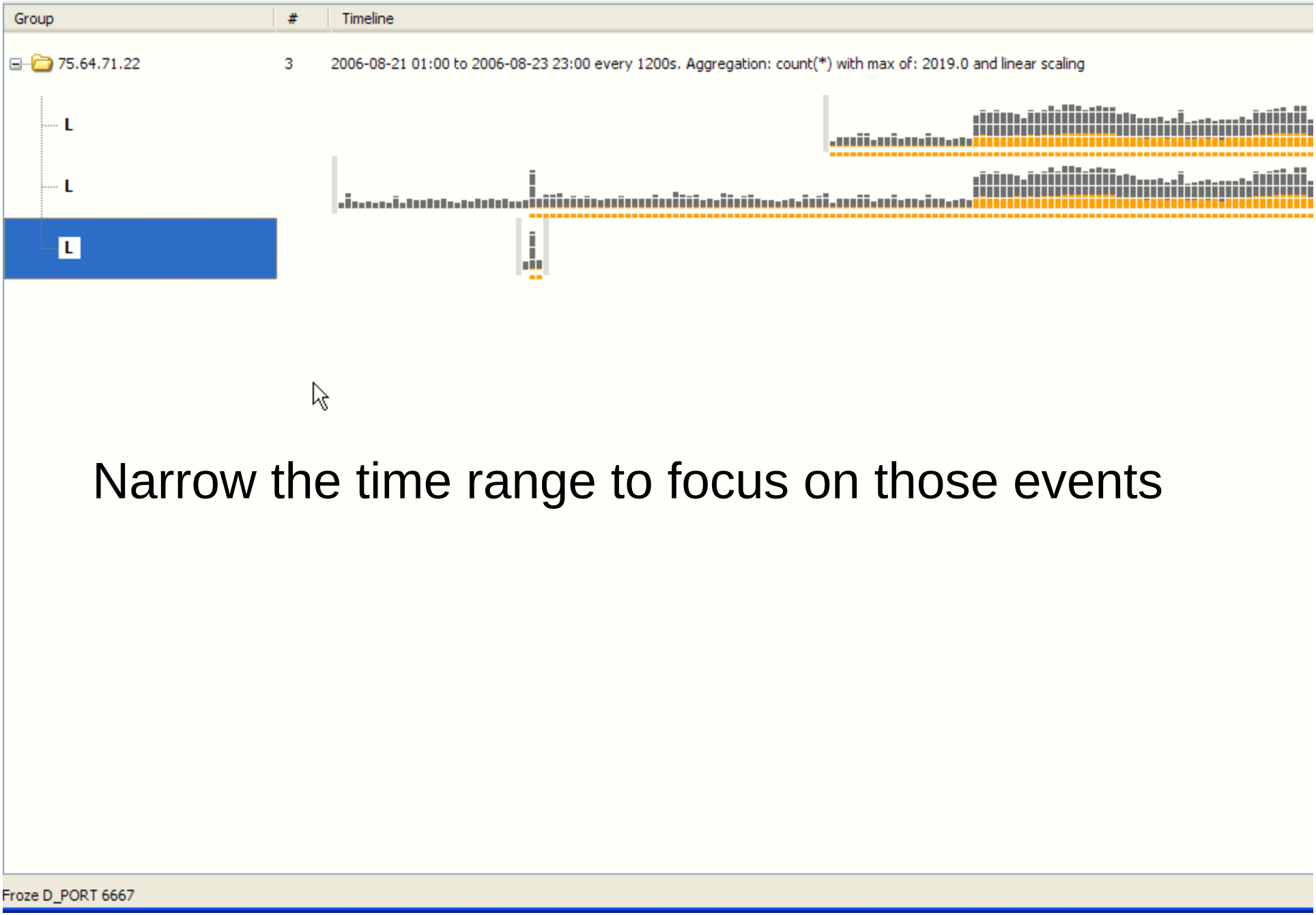
Brushing port 6667,
another IRC Port



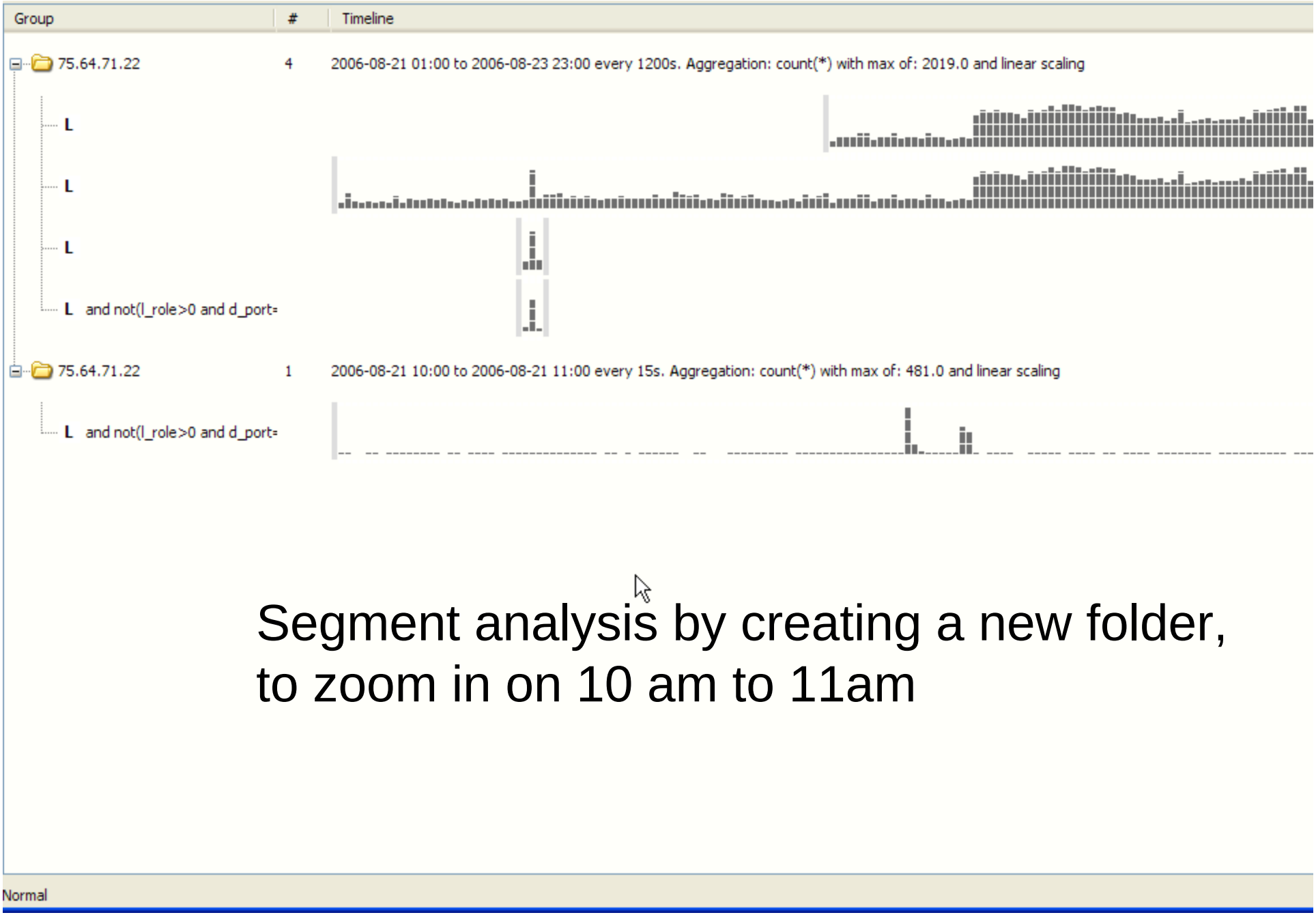
Add another day to time spanned by query to see further back in time.



Brushing port 6667 again
now shows when traffic began

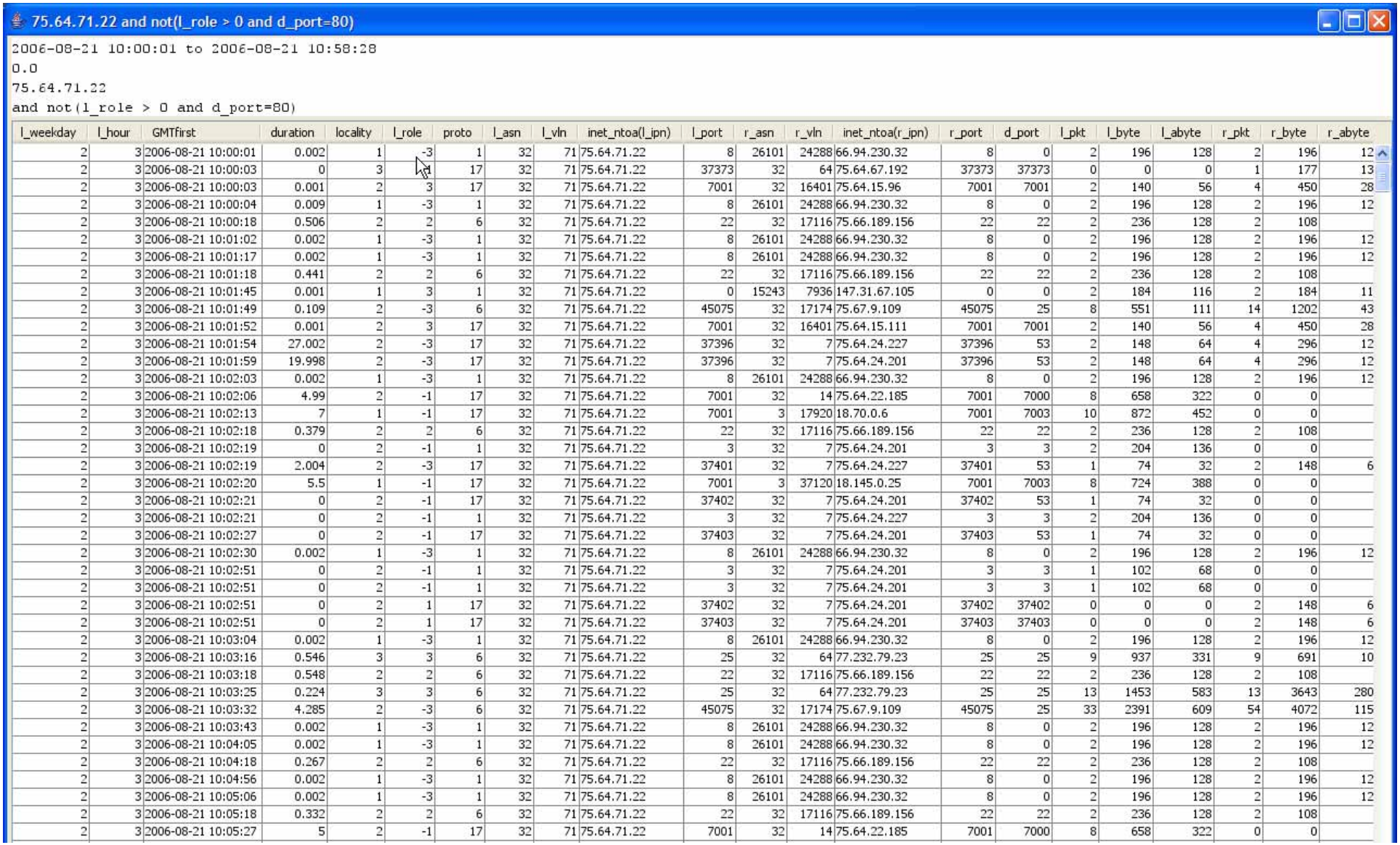


Narrow the time range to focus on those events



Segment analysis by creating a new folder, to zoom in on 10 am to 11am

The Event Table



75.64.71.22 and not(l_role > 0 and d_port=80)

2006-08-21 10:00:01 to 2006-08-21 10:58:28
0.0
75.64.71.22
and not(l_role > 0 and d_port=80)

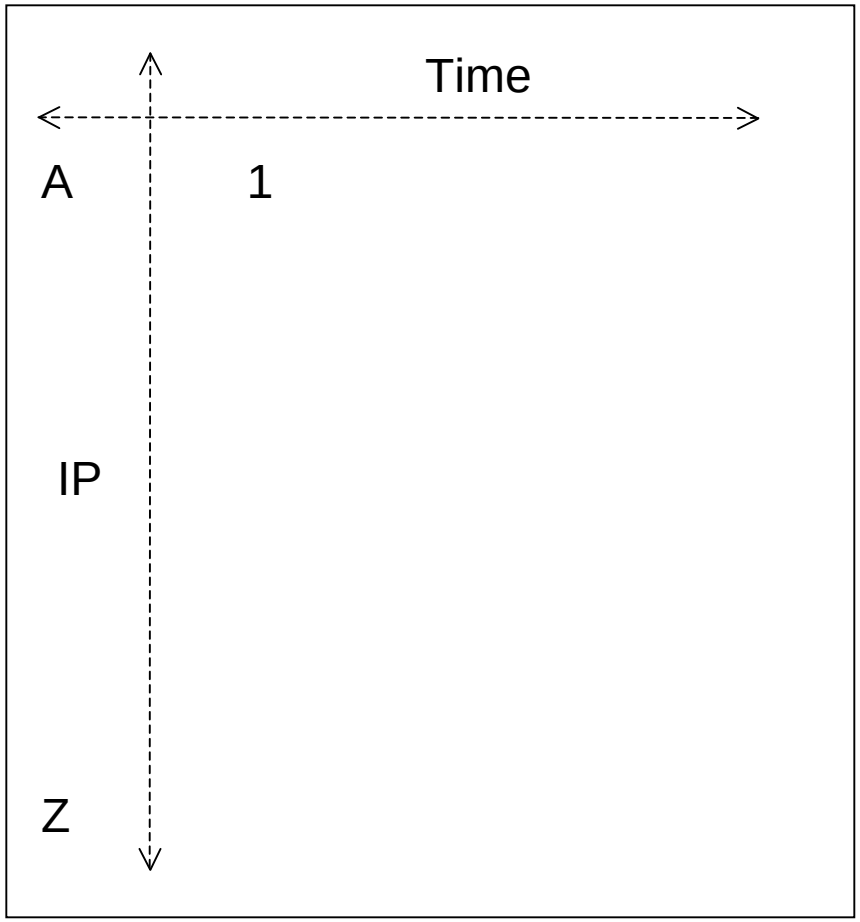
l_weekday	l_hour	GMTfirst	duration	locality	l_role	proto	l_asn	l_vln	inet_ntoa(l_ipn)	l_port	r_asn	r_vln	inet_ntoa(r_ipn)	r_port	d_port	l_pkt	l_byte	l_abyte	r_pkt	r_byte	r_abyte
2	3	2006-08-21 10:00:01	0.002	1	-3	1	32	71	75.64.71.22	8	26101	24288	66.94.230.32	8	0	2	196	128	2	196	12
2	3	2006-08-21 10:00:03	0	3	3	17	32	71	75.64.71.22	37373	32	64	75.64.67.192	37373	37373	0	0	0	1	177	13
2	3	2006-08-21 10:00:03	0.001	2	3	17	32	71	75.64.71.22	7001	32	16401	75.64.15.96	7001	7001	2	140	56	4	450	28
2	3	2006-08-21 10:00:04	0.009	1	-3	1	32	71	75.64.71.22	8	26101	24288	66.94.230.32	8	0	2	196	128	2	196	12
2	3	2006-08-21 10:00:18	0.506	2	2	6	32	71	75.64.71.22	22	32	17116	75.66.189.156	22	22	2	236	128	2	108	
2	3	2006-08-21 10:01:02	0.002	1	-3	1	32	71	75.64.71.22	8	26101	24288	66.94.230.32	8	0	2	196	128	2	196	12
2	3	2006-08-21 10:01:17	0.002	1	-3	1	32	71	75.64.71.22	8	26101	24288	66.94.230.32	8	0	2	196	128	2	196	12
2	3	2006-08-21 10:01:18	0.441	2	2	6	32	71	75.64.71.22	22	32	17116	75.66.189.156	22	22	2	236	128	2	108	
2	3	2006-08-21 10:01:45	0.001	1	3	1	32	71	75.64.71.22	0	15243	7936	147.31.67.105	0	0	2	184	116	2	184	11
2	3	2006-08-21 10:01:49	0.109	2	-3	6	32	71	75.64.71.22	45075	32	17174	75.67.9.109	45075	25	8	551	111	14	1202	43
2	3	2006-08-21 10:01:52	0.001	2	3	17	32	71	75.64.71.22	7001	32	16401	75.64.15.111	7001	7001	2	140	56	4	450	28
2	3	2006-08-21 10:01:54	27.002	2	-3	17	32	71	75.64.71.22	37396	32	7	75.64.24.227	37396	53	2	148	64	4	296	12
2	3	2006-08-21 10:01:59	19.998	2	-3	17	32	71	75.64.71.22	37396	32	7	75.64.24.201	37396	53	2	148	64	4	296	12
2	3	2006-08-21 10:02:03	0.002	1	-3	1	32	71	75.64.71.22	8	26101	24288	66.94.230.32	8	0	2	196	128	2	196	12
2	3	2006-08-21 10:02:06	4.99	2	-1	17	32	71	75.64.71.22	7001	32	14	75.64.22.185	7001	7000	8	658	322	0	0	
2	3	2006-08-21 10:02:13	7	1	-1	17	32	71	75.64.71.22	7001	3	17920	18.70.0.6	7001	7003	10	872	452	0	0	
2	3	2006-08-21 10:02:18	0.379	2	2	6	32	71	75.64.71.22	22	32	17116	75.66.189.156	22	22	2	236	128	2	108	
2	3	2006-08-21 10:02:19	0	2	-1	1	32	71	75.64.71.22	3	32	7	75.64.24.201	3	3	2	204	136	0	0	
2	3	2006-08-21 10:02:19	2.004	2	-3	17	32	71	75.64.71.22	37401	32	7	75.64.24.227	37401	53	1	74	32	2	148	6
2	3	2006-08-21 10:02:20	5.5	1	-1	17	32	71	75.64.71.22	7001	3	37120	18.145.0.25	7001	7003	8	724	388	0	0	
2	3	2006-08-21 10:02:21	0	2	-1	17	32	71	75.64.71.22	37402	32	7	75.64.24.201	37402	53	1	74	32	0	0	
2	3	2006-08-21 10:02:21	0	2	-1	1	32	71	75.64.71.22	3	32	7	75.64.24.227	3	3	2	204	136	0	0	
2	3	2006-08-21 10:02:27	0	2	-1	17	32	71	75.64.71.22	37403	32	7	75.64.24.201	37403	53	1	74	32	0	0	
2	3	2006-08-21 10:02:30	0.002	1	-3	1	32	71	75.64.71.22	8	26101	24288	66.94.230.32	8	0	2	196	128	2	196	12
2	3	2006-08-21 10:02:51	0	2	-1	1	32	71	75.64.71.22	3	32	7	75.64.24.201	3	3	1	102	68	0	0	
2	3	2006-08-21 10:02:51	0	2	-1	1	32	71	75.64.71.22	3	32	7	75.64.24.201	3	3	1	102	68	0	0	
2	3	2006-08-21 10:02:51	0	2	1	17	32	71	75.64.71.22	37402	32	7	75.64.24.201	37402	37402	0	0	0	2	148	6
2	3	2006-08-21 10:02:51	0	2	1	17	32	71	75.64.71.22	37403	32	7	75.64.24.201	37403	37403	0	0	0	2	148	6
2	3	2006-08-21 10:03:04	0.002	1	-3	1	32	71	75.64.71.22	8	26101	24288	66.94.230.32	8	0	2	196	128	2	196	12
2	3	2006-08-21 10:03:16	0.546	3	3	6	32	71	75.64.71.22	25	32	64	77.232.79.23	25	25	9	937	331	9	691	10
2	3	2006-08-21 10:03:18	0.548	2	2	6	32	71	75.64.71.22	22	32	17116	75.66.189.156	22	22	2	236	128	2	108	
2	3	2006-08-21 10:03:25	0.224	3	3	6	32	71	75.64.71.22	25	32	64	77.232.79.23	25	25	13	1453	583	13	3643	280
2	3	2006-08-21 10:03:32	4.285	2	-3	6	32	71	75.64.71.22	45075	32	17174	75.67.9.109	45075	25	33	2391	609	54	4072	115
2	3	2006-08-21 10:03:43	0.002	1	-3	1	32	71	75.64.71.22	8	26101	24288	66.94.230.32	8	0	2	196	128	2	196	12
2	3	2006-08-21 10:04:05	0.002	1	-3	1	32	71	75.64.71.22	8	26101	24288	66.94.230.32	8	0	2	196	128	2	196	12
2	3	2006-08-21 10:04:18	0.267	2	2	6	32	71	75.64.71.22	22	32	17116	75.66.189.156	22	22	2	236	128	2	108	
2	3	2006-08-21 10:04:56	0.002	1	-3	1	32	71	75.64.71.22	8	26101	24288	66.94.230.32	8	0	2	196	128	2	196	12
2	3	2006-08-21 10:05:06	0.002	1	-3	1	32	71	75.64.71.22	8	26101	24288	66.94.230.32	8	0	2	196	128	2	196	12
2	3	2006-08-21 10:05:18	0.332	2	2	6	32	71	75.64.71.22	22	32	17116	75.66.189.156	22	22	2	236	128	2	108	
2	3	2006-08-21 10:05:27	5	2	-1	17	32	71	75.64.71.22	7001	32	14	75.64.22.185	7001	7000	8	658	322	0	0	

From Event Table to Event Plot

Event Table

1	Time	A	…	Measures

Event Plot

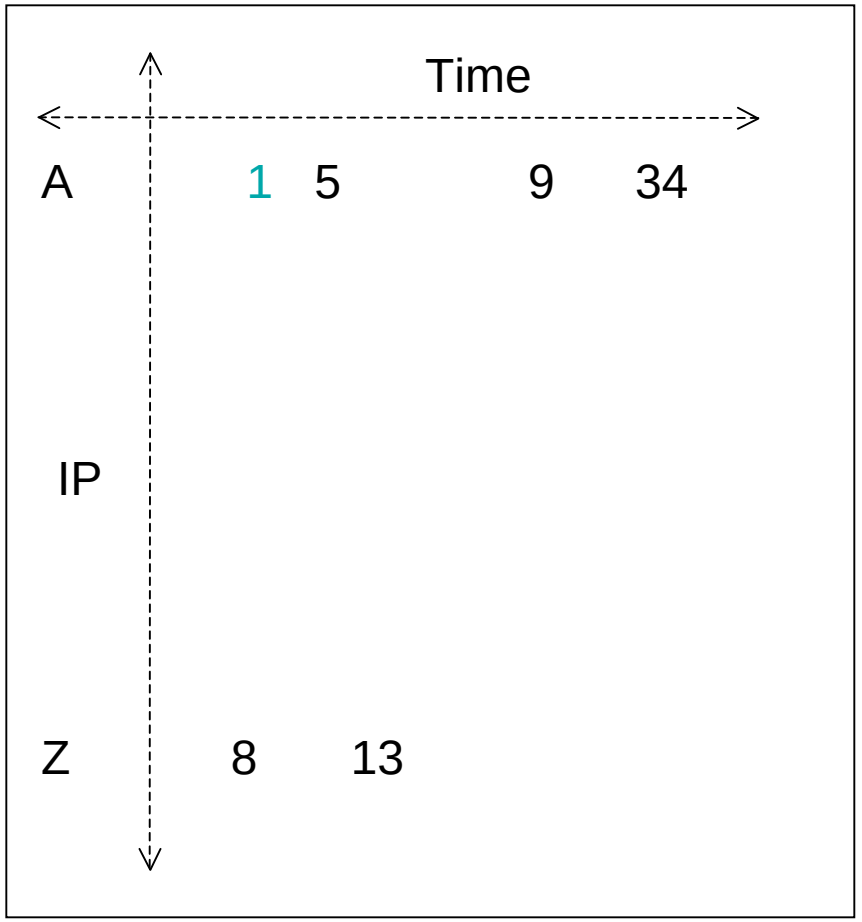


From Event Table to Event Plot

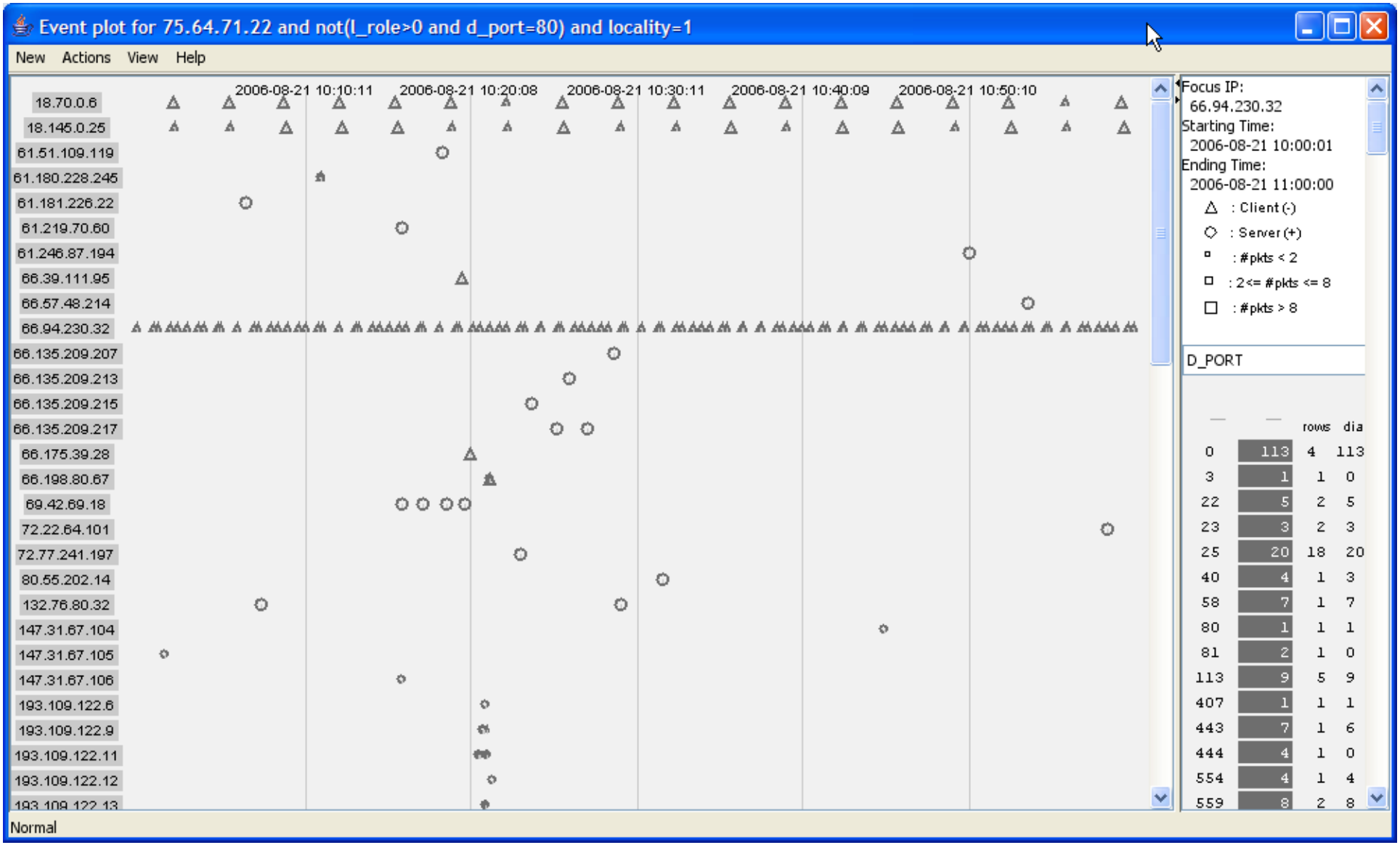
Event Table

1	Time	A	…	Measures
. .				
#	Time	IP	…	Measures
. .				
n	Time	Z	…	Measures

Event Plot

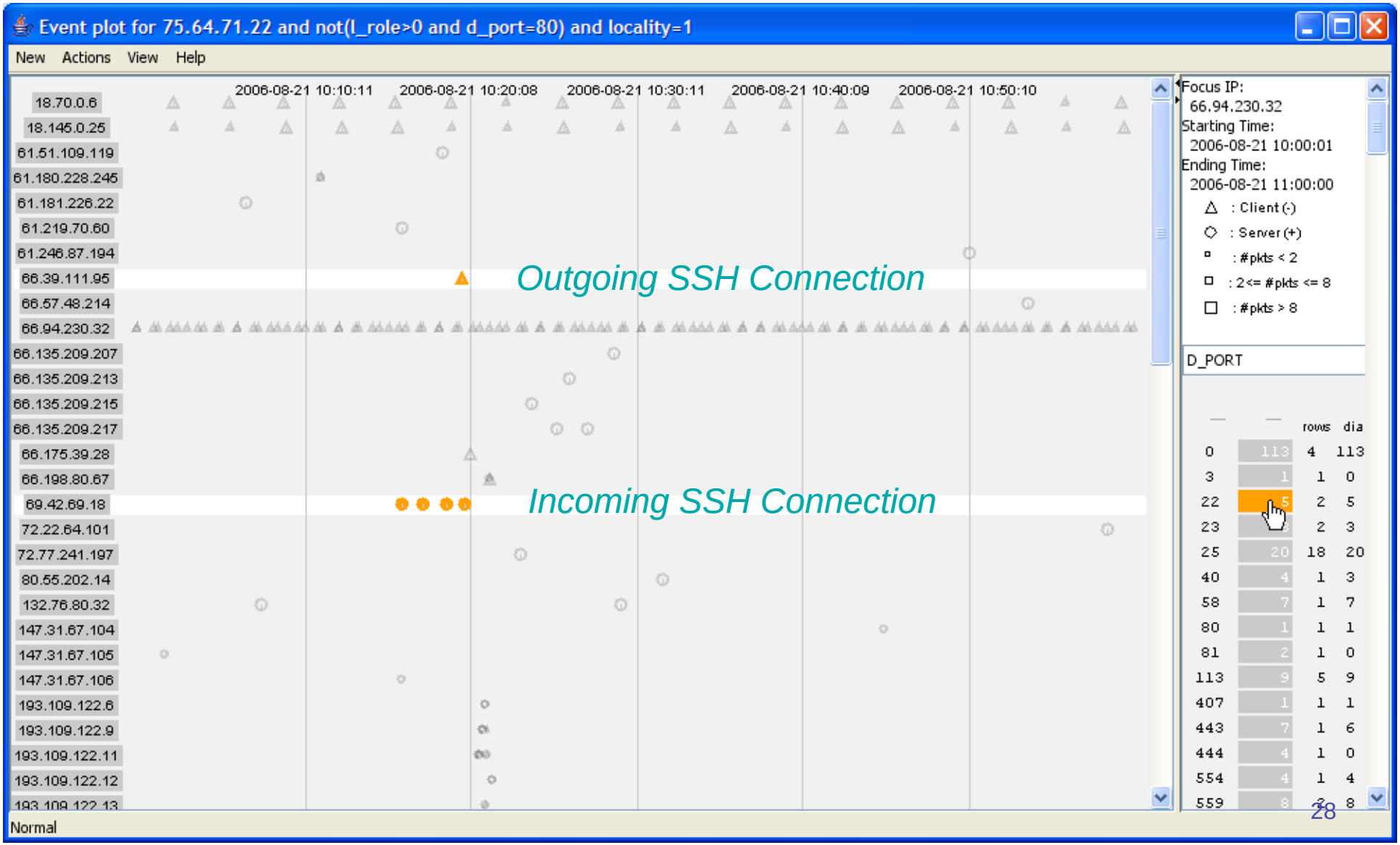


Event Plot

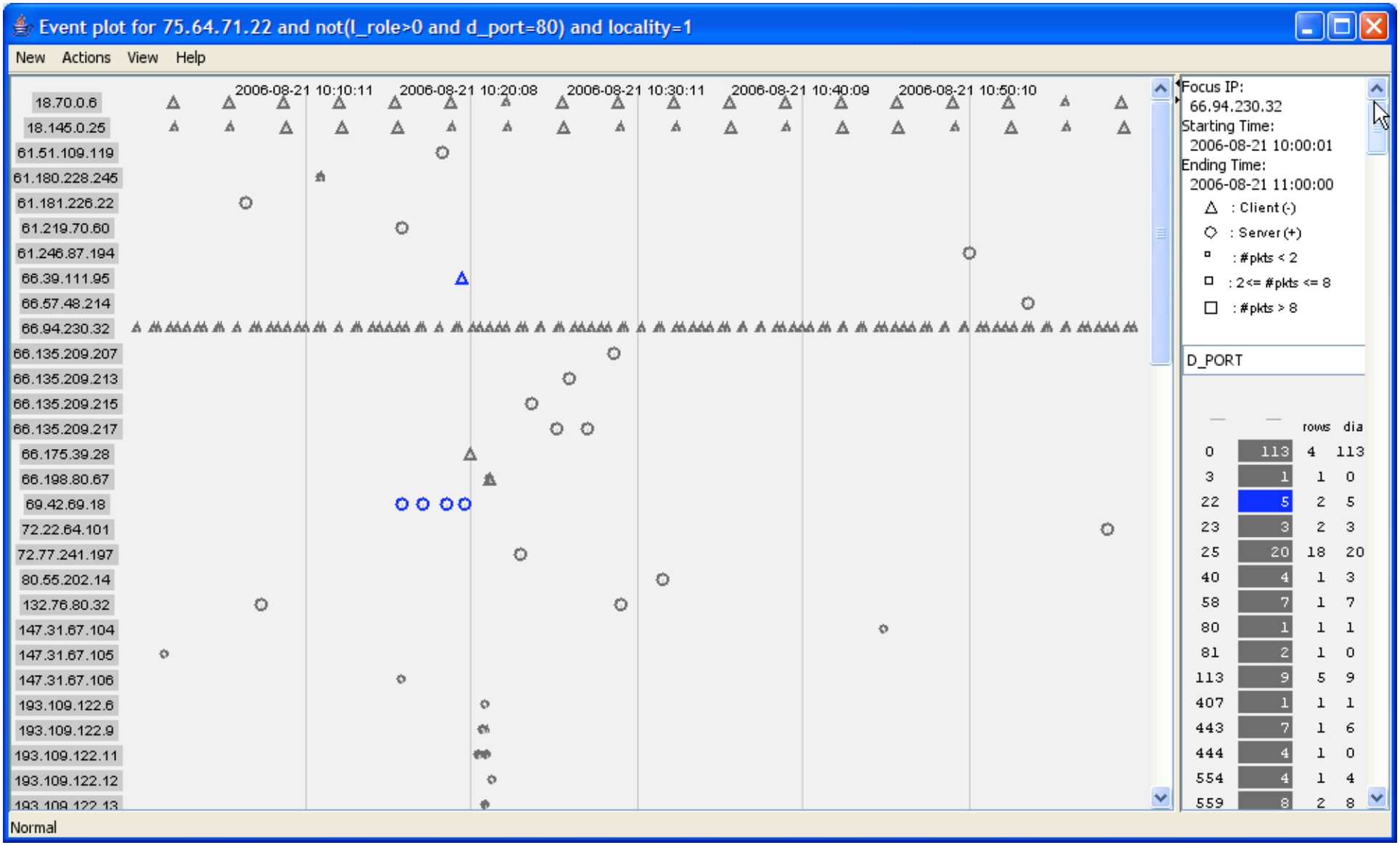

Event plot for 75.64.71.22 and not(l_role>0 and d_port=80) and locality=1
New
Actions
View
Help
2006-08-21 10:10:11
2006-08-21 10:20:08
2006-08-21 10:30:11
2006-08-21 10:40:09
2006-08-21 10:50:10
18.70.0.6
18.145.0.25
61.51.109.119
61.180.228.245
61.181.226.22
61.219.70.60
61.246.87.194
66.39.111.95
66.57.48.214
66.94.230.32
66.135.209.207
66.135.209.213
66.135.209.215
66.135.209.217
66.175.39.28
66.198.80.67
69.42.69.18
72.22.64.101
72.77.241.197
80.55.202.14
132.76.80.32
147.31.67.104
147.31.67.105
147.31.67.106
193.109.122.6
193.109.122.9
193.109.122.11
193.109.122.12
193.109.122.13
Focus IP:
66.94.230.32
Starting Time:
2006-08-21 10:00:01
Ending Time:
2006-08-21 11:00:00
: Client (-)
: Server (+)
: #pkts < 2
: 2<= #pkts <= 8
: #pkts > 8
D_PORT
rows
dia
0 113 4 113
3 1 1 0
22 5 2 5
23 3 2 3
25 20 18 20
40 4 1 3
58 7 1 7
80 1 1 1
81 2 1 0
113 9 5 9
407 1 1 1
443 7 1 6
444 4 1 0
554 4 1 4
559 8 2 8
Normal

Where is the actual intrusion?

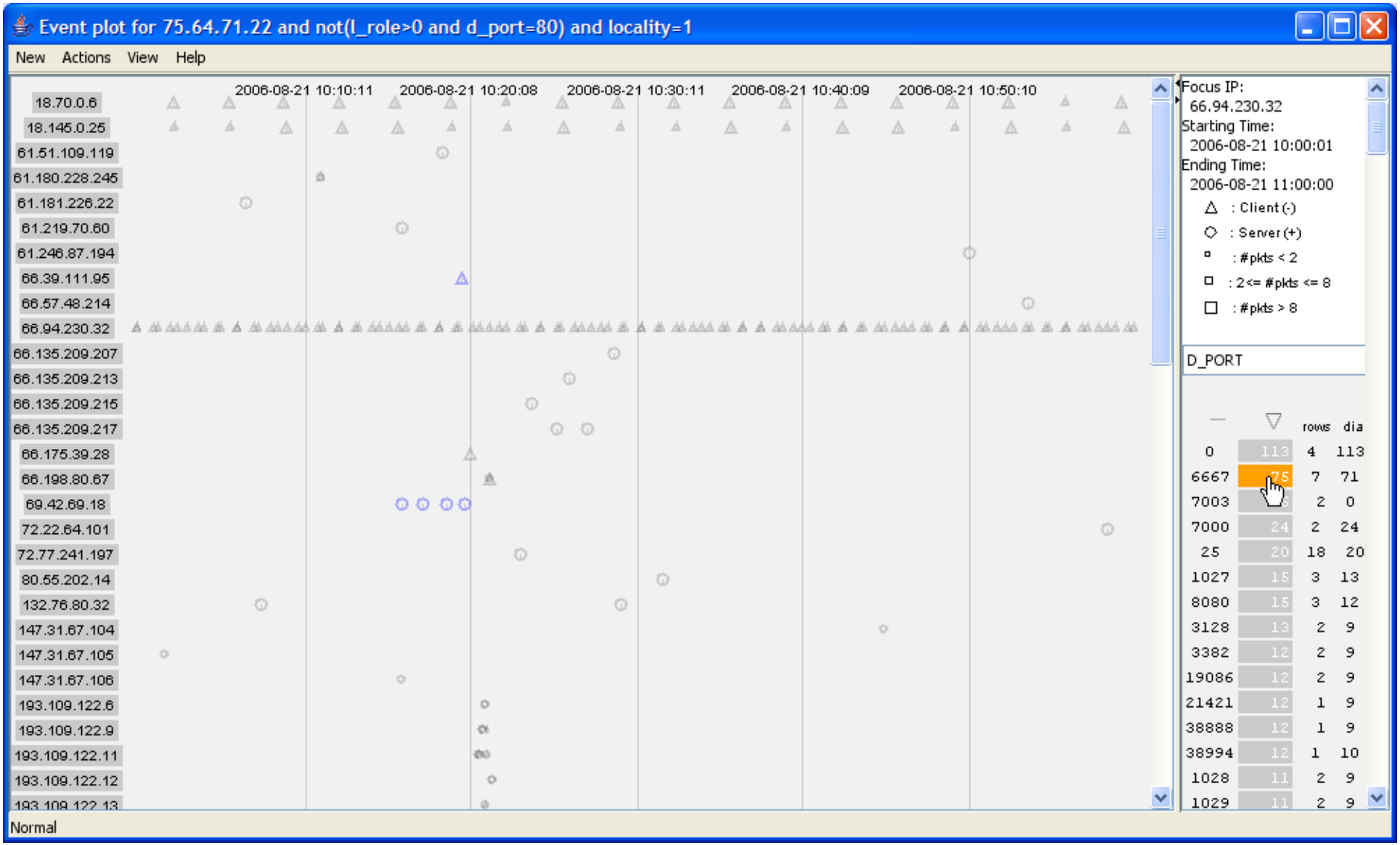
Initial SSH Connection



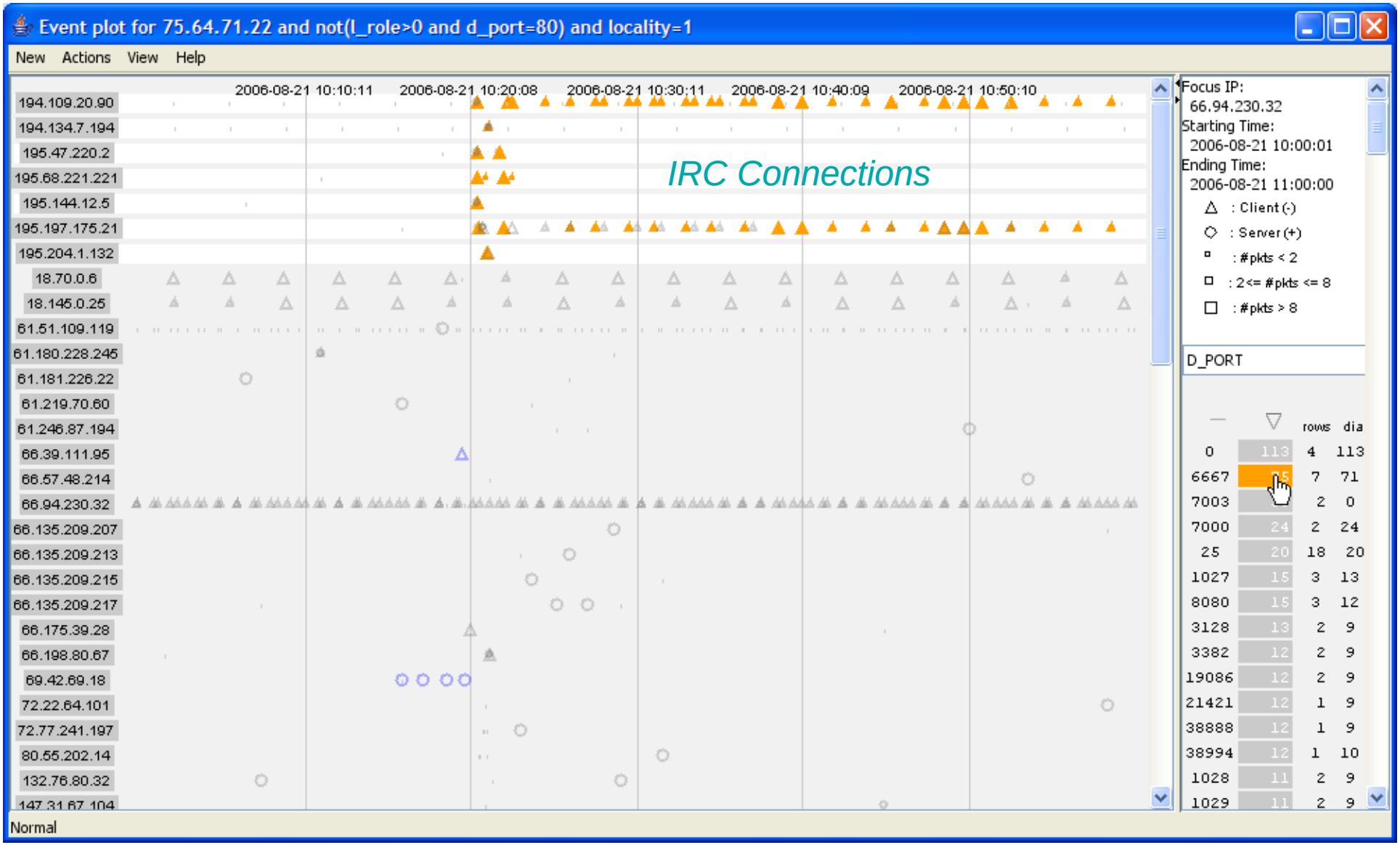
Initial SSH Connection



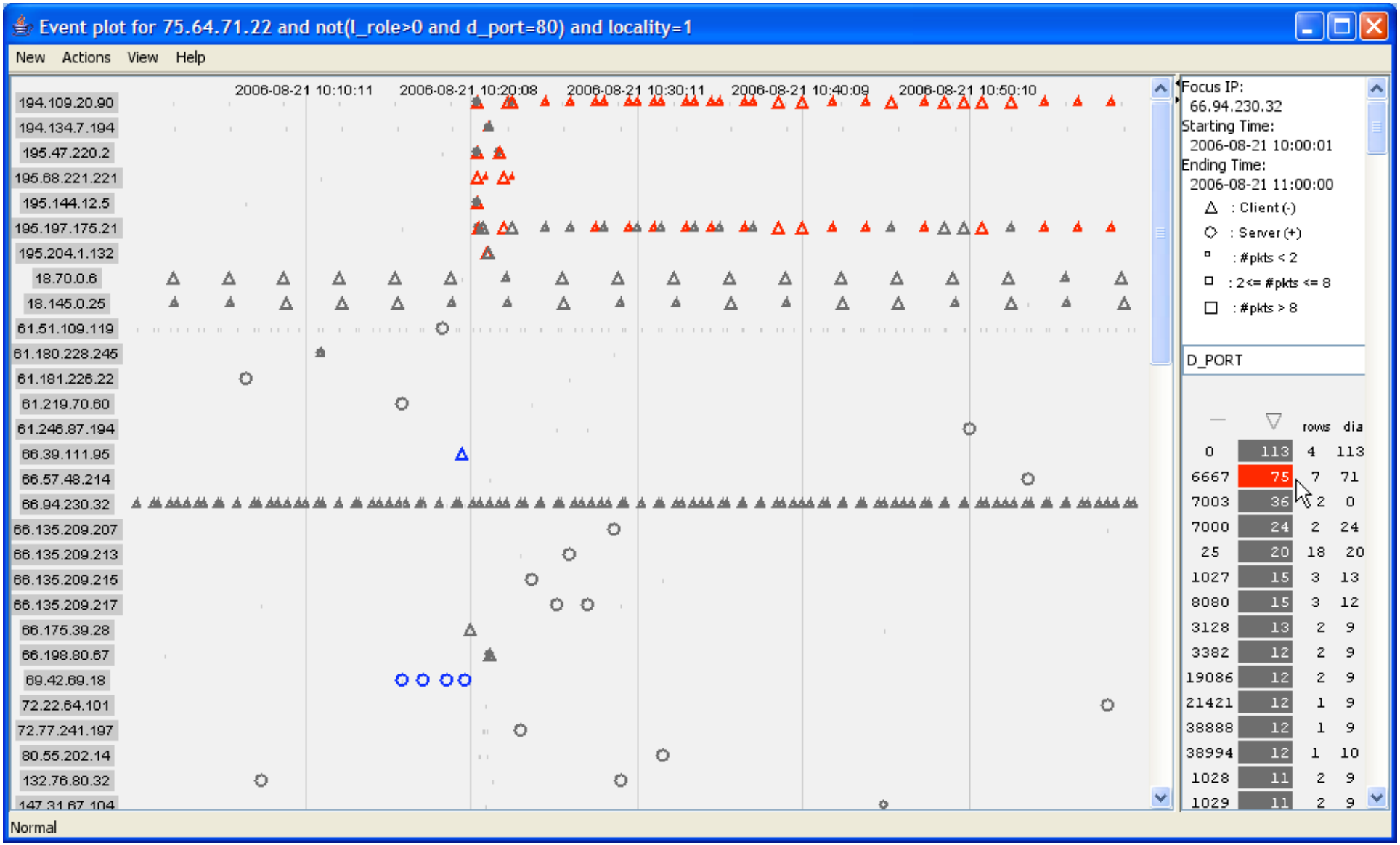
IRC Traffic on port 6667



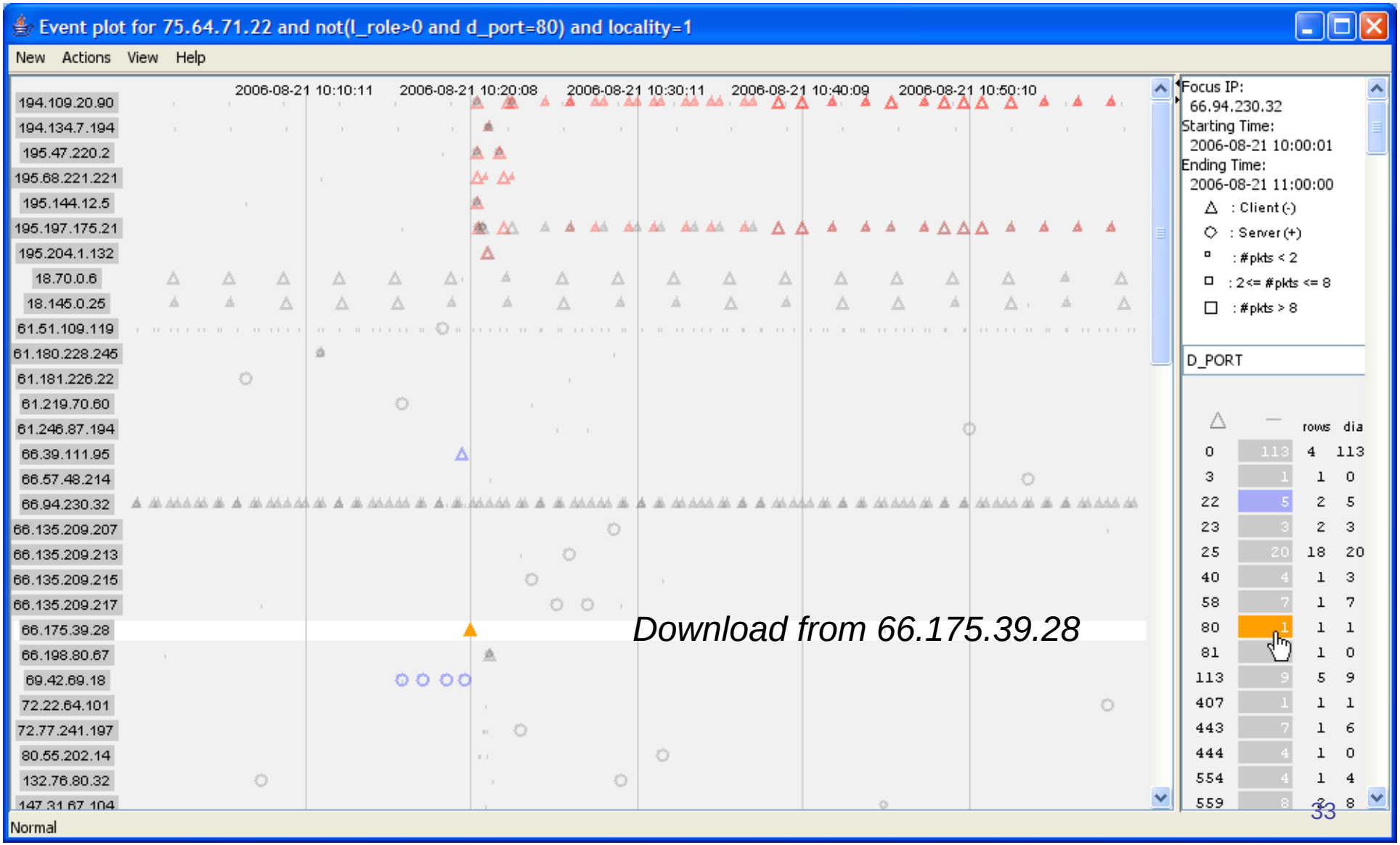
IRC Traffic on port 6667



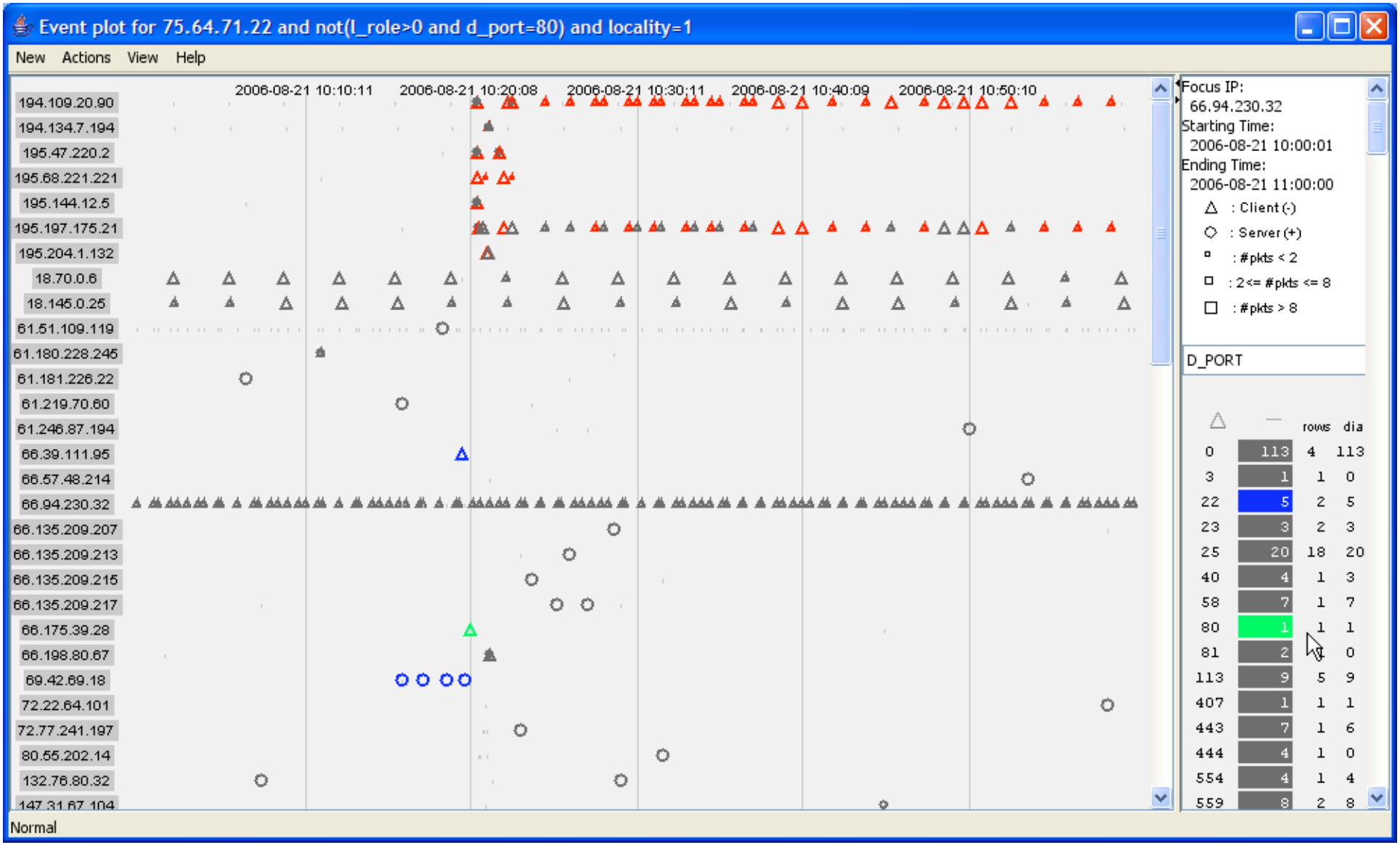
IRC Traffic on port 6667



Download of Exploit Tools



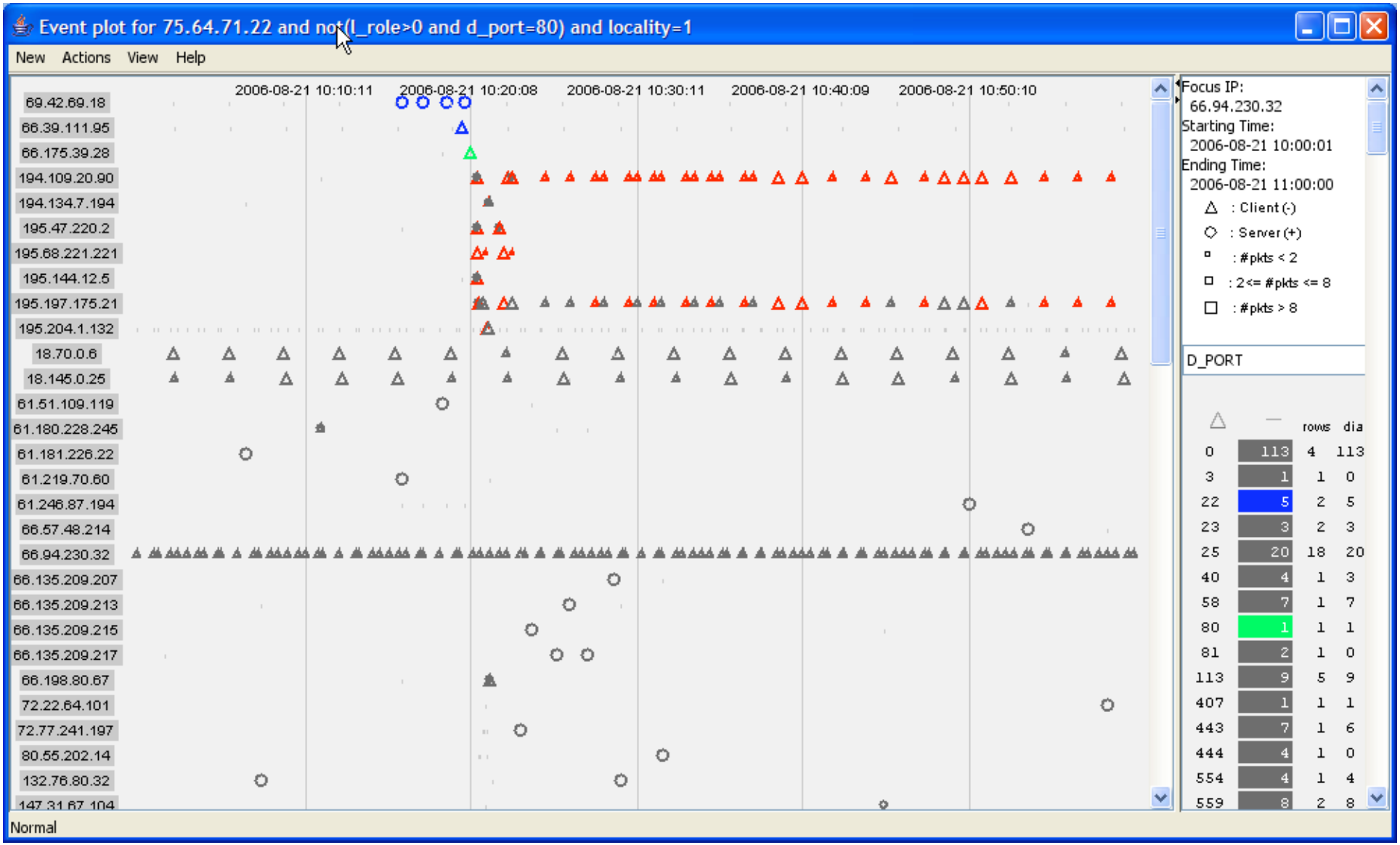
Download of Exploit Tools



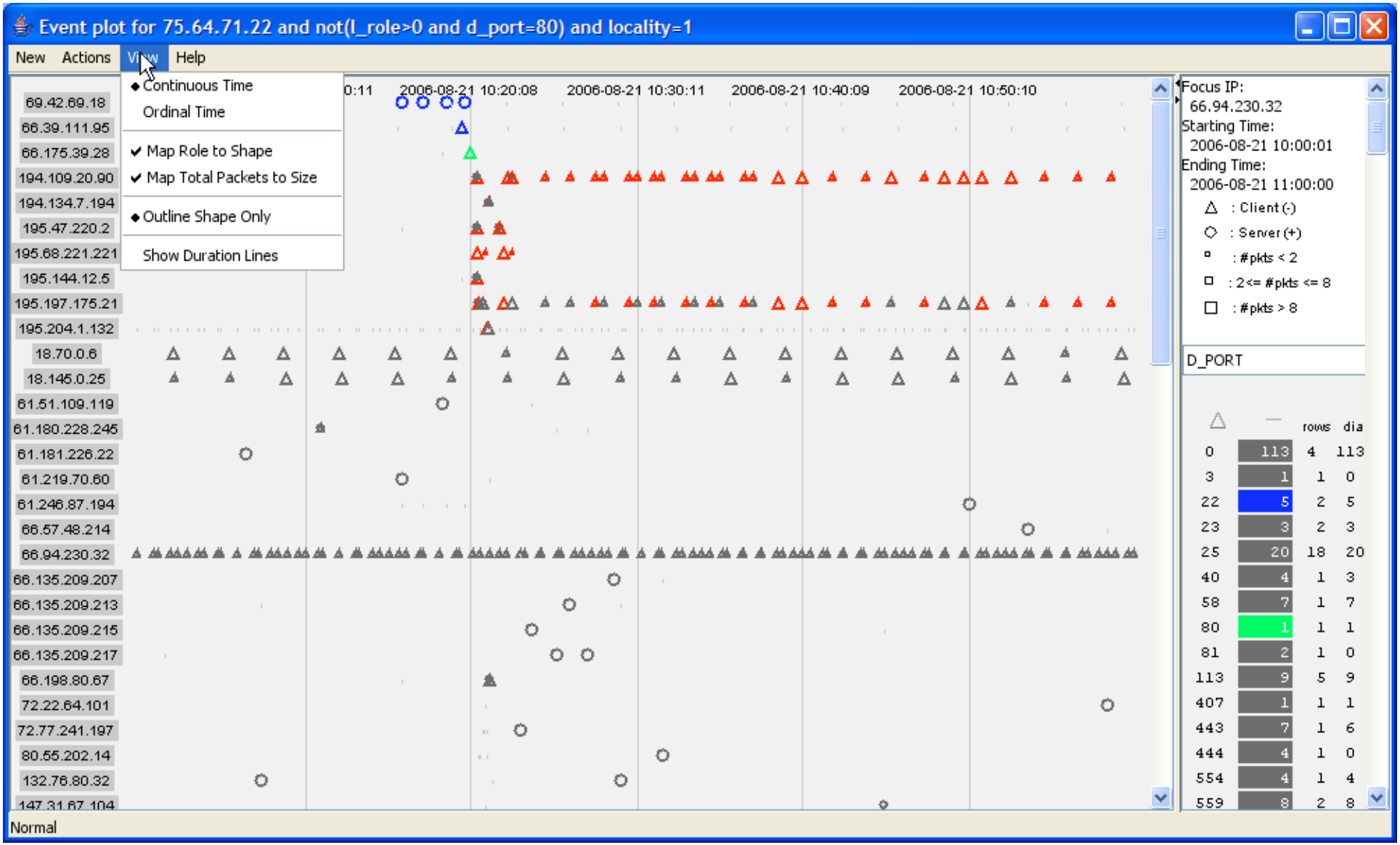
Reordering rows to tell story

- Initial SSH Connection
- Download of Exploit Tools
- IRC Servers Contacted

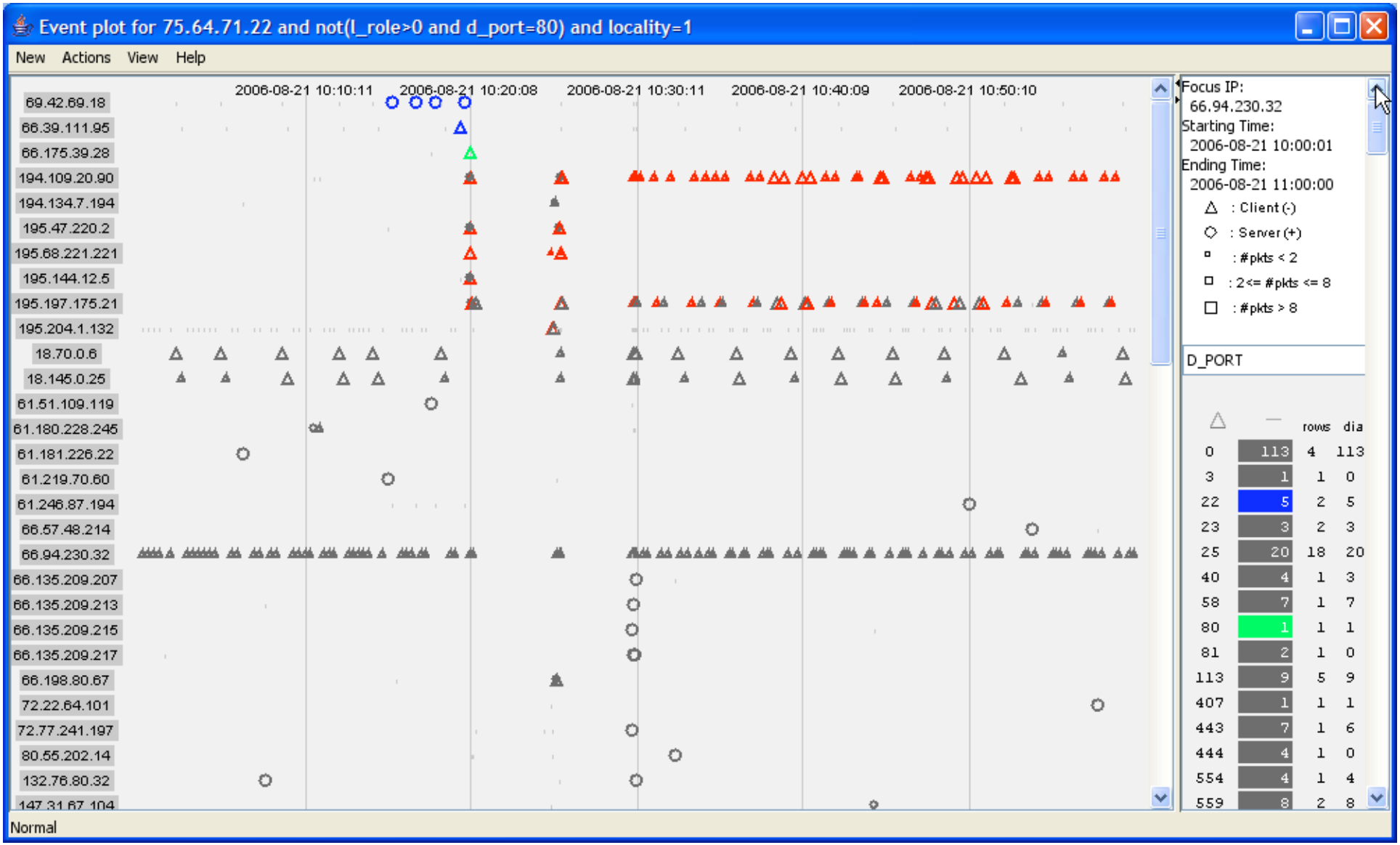
Reordered Rows



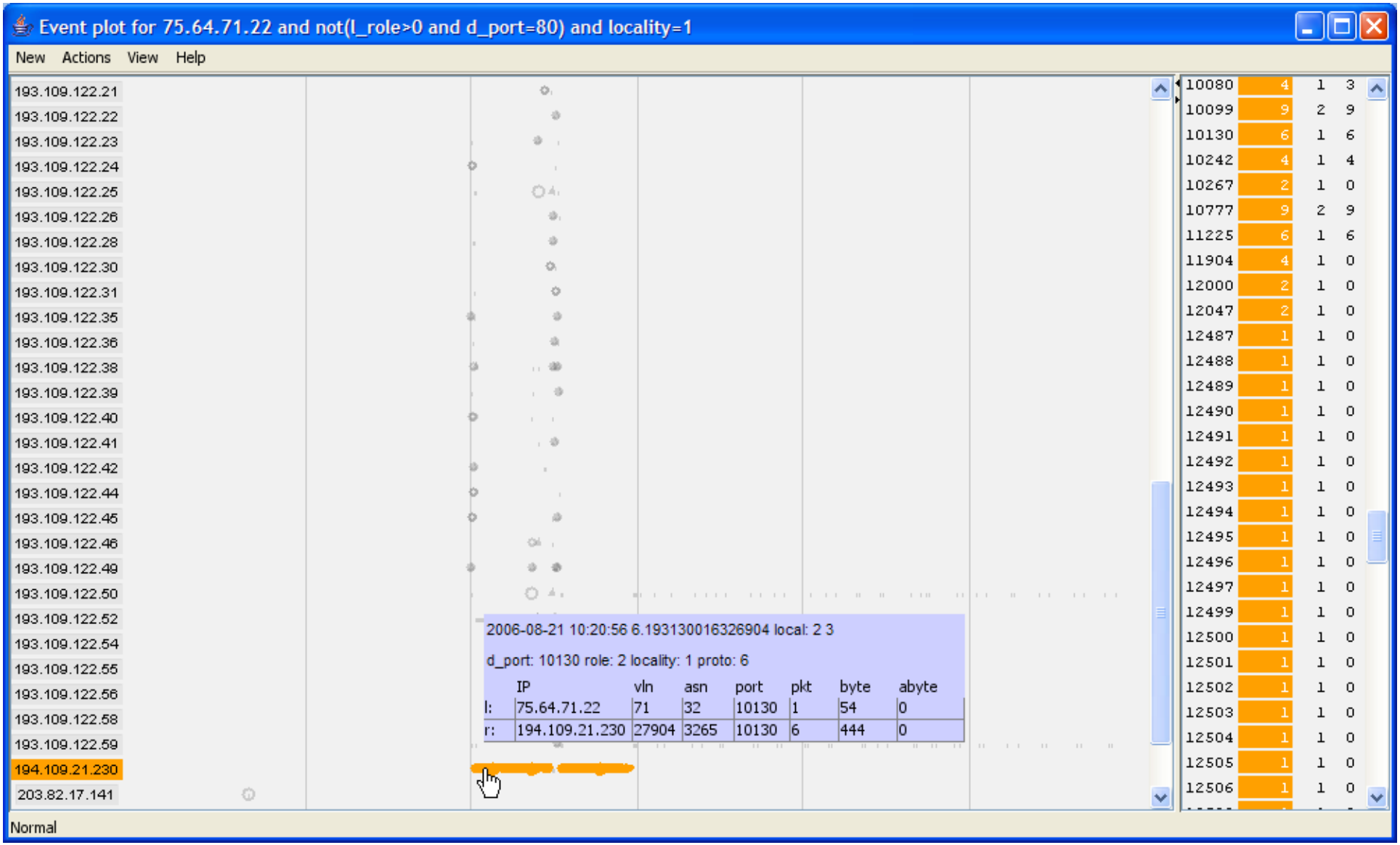
Switch to Ordinal Time



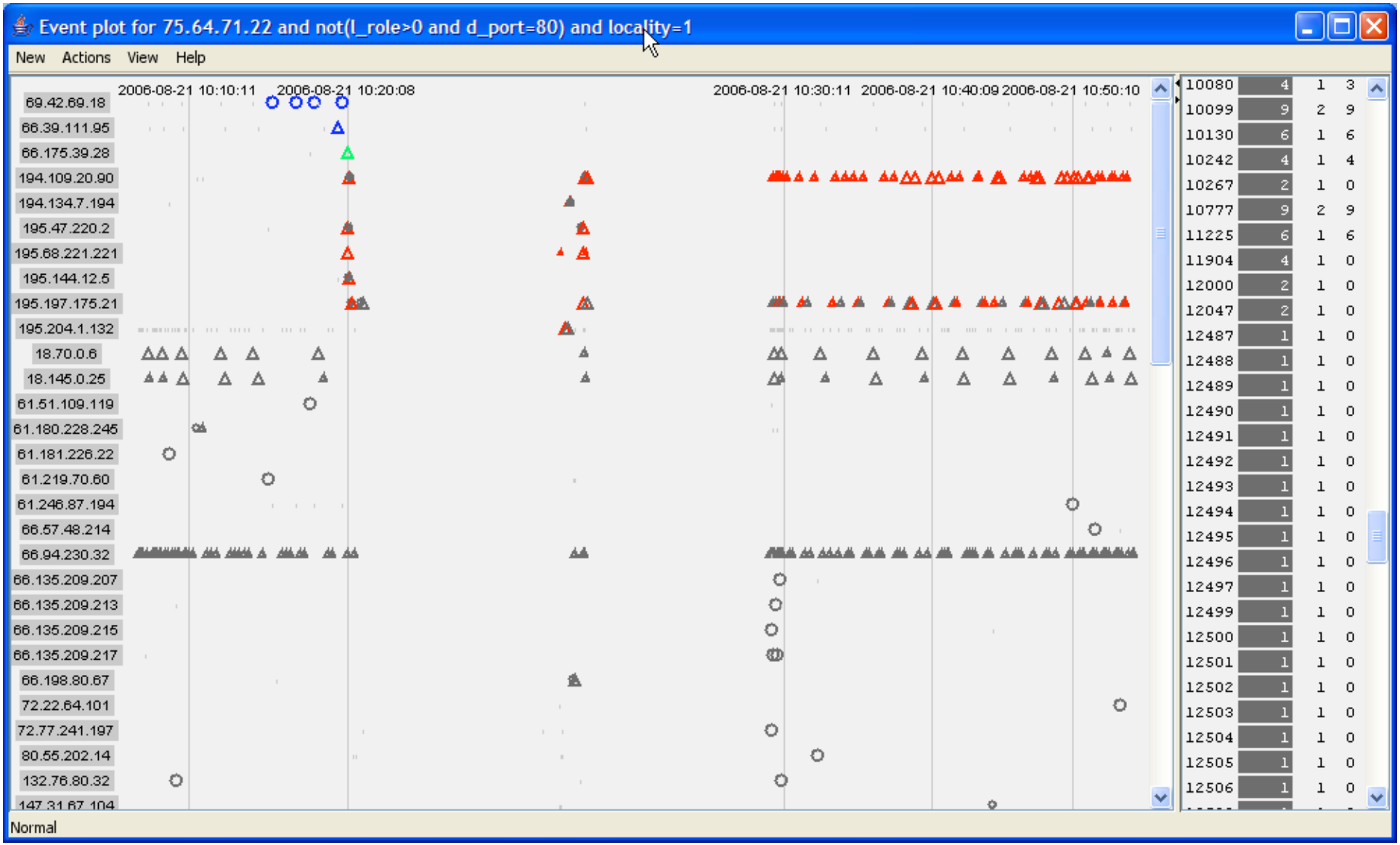
Switch to Ordinal Time



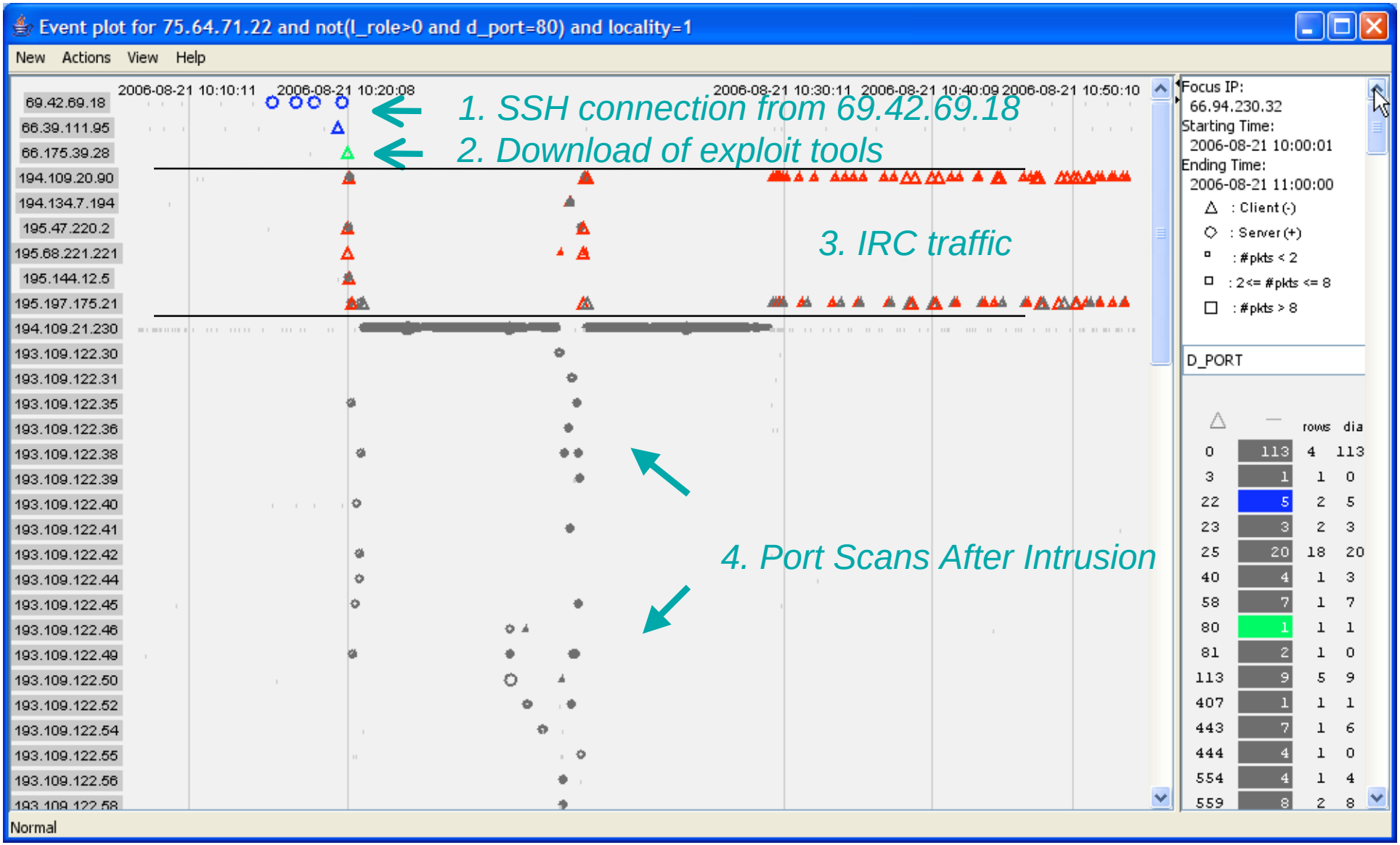
Mine the Gap


Event plot for 75.64.71.22 and not(l_role>0 and d_port=80) and locality=1
New
Actions
View
Help
193.109.122.21
193.109.122.22
193.109.122.23
193.109.122.24
193.109.122.25
193.109.122.26
193.109.122.28
193.109.122.30
193.109.122.31
193.109.122.35
193.109.122.36
193.109.122.38
193.109.122.39
193.109.122.40
193.109.122.41
193.109.122.42
193.109.122.44
193.109.122.45
193.109.122.46
193.109.122.49
193.109.122.50
193.109.122.52
193.109.122.54
193.109.122.55
193.109.122.56
193.109.122.58
193.109.122.59
194.109.21.230
203.82.17.141
2006-08-21 10:20:56 6.193130016326904 local: 2 3
d_port: 10130 role: 2 locality: 1 proto: 6
IP vln asn port pkt byte abyte
l: 75.64.71.22 71 32 10130 1 54 0
r: 194.109.21.230 27904 3265 10130 6 444 0
10080 4 1 3
10099 9 2 9
10130 6 1 6
10242 4 1 4
10267 2 1 0
10777 9 2 9
11225 6 1 6
11904 4 1 0
12000 2 1 0
12047 2 1 0
12487 1 1 0
12488 1 1 0
12489 1 1 0
12490 1 1 0
12491 1 1 0
12492 1 1 0
12493 1 1 0
12494 1 1 0
12495 1 1 0
12496 1 1 0
12497 1 1 0
12499 1 1 0
12500 1 1 0
12501 1 1 0
12502 1 1 0
12503 1 1 0
12504 1 1 0
12505 1 1 0
12506 1 1 0
Normal

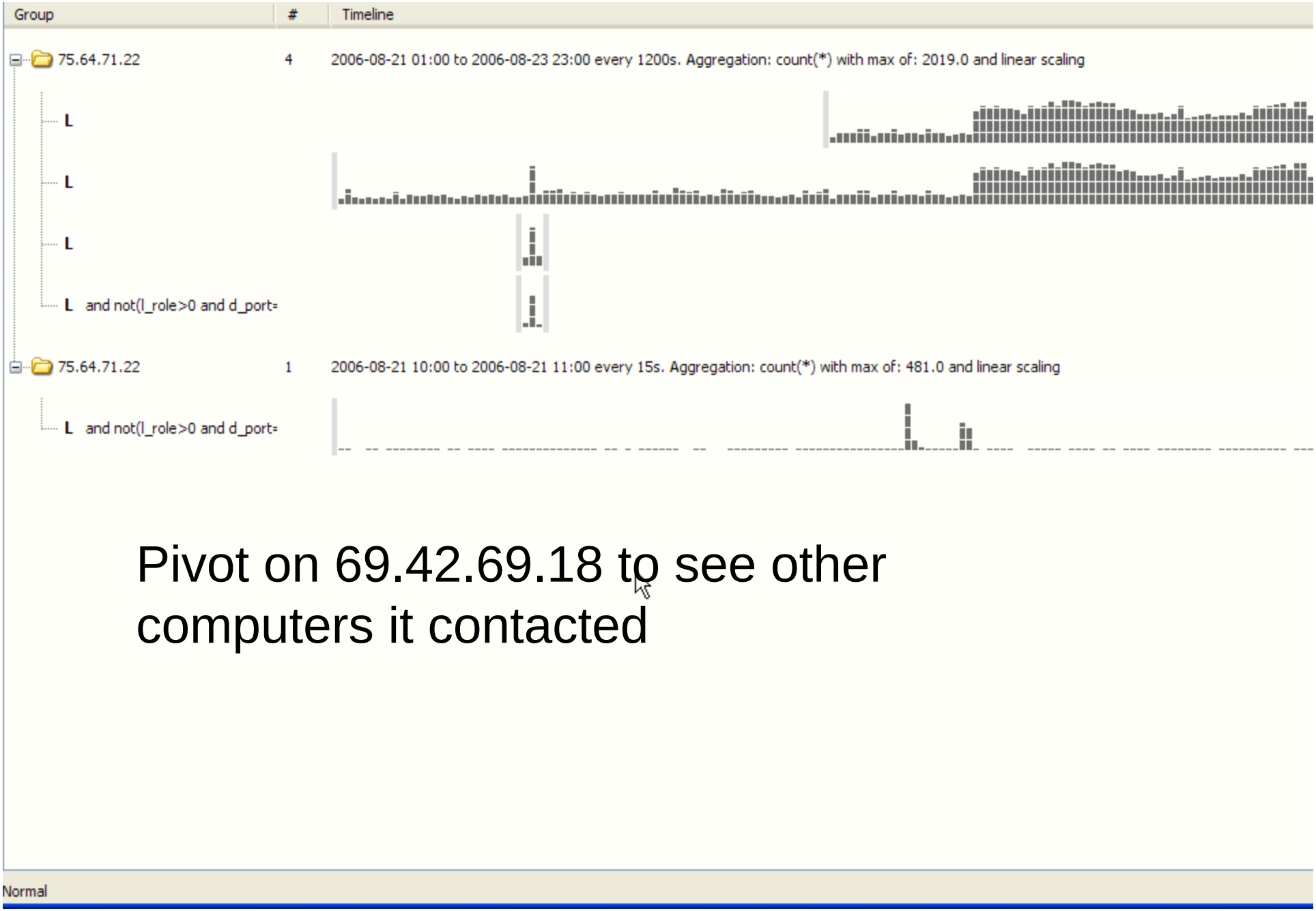
Explain the Gap



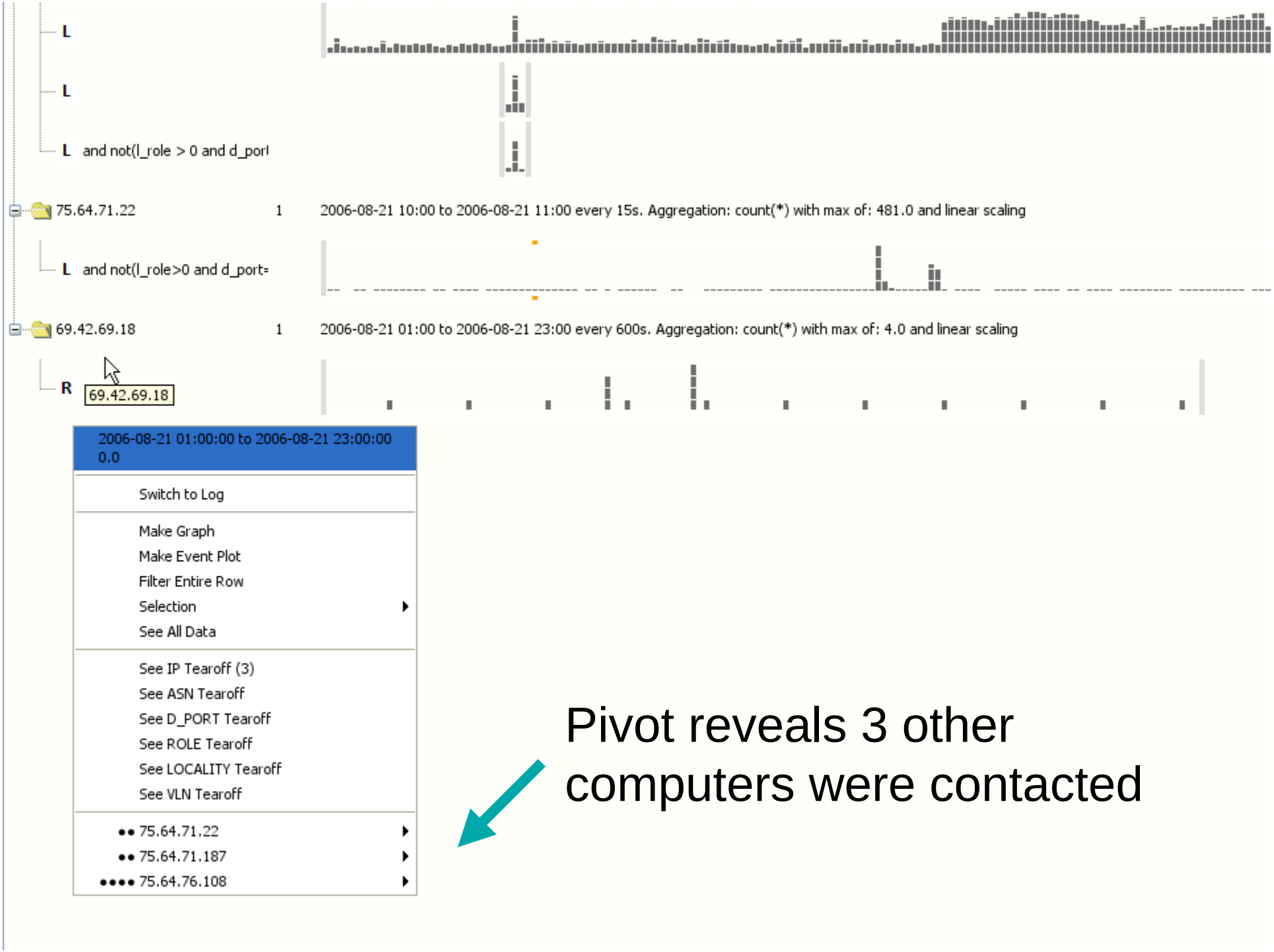
Intrusion Narrative



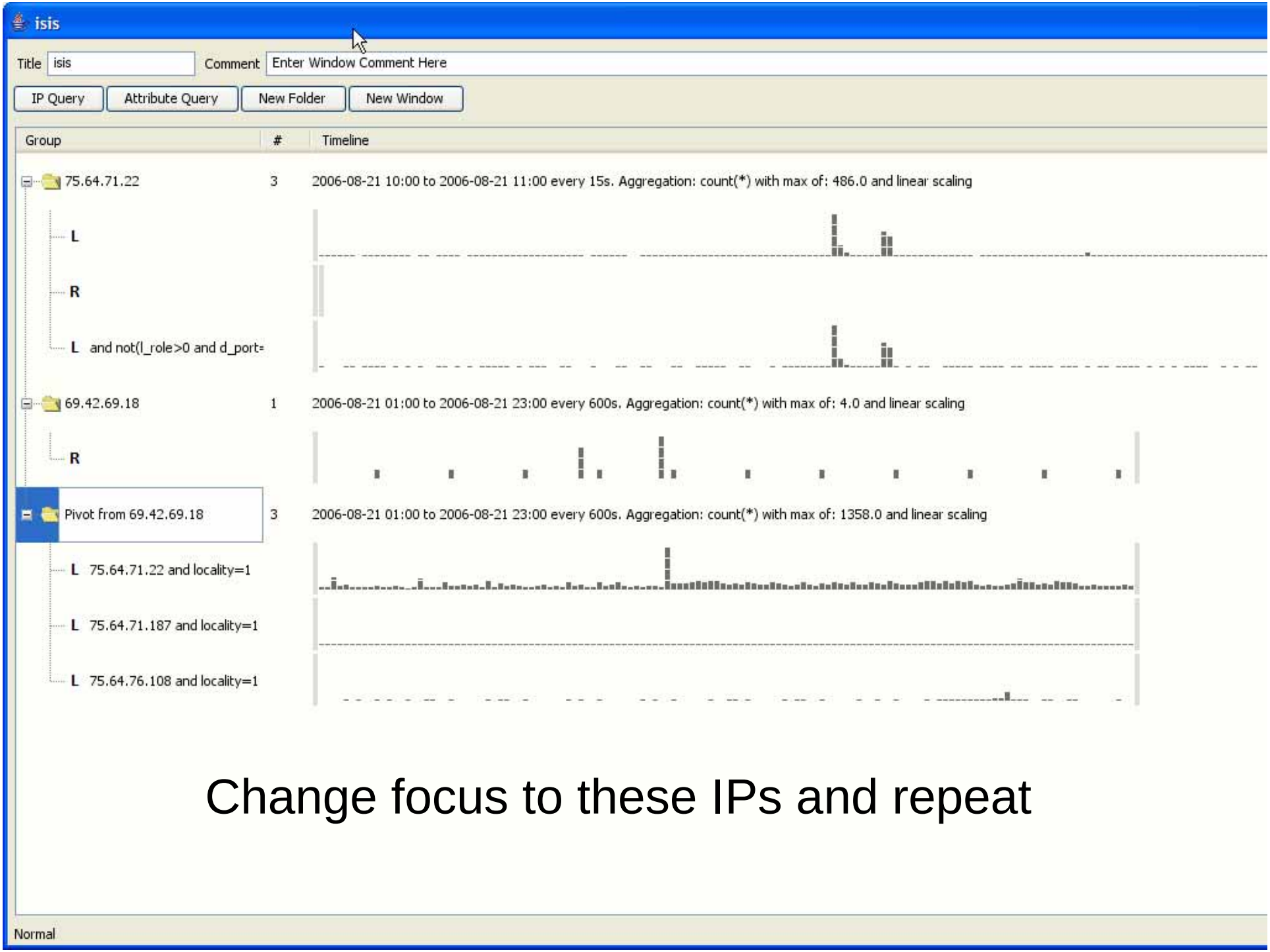
Were there other intrusions?



Pivot on 69.42.69.18 to see other computers it contacted



Pivot reveals 3 other computers were contacted



Change focus to these IPs and repeat

Feature Summary

- Timelines provide overview, zoom, and filter to show aggregate patterns in the data and provide an exploration history
- Tearoffs and brushing break down timelines along different dimensions to reveal relationships
- Text tables and pop-ups provide detail-on-demand
- Event plot provides detail for sequencing with visual grouping
- Reordering rows of timelines and event plots creates juxtapositions for comparisons and narrative
- Ordinal time emphasizes event sequences

Isis

- Future work
  - Pilot deployment at US-CERT
  - Combine with other tools
  - Query tables of a billion events at interactive speeds

- Acknowledgements
  - Stanford EE/CS Networking
  - DHS Science + Technology Directorate
  - National Visual Analytics Center

- Contact
  - John Gerth <gerth@graphics.stanford.edu>

The Analyst's Soup

- Disparate data sources
  - Traffic counters
  - Event logs
  - Flow records
  - Packet Traces
- Overlapping Tasks
  - *Perception*: Monitor
  - *Comprehension*: Explore and Investigate
  - *Projection*: Forecast and Present

Network Flows

- Advantages
  - Uniform and increasingly available
  - Mitigate privacy concerns
  - Hard to subvert
  - Largely insensitve to encryption
- Disadvantages
  - Still voluminous
  - Aggregate measure
  - Lack Content

Data Infrastructure

- Network Flow Sensor
  - Span ports from two backbone switches
  - See all layer 3 traffic for three buildings
  - 20-30M Argus flows/day (~1GB compressed)
- MySQL Database
  - Transform src/dst to local/remote
  - Convenience columns for locality, role, dst port
  - Index most dimensions (adds about 50%)
  - Tables + indices ~2GB/day