Happy Packets: Some Initial Results

NANOG/SF 2004.05.24

Randy Bush <randy@psg.com>
Timothy G. Griffin <tim.griffin@intel.com>
Z. Morley Mao <zmao@eecs.umich.edu>
Eric Purpus <epurpus@cs.uoregon.edu>
Dan Stutsbach <dstutsba@cs.uoregon.edu>

<http://psg.com/~randy/040524.nanog-happy.pdf>

Central Question

- What is the relationship between **control plane** instability and **data plane** instability?

- Related Questions:
  - Is the quantity of BGP updates good or bad?
  - Who wants to see zero BGP updates?

Internet Weather

We frequently hear comments such as

- Internet routing is fragile, collapsing, ...,
- BGP is broken or is not working well,
- Day X was a bad routing day on the internet,
- Change X to protocol Y will improve routing,
- Etc.

And we often measure routing dynamics and say that some measurement is better or worse than another

Internet [Routing] Instability

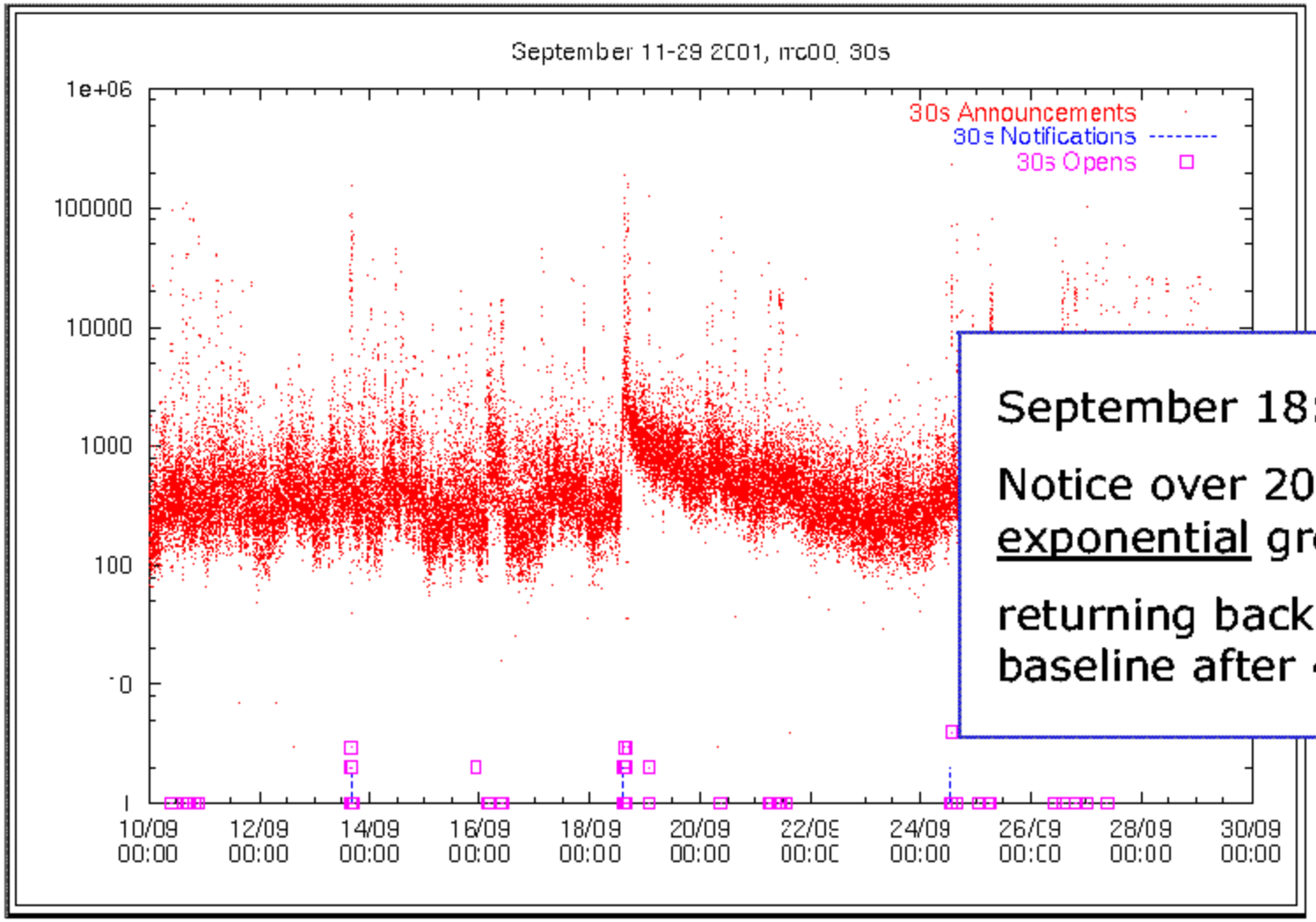
- We are told that a lot of BGP updates is equated with internet instability
- “There are too many BGP updates, so BGP must be broken.”

White Blood Cells

- Perhaps BGP announcements are like white blood cells
- Their presence may signal a problem
- But they are often part of the cure, not necessarily part of the problem

A view on content of the same messages

Number of prefix announcements in 30 sec intervals



September 18:

Notice over 20-fold exponential growth

returning back to baseline after 4 days!

renesys

Big Events

- The Renesys folk and others are looking at big events
- We are at looking at single announcements
- So our work does not contradict Renesys, but it does suggest we consider some of the assumptions (see Lan Wang et alia)
- And we are measuring data plane performance waiting for the next big event

Routing Quality

- But what is *good* routing? How can we say one measurement shows routing is better than another unless we have **metrics** for routing quality?
- We often work on the assumption that number of prefixes, speed or completeness of convergence, etc. are measures of routing quality

Happy Packets

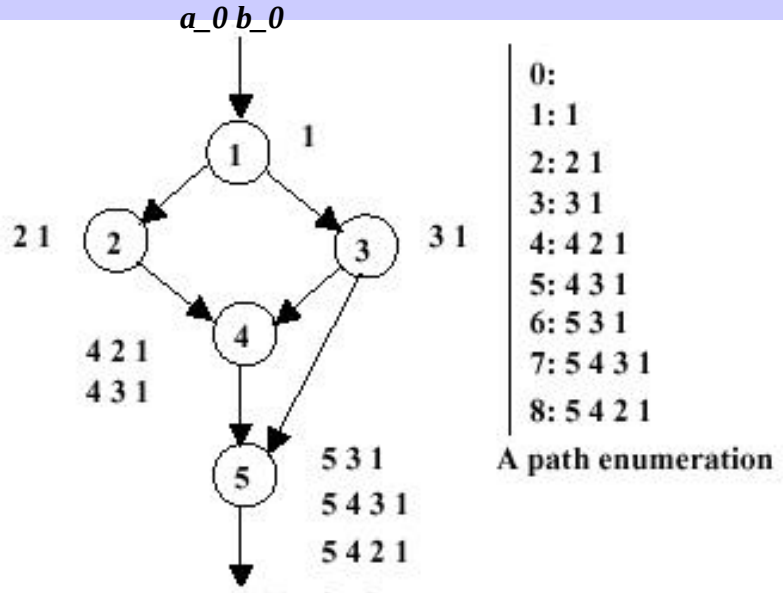
- The measure which counts is whether the users' packets reach their destination
- If the users' packets are happy, the routing system, and other components, are doing their job
- We call these ***Happy Packets***
- There are well-known metrics for the data plane, Delay, Drop, Jitter, and Reordering
- So we set out to measure Control Plane quality by measuring the Data Plane

Router Scaling

- While data plane performance is the goal, we can't have routers falling over processing chatty BGP
- But, as long as network BGP growth increases load on the routers below Moore's law, it is not clear we are in danger

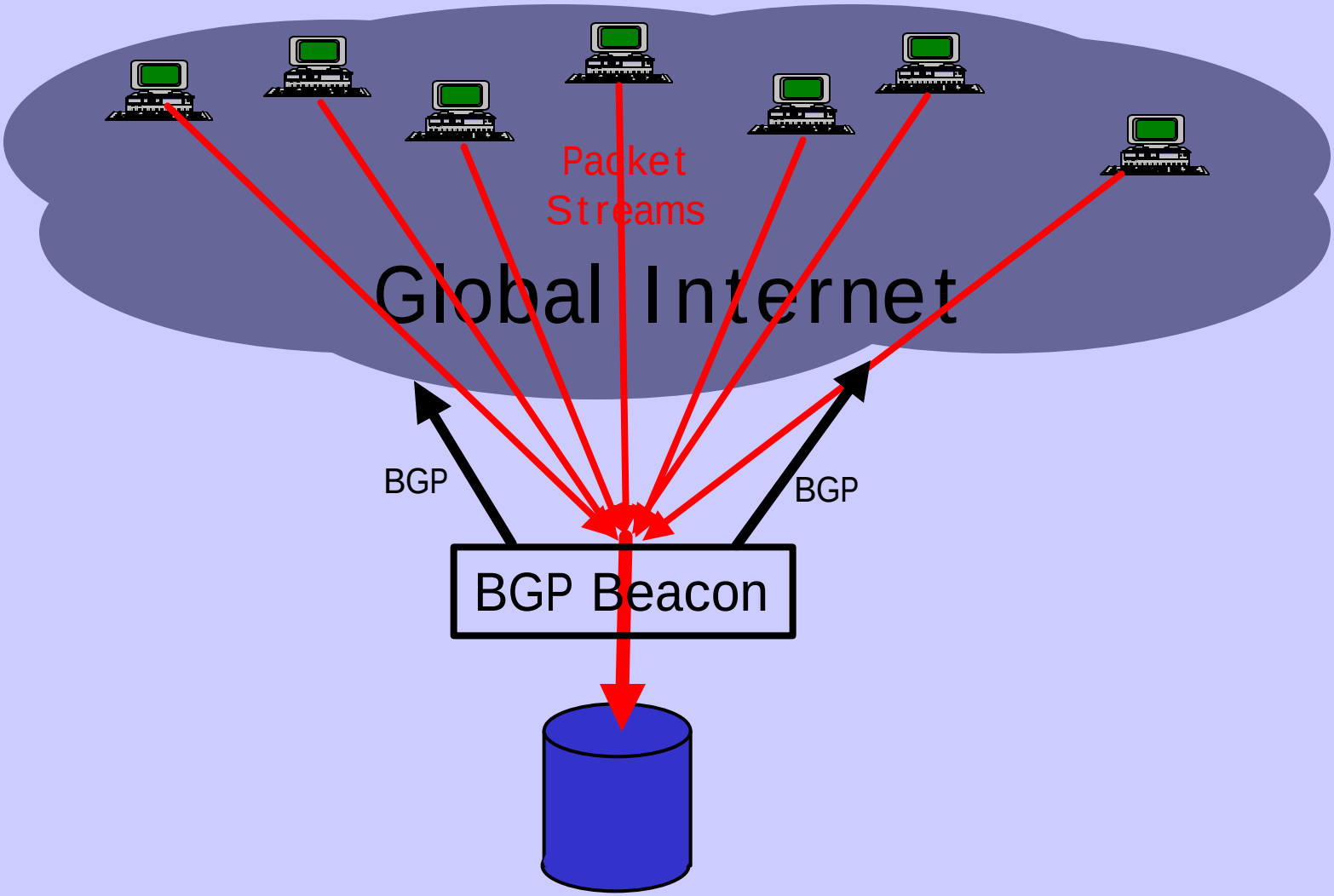
Why I'm Going Crazy Trying to Interpret those BGP Updates?

It is easy to construct a 5 node BGP system where a simple Announce/Withdraw signal (a_0 b_0) at one node can produce any of these 52 output signals at another...



a_6b_6	$a_7a_6b_6$	$a_8a_6b_6$	$a_8a_6a_8b_8a_7b_7$
$a_6a_7b_7$	$a_7a_6a_7b_7$	$a_8a_6a_7b_7$	$a_8a_7a_6a_7a_8b_8$
$a_6a_8b_8$	$a_7a_6a_8b_8$	$a_8a_6a_8b_8$	$a_8a_7a_8a_6a_8b_8$
$a_6a_7a_8b_8$	$a_7a_8a_6b_6$	$a_8a_7a_6b_6$	$a_8a_7a_8b_8a_6b_6$
$a_6a_8a_7b_7$	$a_7b_7a_6b_6$	$a_8b_8a_6b_6$	$a_8b_8a_6b_6a_7b_7$
$a_6b_6a_7b_7$	$a_7a_6a_7a_8b_8$	$a_8a_6a_7a_8b_8$	$a_8b_8a_7a_6a_7b_7$
$a_6b_6a_8b_8$	$a_7a_6b_6a_8b_8$	$a_8a_6a_8a_7b_7$	$a_8b_8a_7b_7a_6b_6$
$a_6a_7b_7a_8b_8$	$a_7a_8a_6a_8b_8$	$a_8a_6b_6a_7b_7$	
$a_6a_8a_7a_8b_8$	$a_7a_8b_8a_6b_6$	$a_8a_7a_6a_7b_7$	
$a_6a_8b_8a_7b_7$	$a_7b_7a_6a_8b_8$	$a_8a_7a_6a_8b_8$	
$a_6b_6a_7a_8b_8$	$a_7b_7a_8a_6b_6$	$a_8a_7a_8a_6b_6$	
$a_6b_6a_8a_7b_7$	$a_7a_6a_7b_7a_8b_8$	$a_8a_7b_7a_6b_6$	
$a_6b_6a_7b_7a_8b_8$	$a_7b_7a_6b_6a_8b_8$	$a_8b_8a_6a_7b_7$	
$a_6b_6a_8a_7a_8b_8$	$a_7b_7a_8a_6a_8b_8$	$a_8b_8a_7a_6b_6$	
$a_6b_6a_8b_8a_7b_7$	$a_7b_7a_8b_8a_6b_6$	$a_8a_6a_8a_7a_8b_8$	

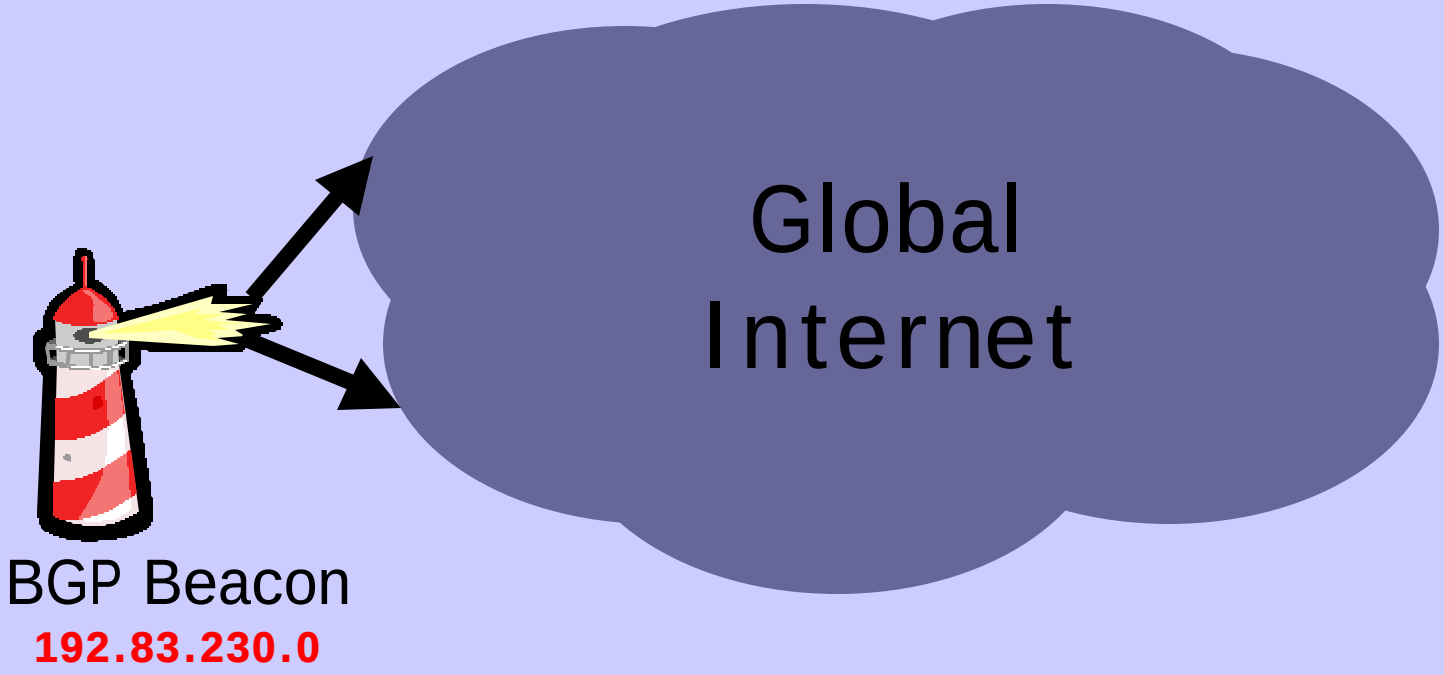
Experimental Setup



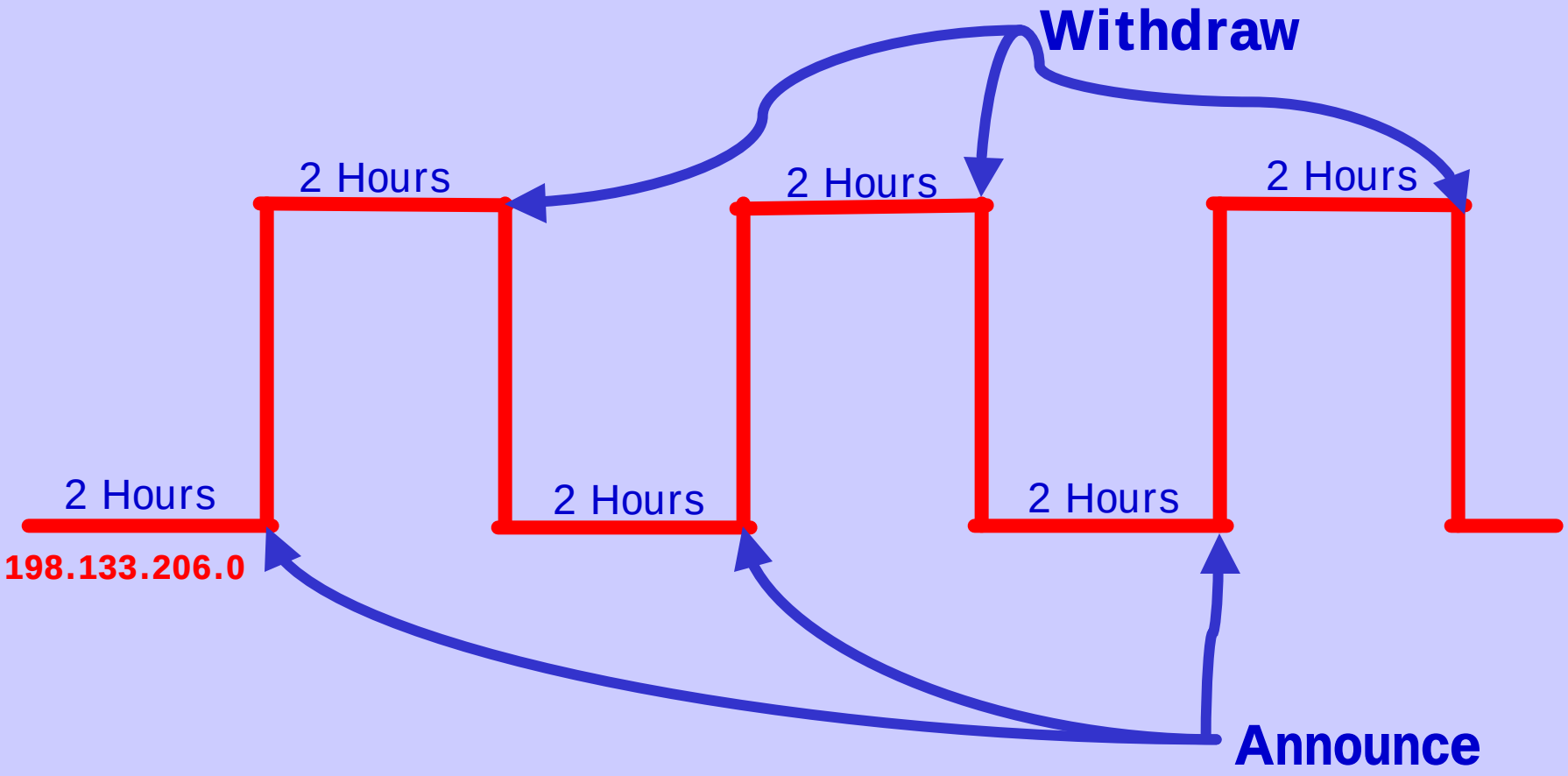
BGP Beacon

BGP Beacon:

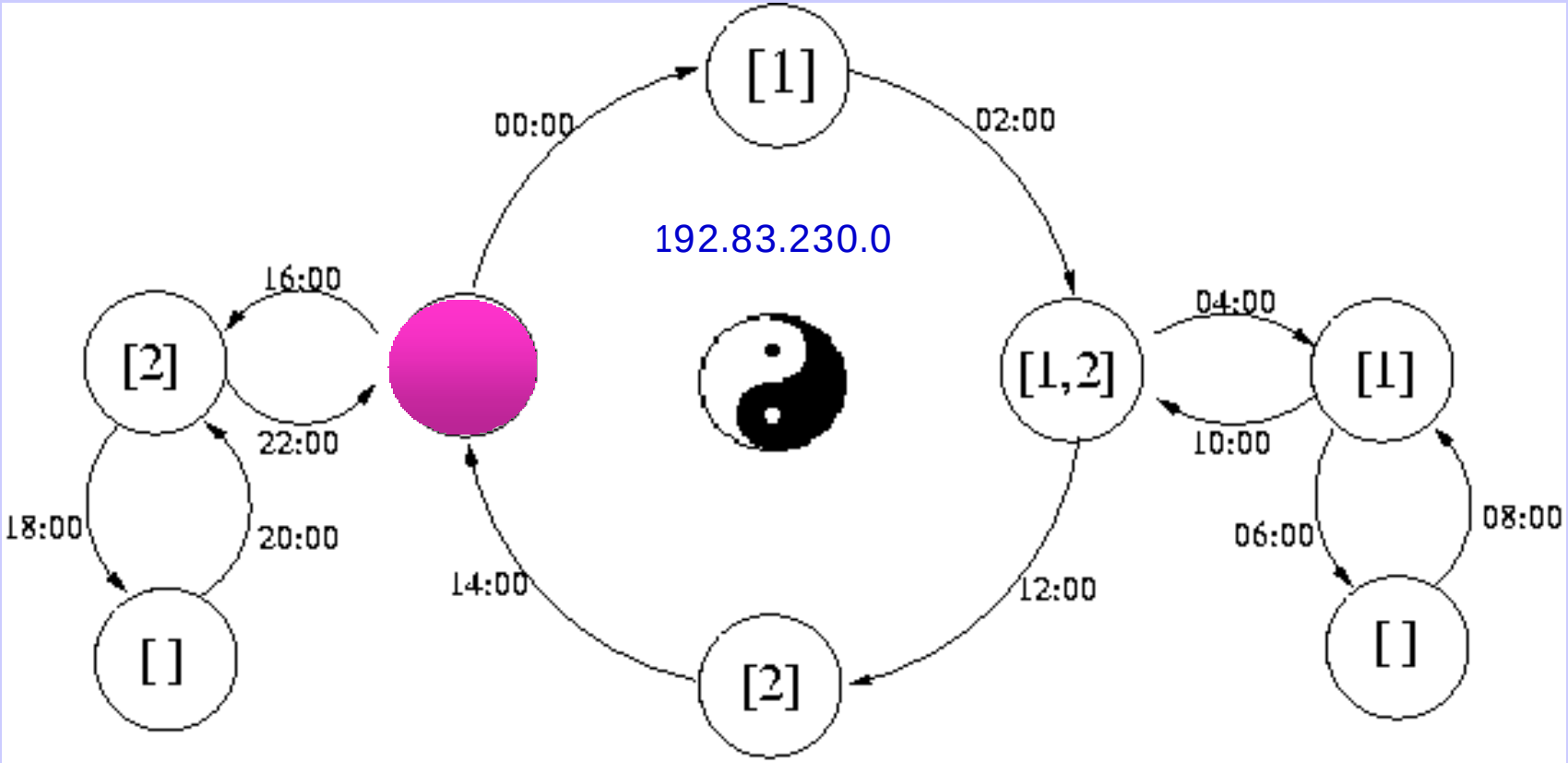
A prefix that is Announced and Withdrawn at **well-known** times



BGP Beacons
Announce & Withdraw

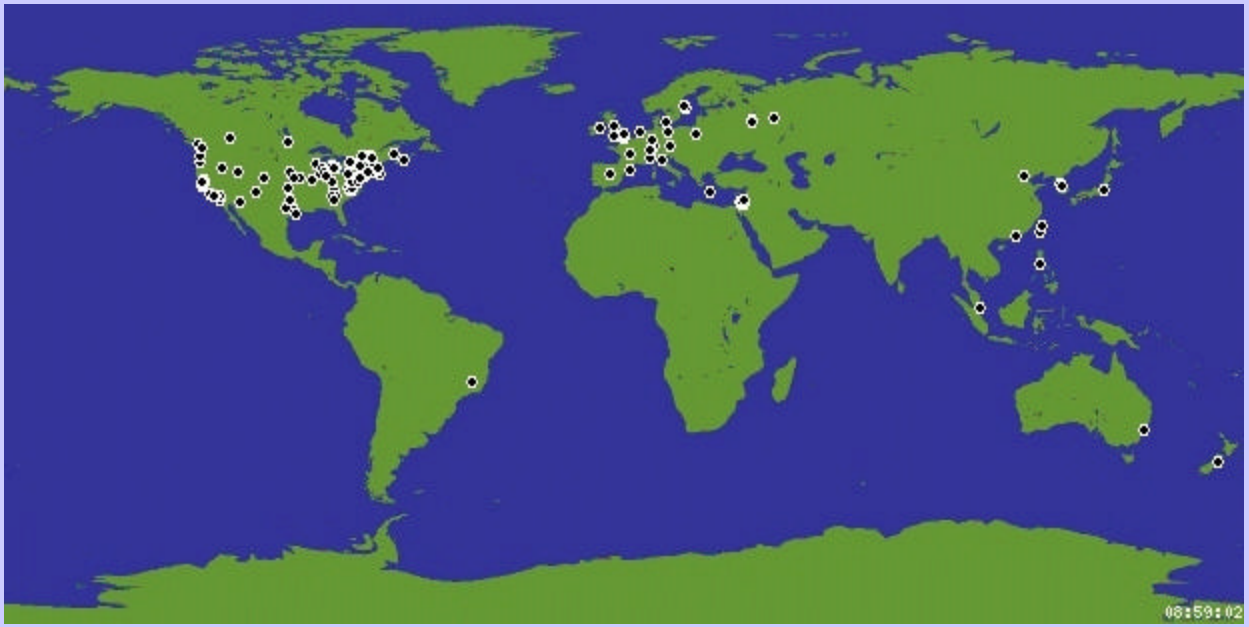


Multi-Homed Beacon

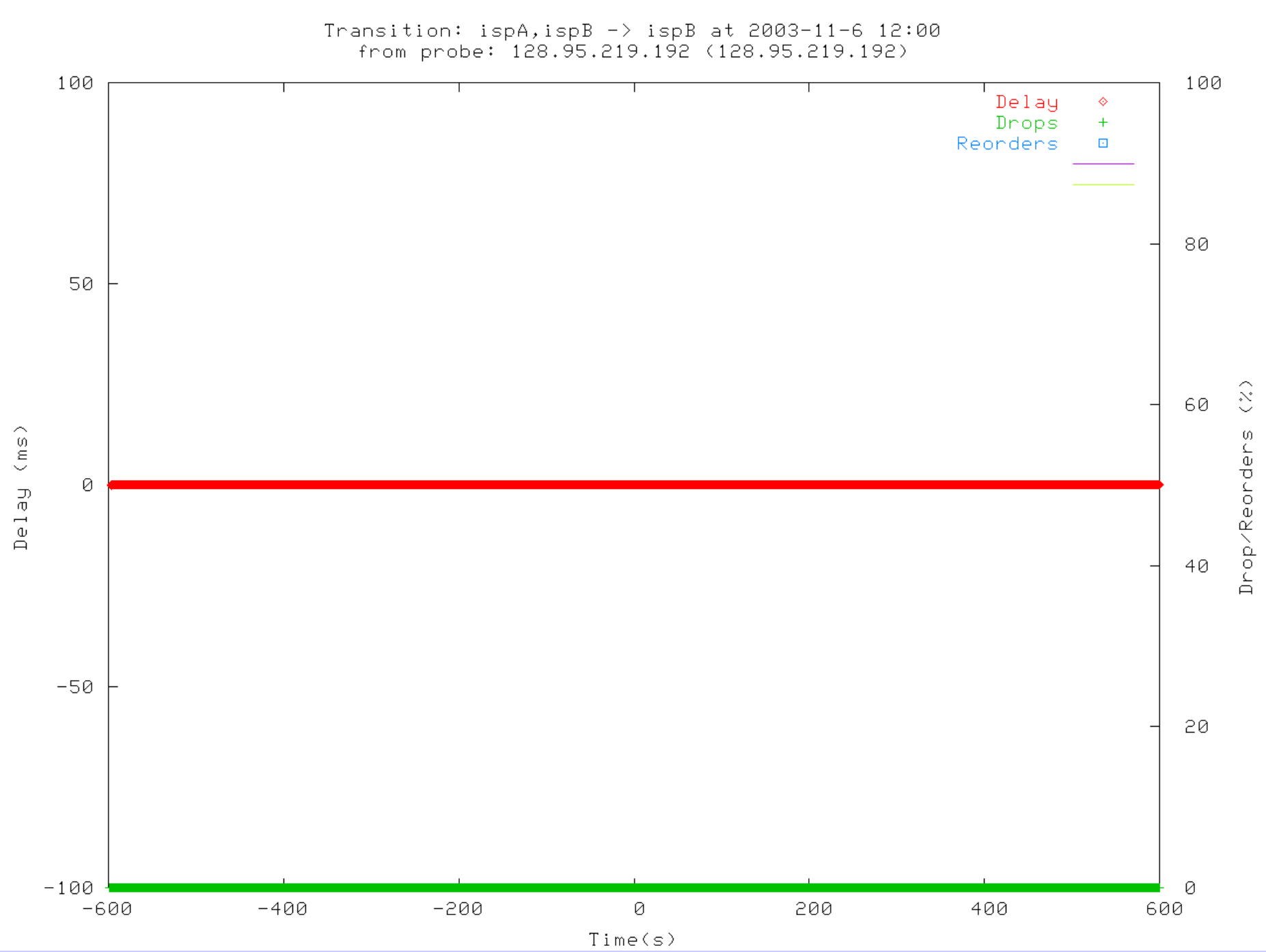


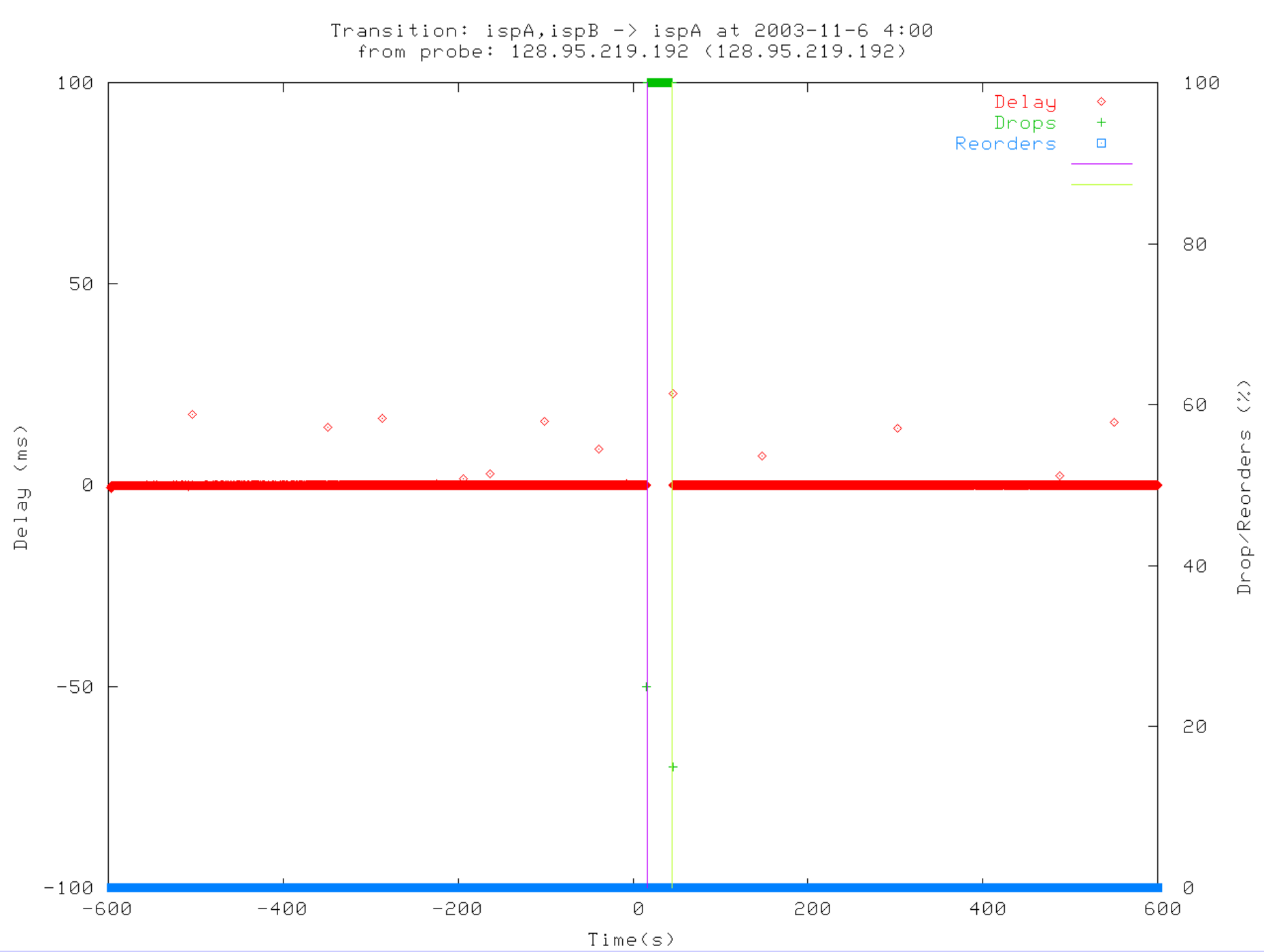
[1]: ISP A
[2]: ISP B
[1,2]: ISP A, ISP B
Time in GMT

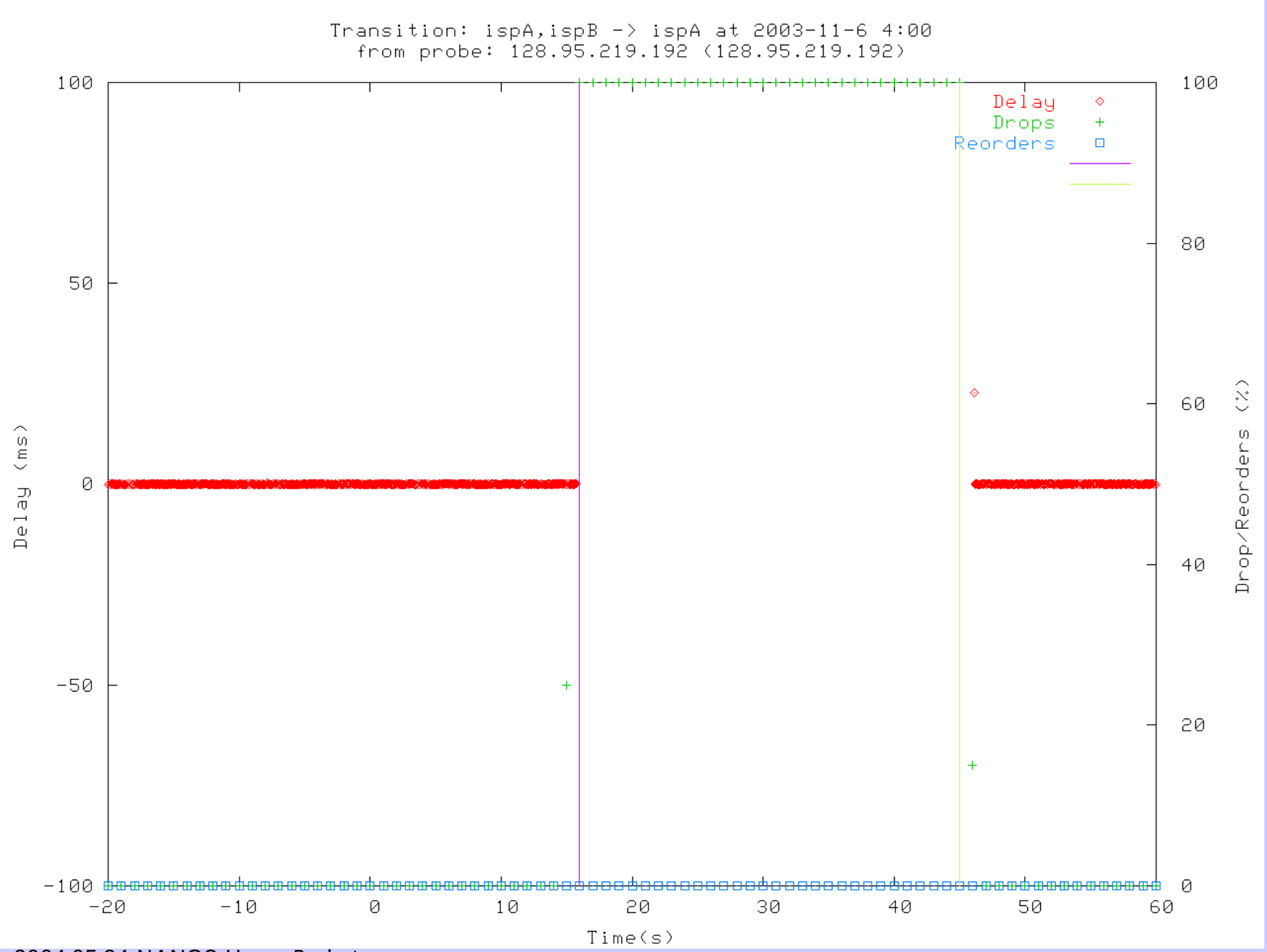
Packet Stream Sources on PlanetLab (and RON)

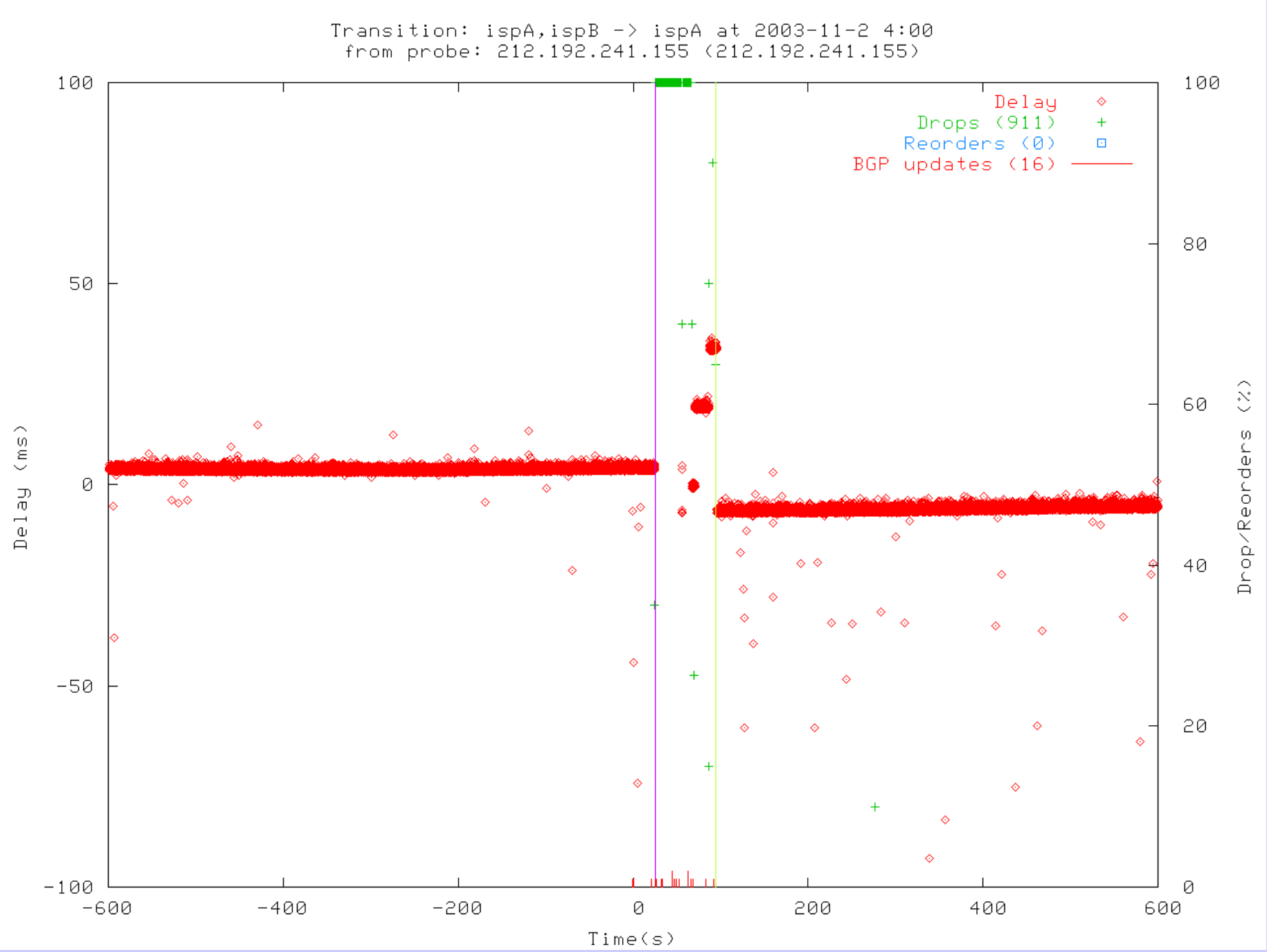


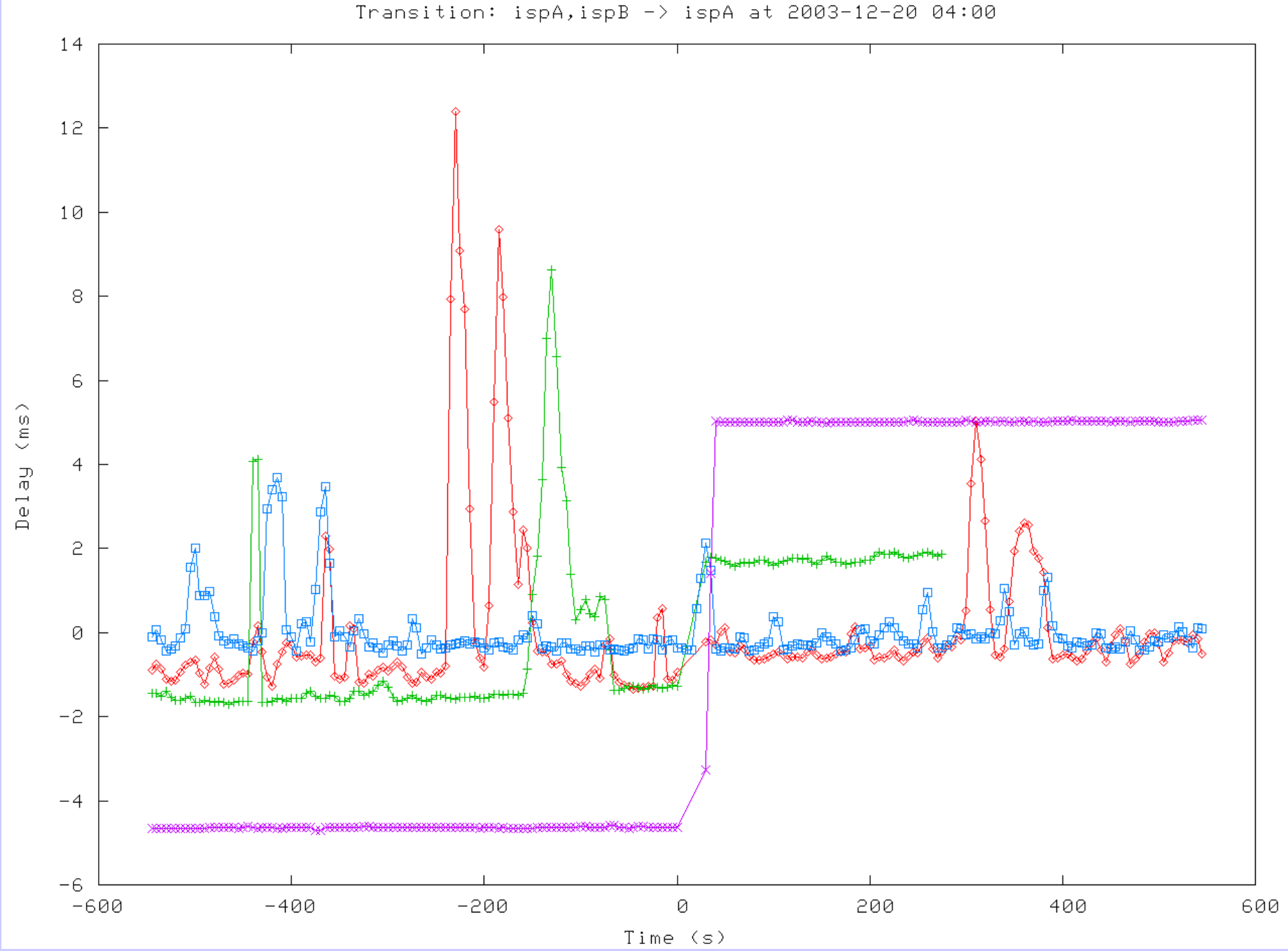
370 nodes at 155 sites
Biased toward R&E Networks
<http://planet-lab.org>

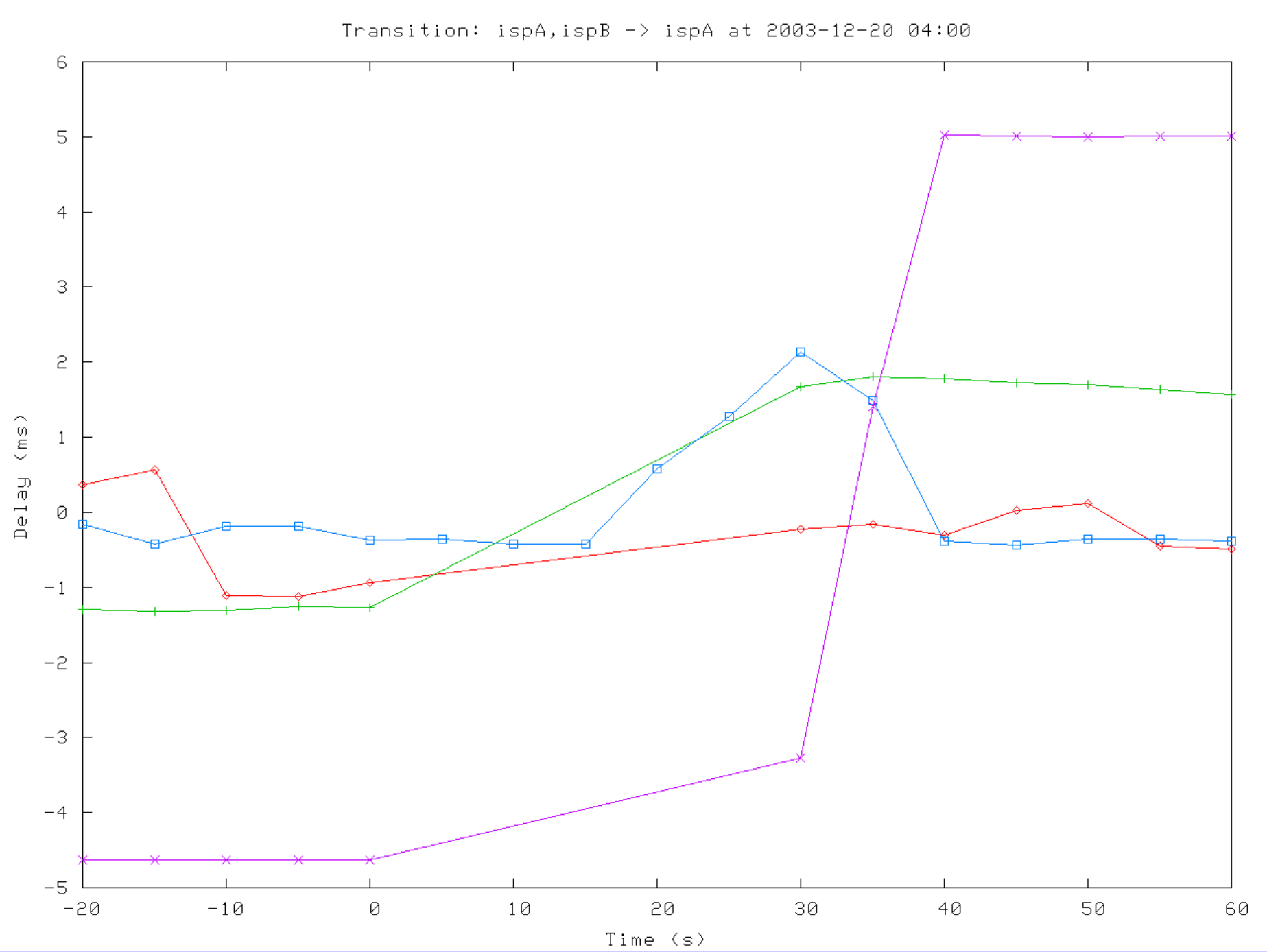

Transition: ispA,ispB -> ispB at 2003-11-6 12:00
from probe: 128.95.219.192 (128.95.219.192)
Delay
Drops
Reorders
100
50
0
-50
-100
Delay (ms)
100
80
60
40
20
0
Drop/Reorders (%)
-600
-400
-200
0
200
400
600
Time(s)

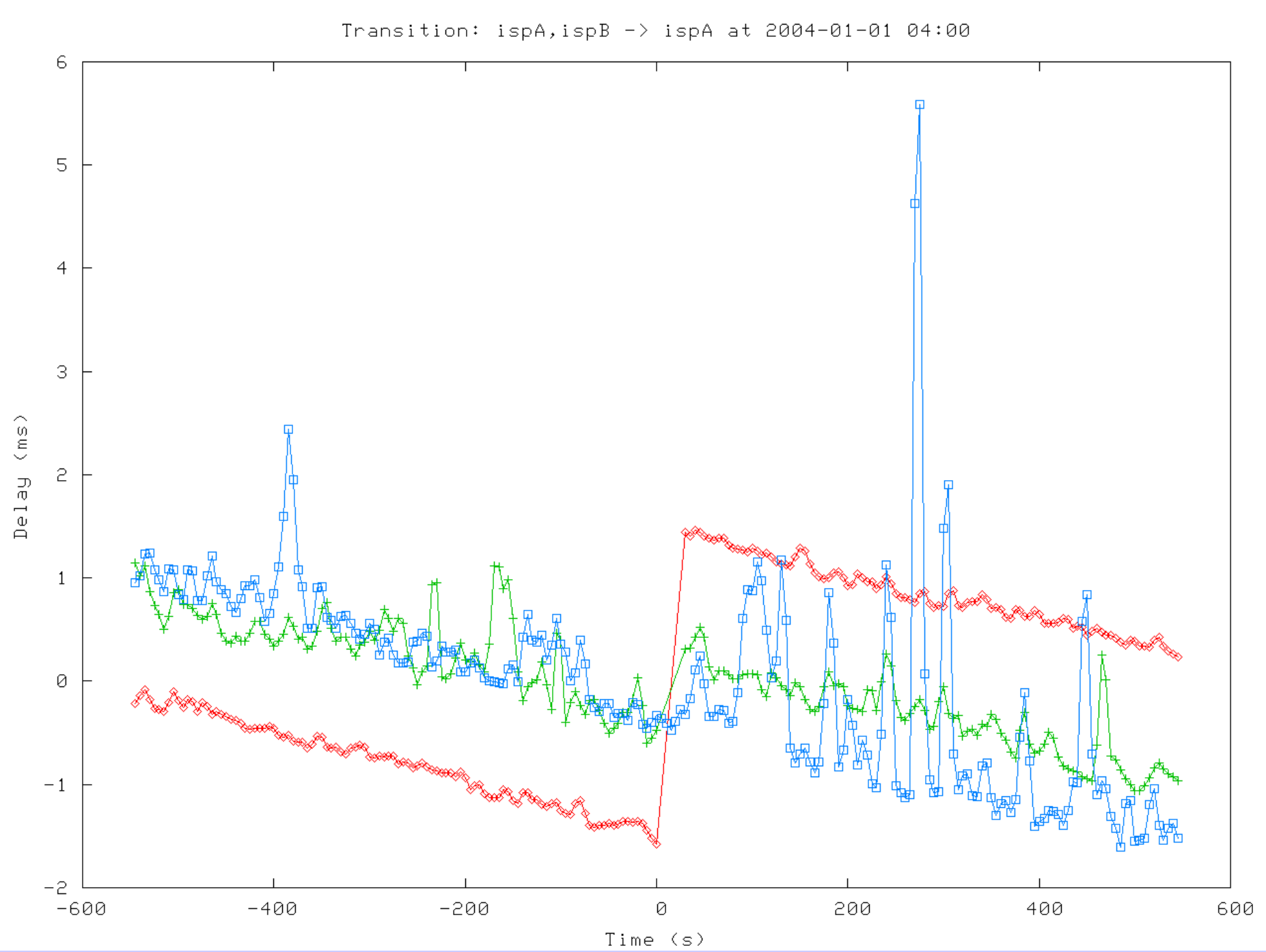

Transition: ispA,ispB -> ispA at 2003-11-6 4:00
from probe: 128.95.219.192 (128.95.219.192)
Delay
Drops
Reorders
100
50
0
-50
-100
Delay (ms)
100
80
60
40
20
0
Drop/Reorders (%)
-600
-400
-200
0
200
400
600
Time(s)

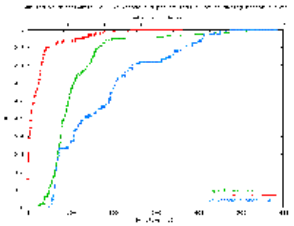
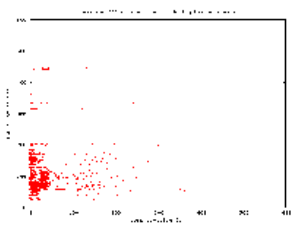
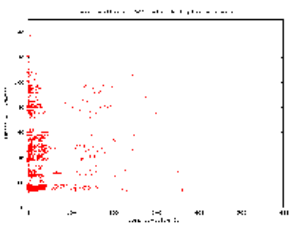
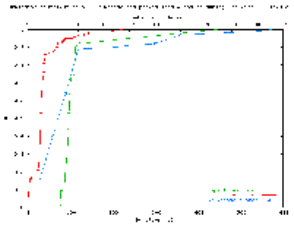
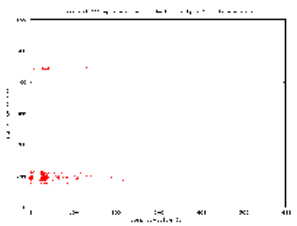
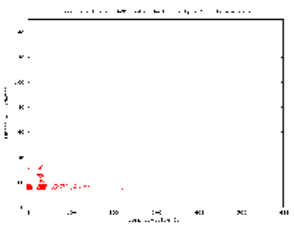
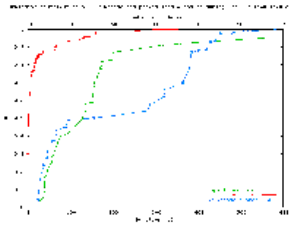
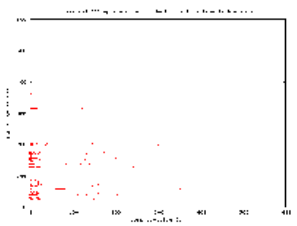
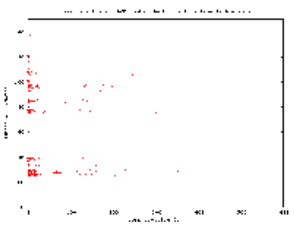
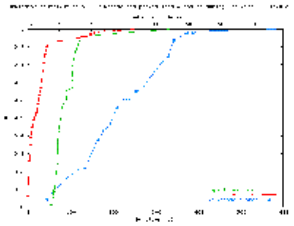
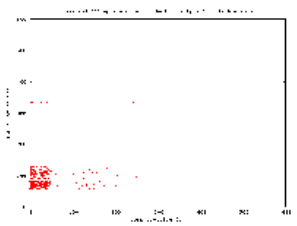
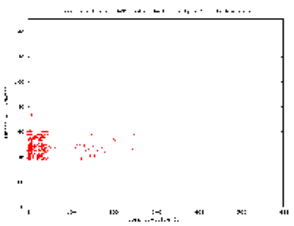
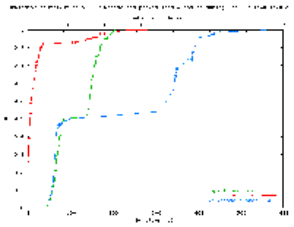
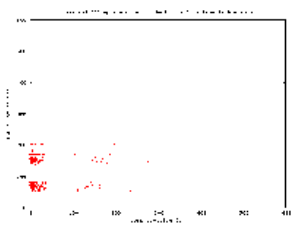
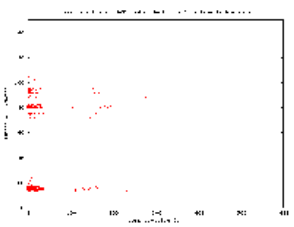

Transition: ispA,ispB -> ispA at 2003-11-6 4:00
from probe: 128.95.219.192 (128.95.219.192)
Delay
Drops
Reorders
Delay (ms)
Drop/Reorders (%)
Time(s)
100
50
0
-50
-100
80
60
40
20
-20
-10
10
20
30
40
50
60

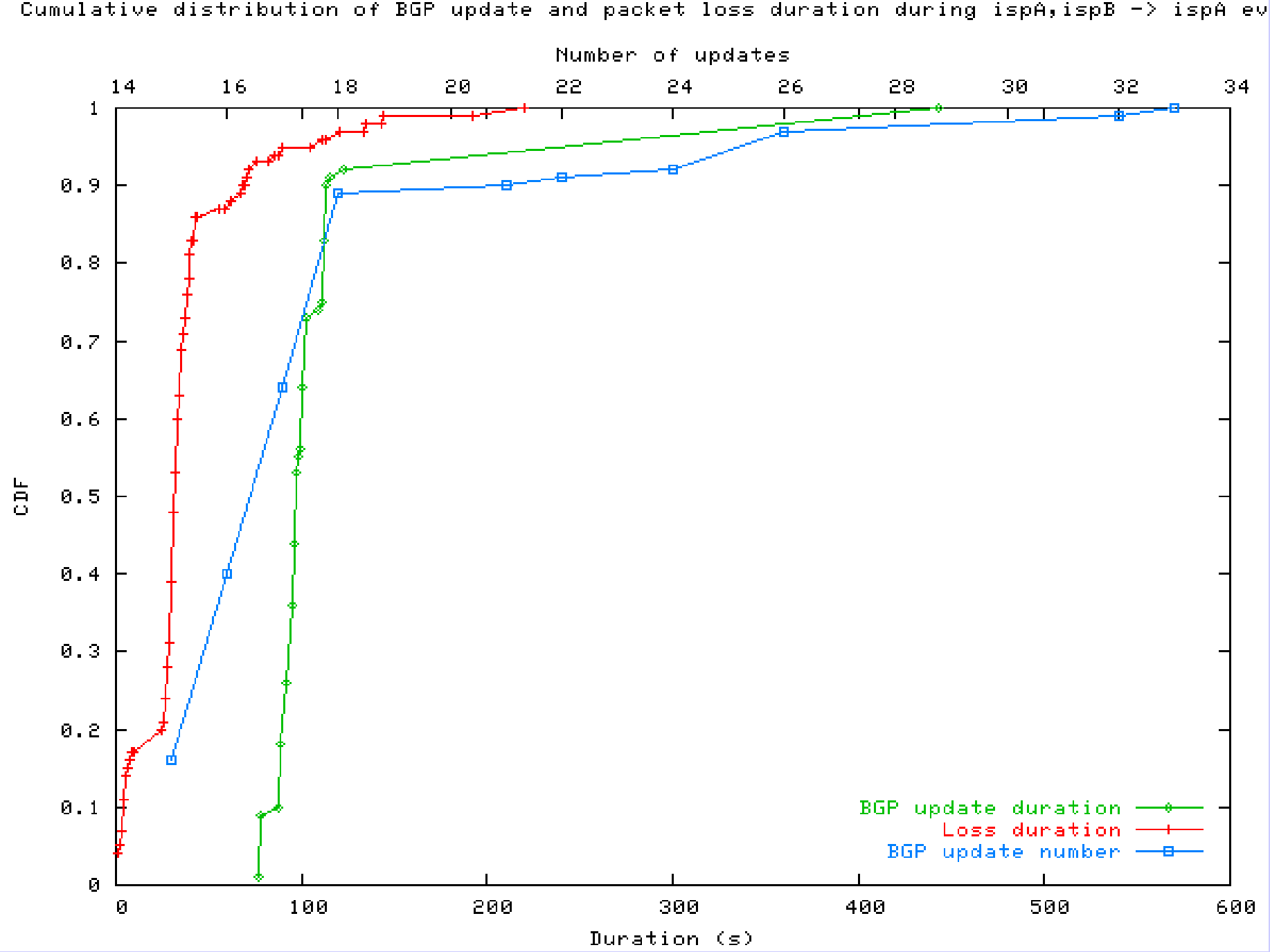

Transition: ispA,ispB -> ispA at 2003-11-2 4:00
from probe: 212.192.241.155 (212.192.241.155)
Delay
Drops (911)
Reorders (0)
BGP updates (16)
Delay (ms)
Drop/Reorders (%)
Time(s)
100
50
0
-50
-100
80
60
40
20
-600
-400
-200
200
400
600

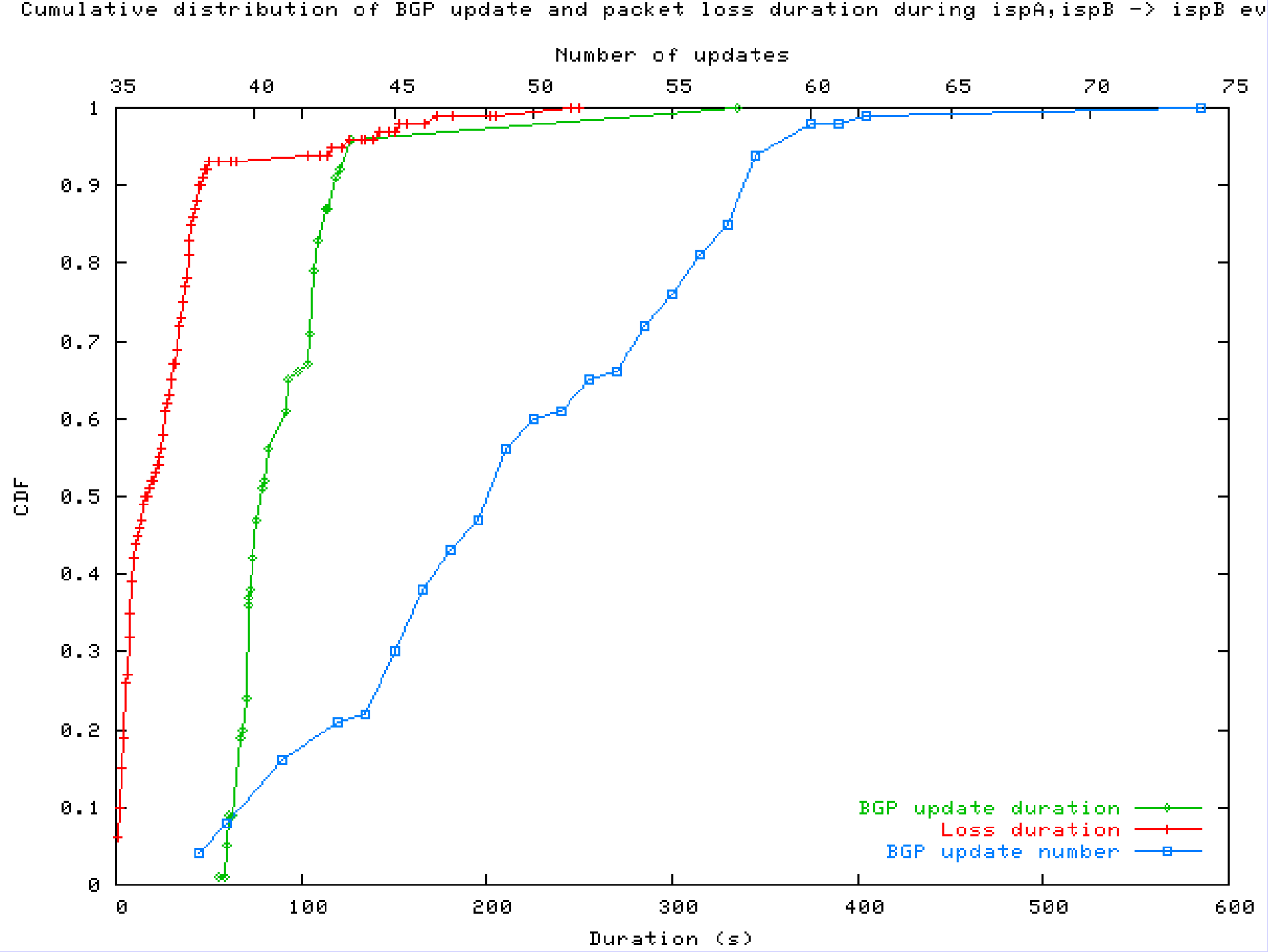

Transition: ispA,ispB -> ispA at 2003-12-20 04:00
Delay (ms)
Time (s)
14
12
10
8
6
4
2
0
-2
-4
-6
-600
-400
-200
0
200
400
600

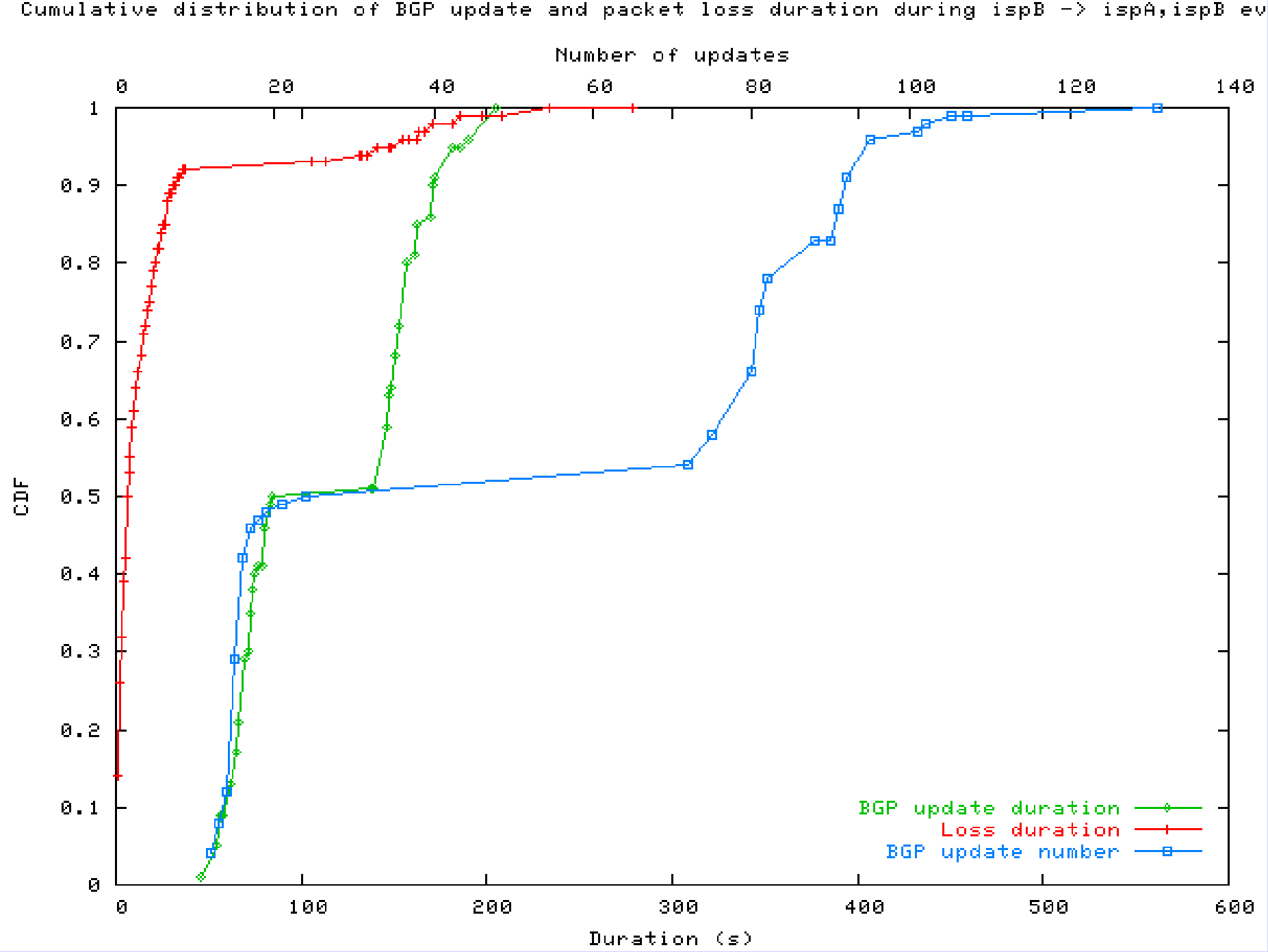

Transition: ispA,ispB -> ispA at 2003-12-20 04:00
Delay (ms)
Time (s)
6
5
4
3
2
1
0
-1
-2
-3
-4
-5
-20
-10
0
10
20
30
40
50
60

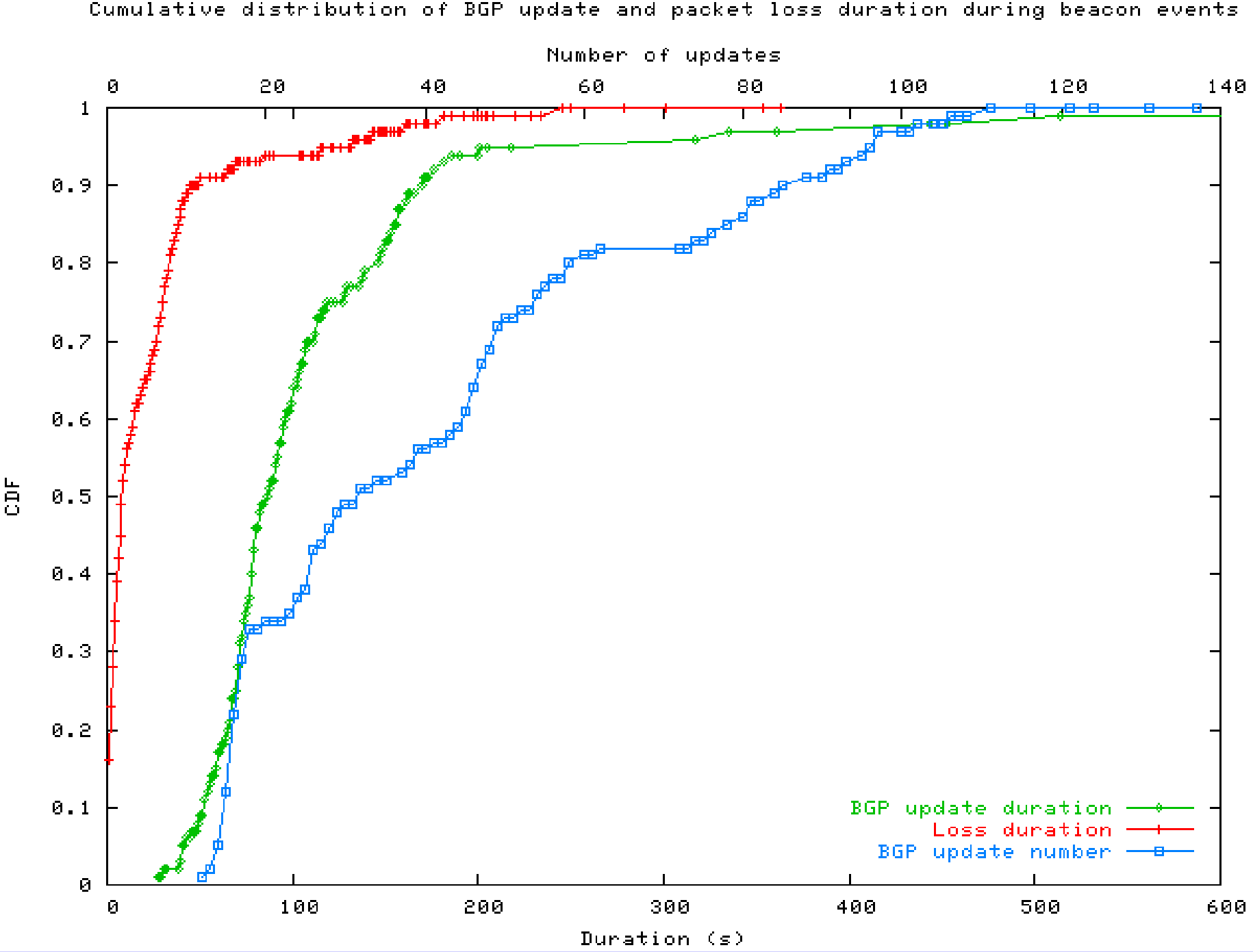

Transition: ispA,ispB -> ispA at 2004-01-01 04:00
Delay (ms)
Time (s)
6
5
4
3
2
1
0
-1
-2
-600
-400
-200
0
200
400
600

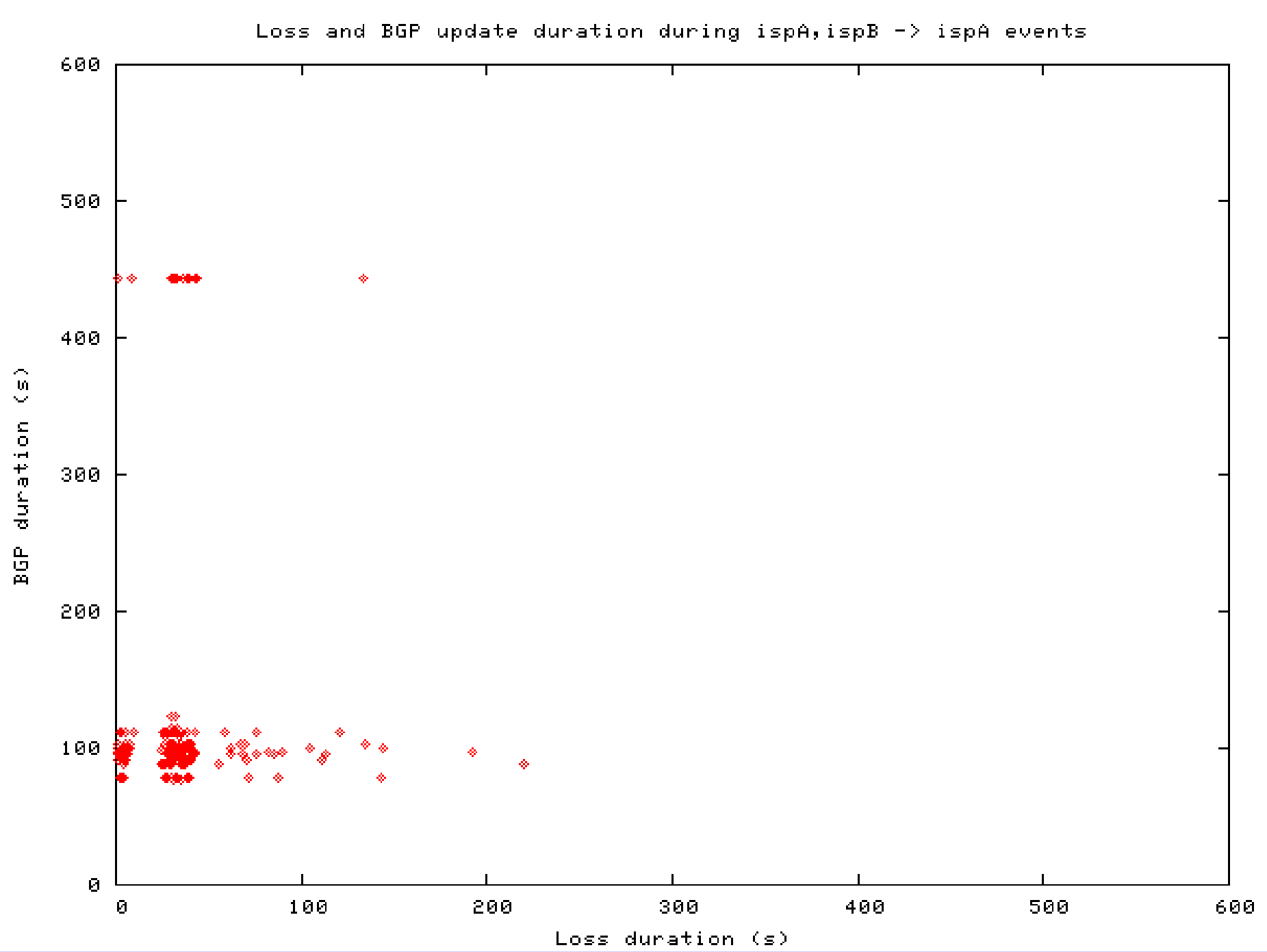
Event	CDF	Loss duration vs BGP duration	Loss duration vs BGP count
All	 [ps][eps]	 [ps][eps]	 [ps][eps]
ispA,ispB -> ispA	 [ps][eps]	 [ps][eps]	 [ps][eps]
ispA -> ispA,ispB	 [ps][eps]	 [ps][eps]	 [ps][eps]
ispA,ispB -> ispB	 [ps][eps]	 [ps][eps]	 [ps][eps]
ispB -> ispA,ispB	 [ps][eps]	 [ps][eps]	 [ps][eps]

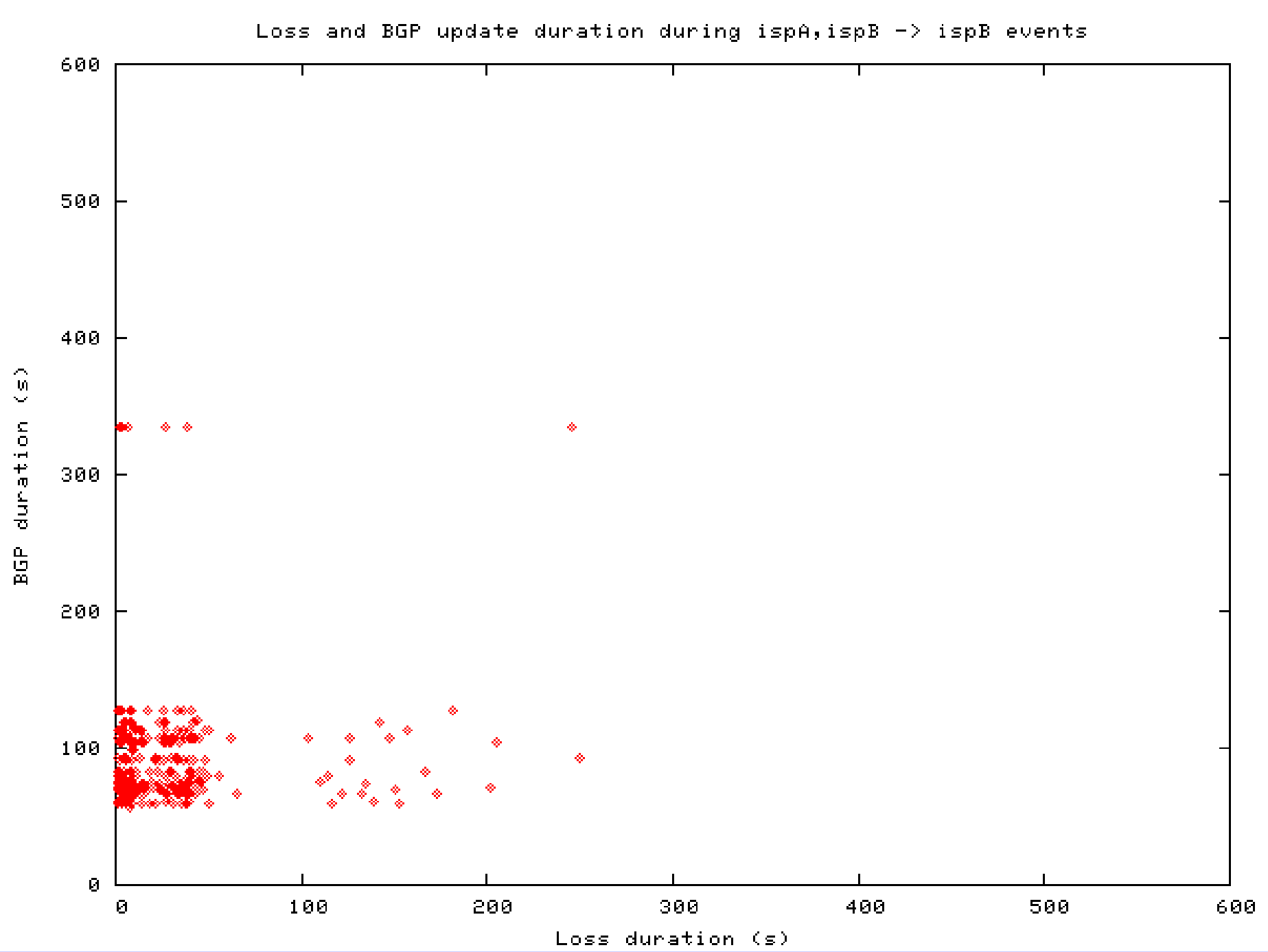

Cumulative distribution of BGP update and packet loss duration during ispA,ispB -> ispA ev
Number of updates
14
16
18
20
22
24
26
28
30
32
34
CDF
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
100
200
300
400
500
600
Duration (s)
BGP update duration
Loss duration
BGP update number

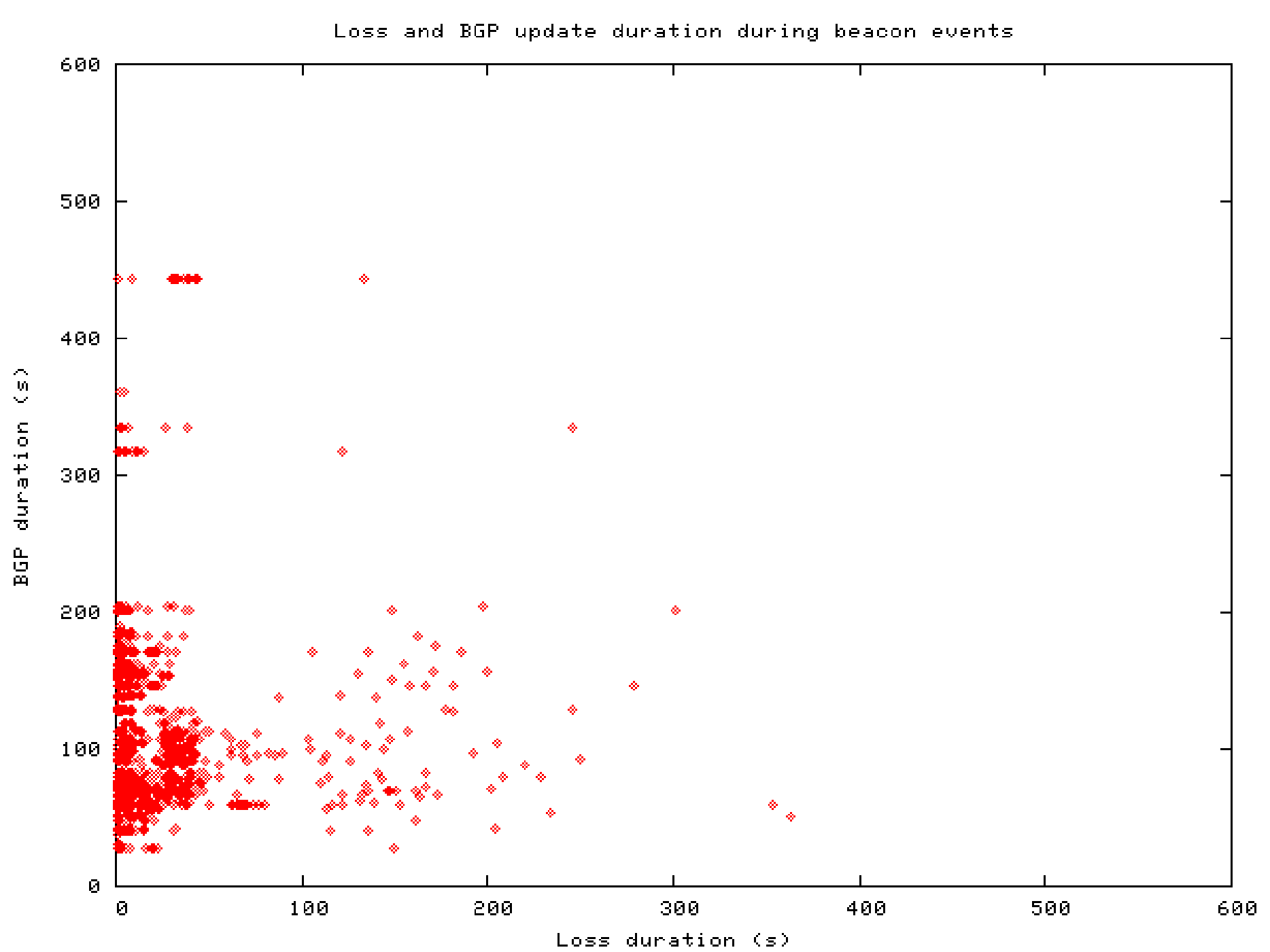

Cumulative distribution of BGP update and packet loss duration during ispA,ispB -> ispB ev
Number of updates
35
40
45
50
55
60
65
70
75
CDF
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
100
200
300
400
500
600
Duration (s)
BGP update duration
Loss duration
BGP update number

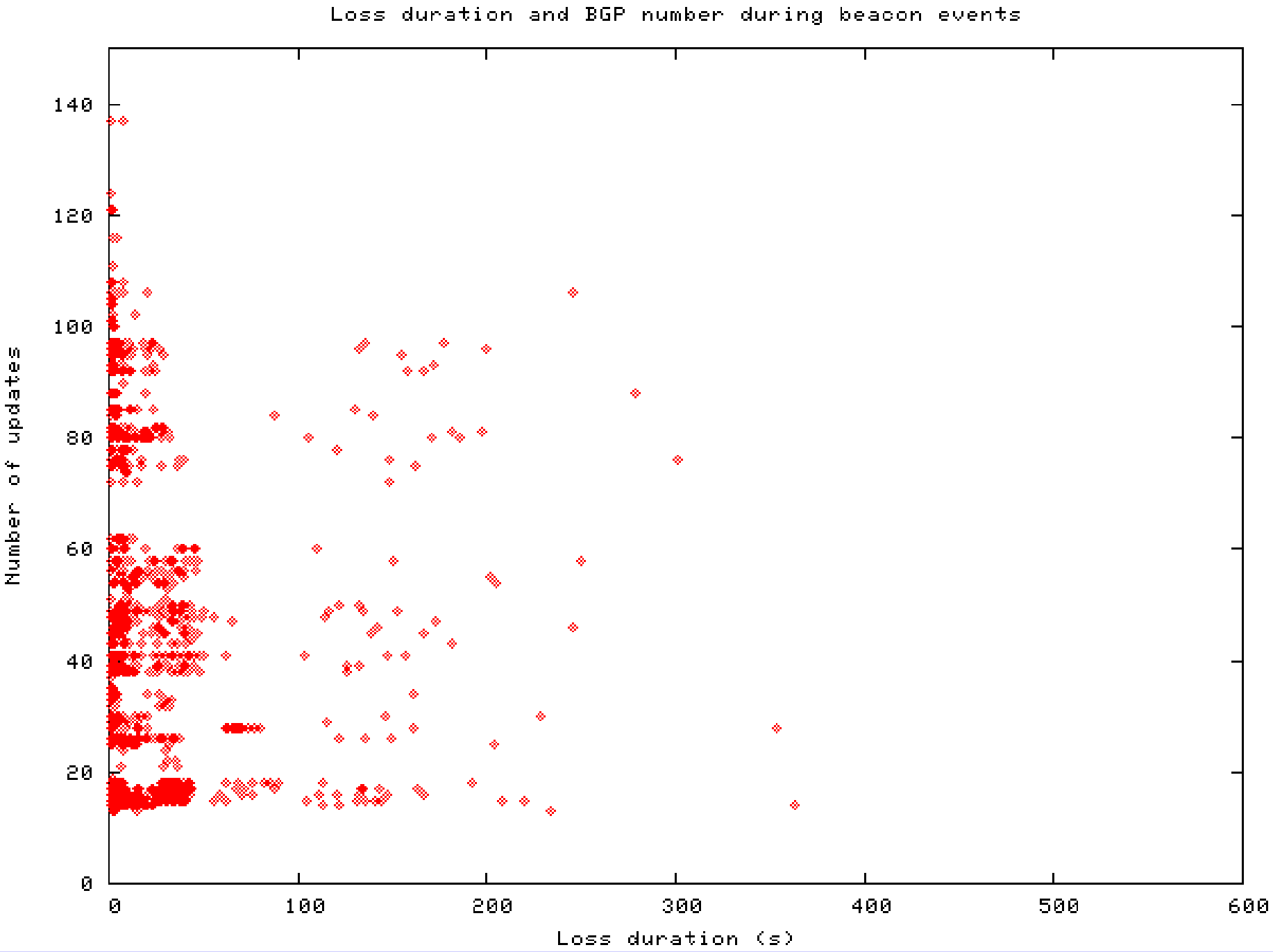

Cumulative distribution of BGP update and packet loss duration during ispB -> ispA,ispB ev
Number of updates
0
20
40
60
80
100
120
140
CDF
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
100
200
300
400
500
600
Duration (s)
BGP update duration
Loss duration
BGP update number

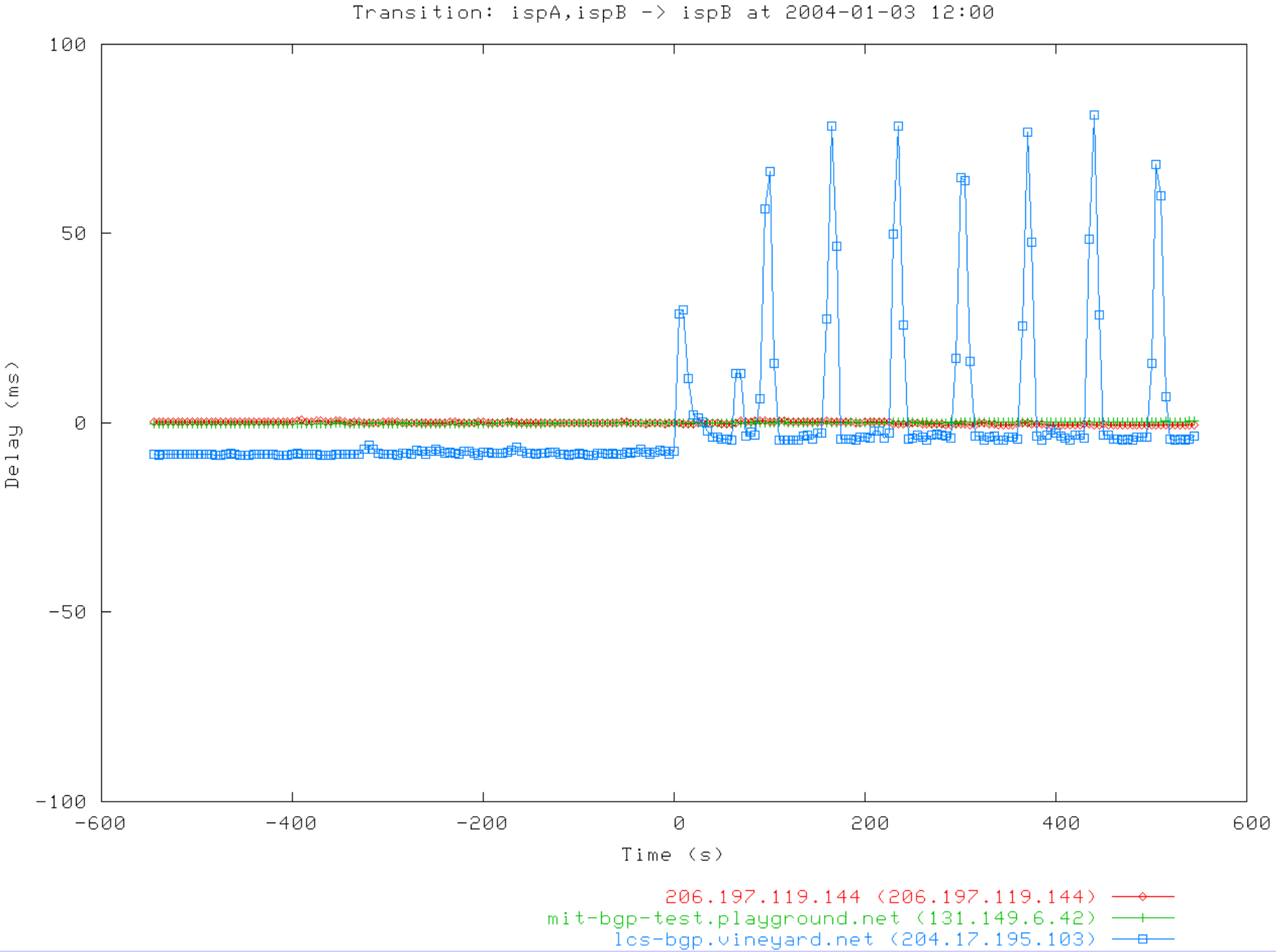

Cumulative distribution of BGP update and packet loss duration during beacon events
Number of updates
0
20
40
60
80
100
120
140
CDF
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
100
200
300
400
500
600
Duration (s)
BGP update duration
Loss duration
BGP update number

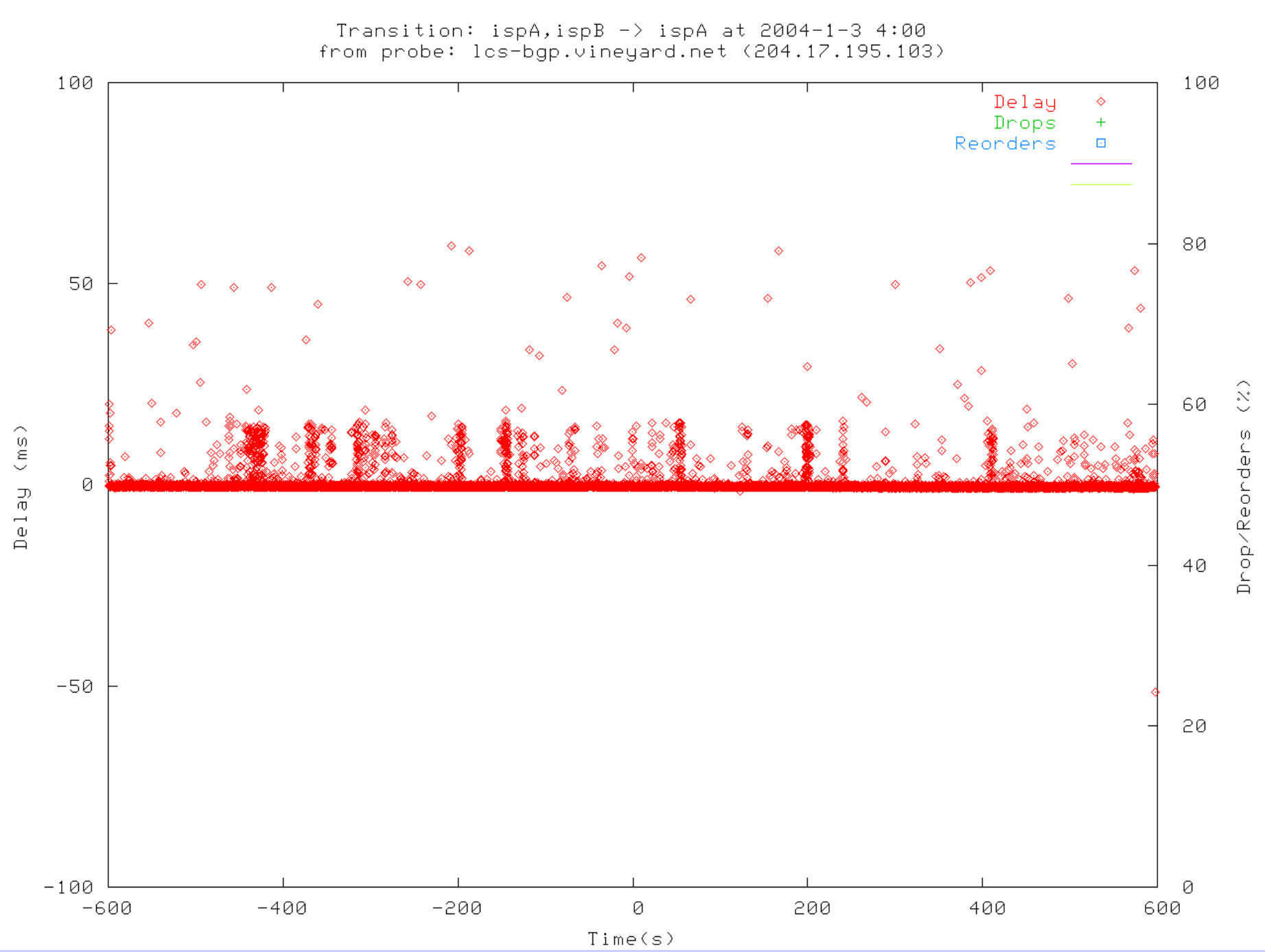

Loss and BGP update duration during ispA,ispB -> ispA events
BGP duration (s)
Loss duration (s)
0
100
200
300
400
500
600

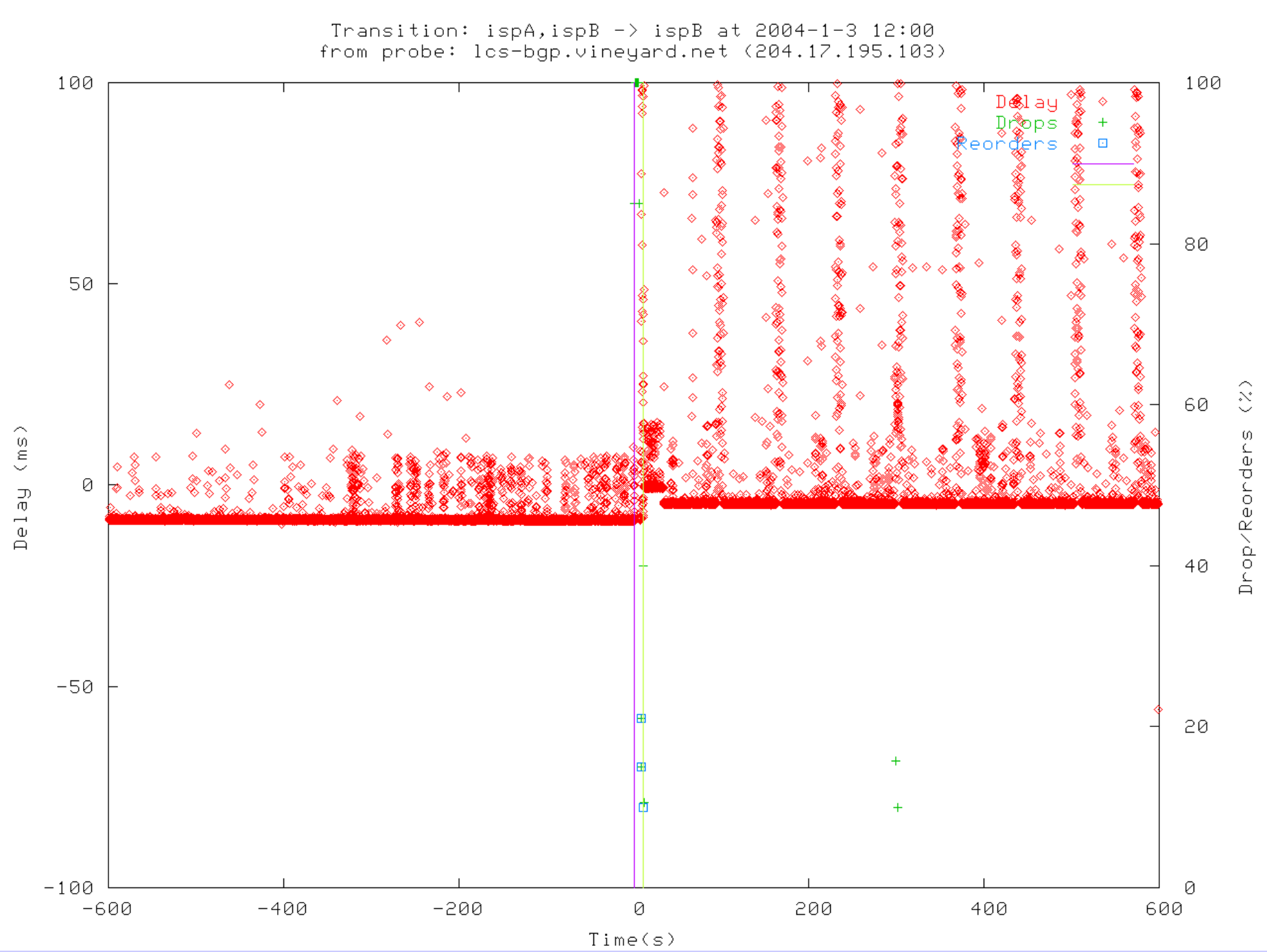

Loss and BGP update duration during ispA,ispB -> ispB events
BGP duration (s)
Loss duration (s)
0
100
200
300
400
500
600

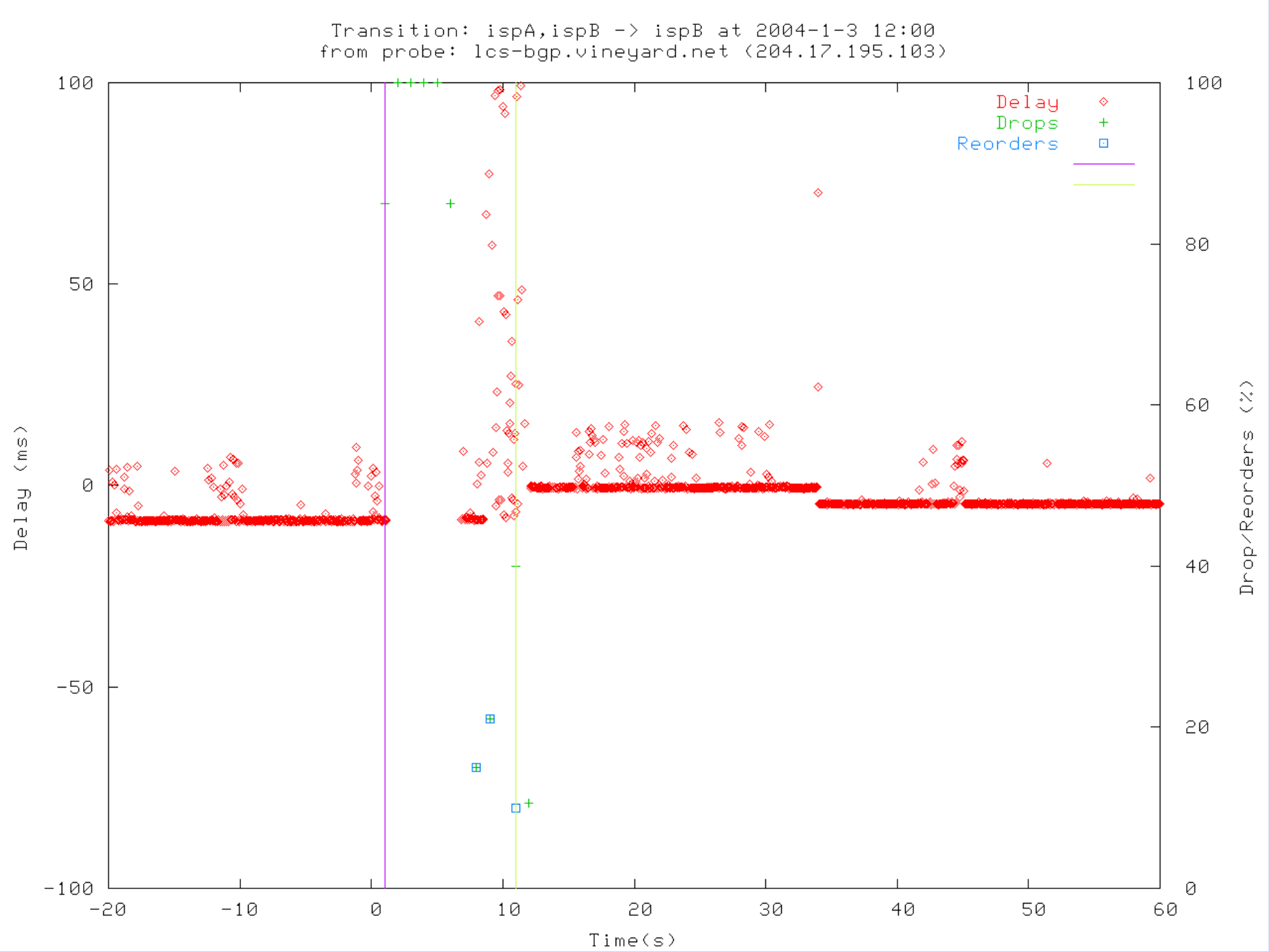

Loss and BGP update duration during beacon events
BGP duration (s)
Loss duration (s)
0
100
200
300
400
500
600

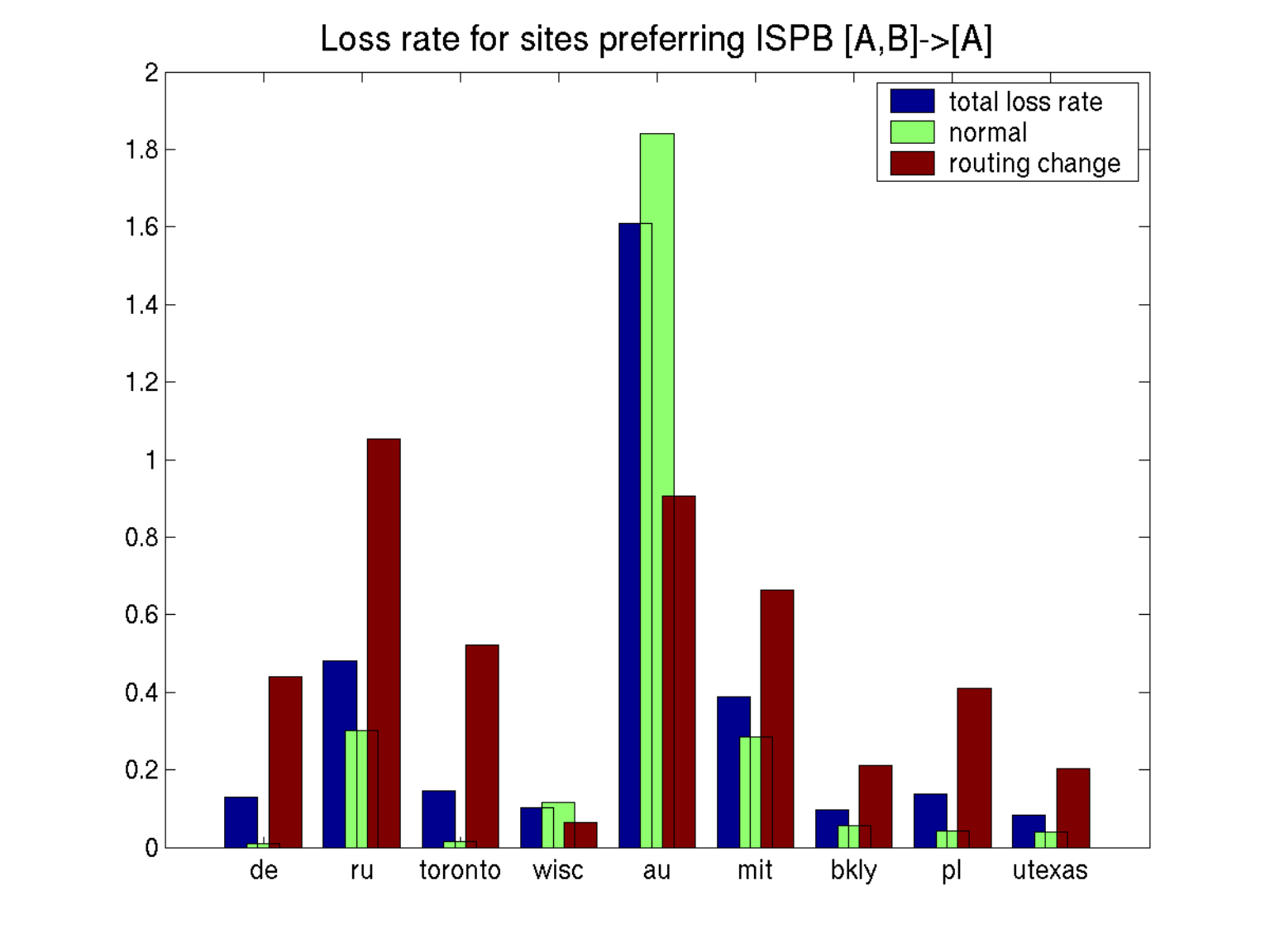

Loss duration and BGP number during beacon events
Number of updates
0
20
40
60
80
100
120
140
0
100
200
300
400
500
600
Loss duration (s)

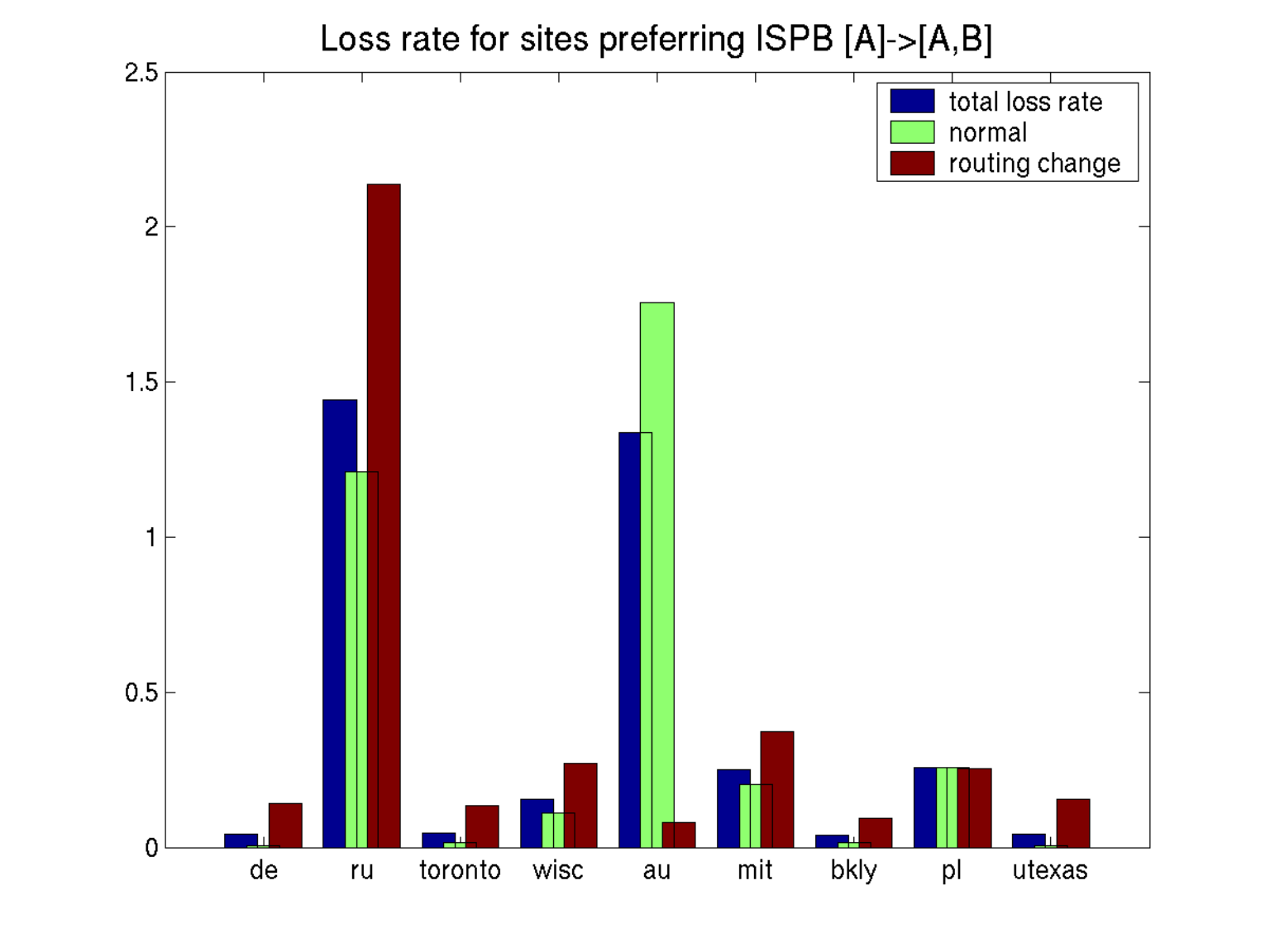

Transition: ispA,ispB -> ispB at 2004-01-03 12:00
100
50
0
-50
-100
Delay (ms)
-600
-400
-200
0
200
400
600
Time (s)
206.197.119.144 (206.197.119.144)
mit-bgp-test.playground.net (131.149.6.42)
lcs-bgp.vineyard.net (204.17.195.103)

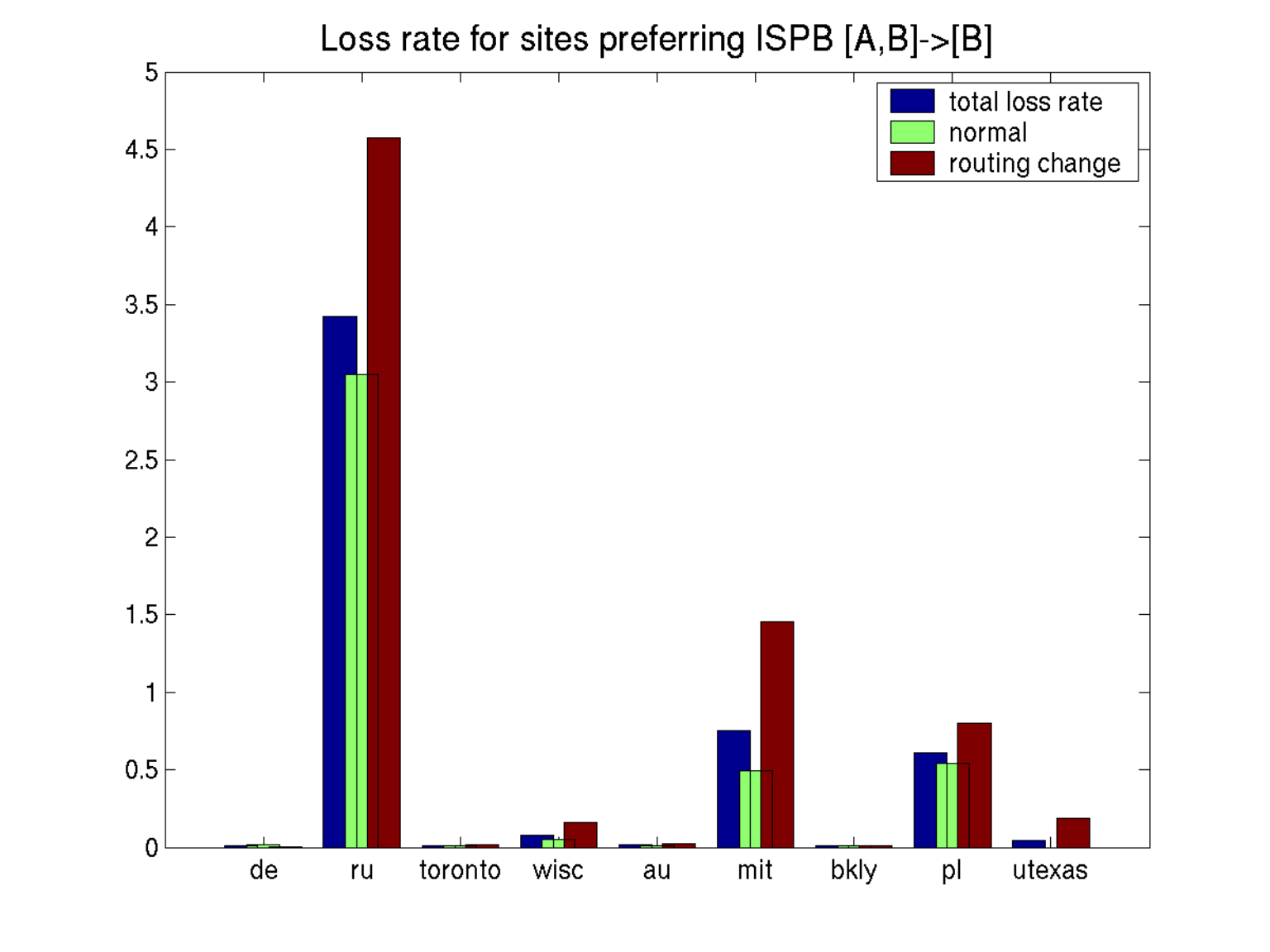

Transition: ispA,ispB -> ispA at 2004-1-3 4:00
from probe: lcs-bgp.vineyard.net (204.17.195.103)
Delay
Drops
Reorders
100
50
0
-50
-100
Delay (ms)
100
80
60
40
20
0
Drop/Reorders (%)
-600
-400
-200
0
200
400
600
Time(s)

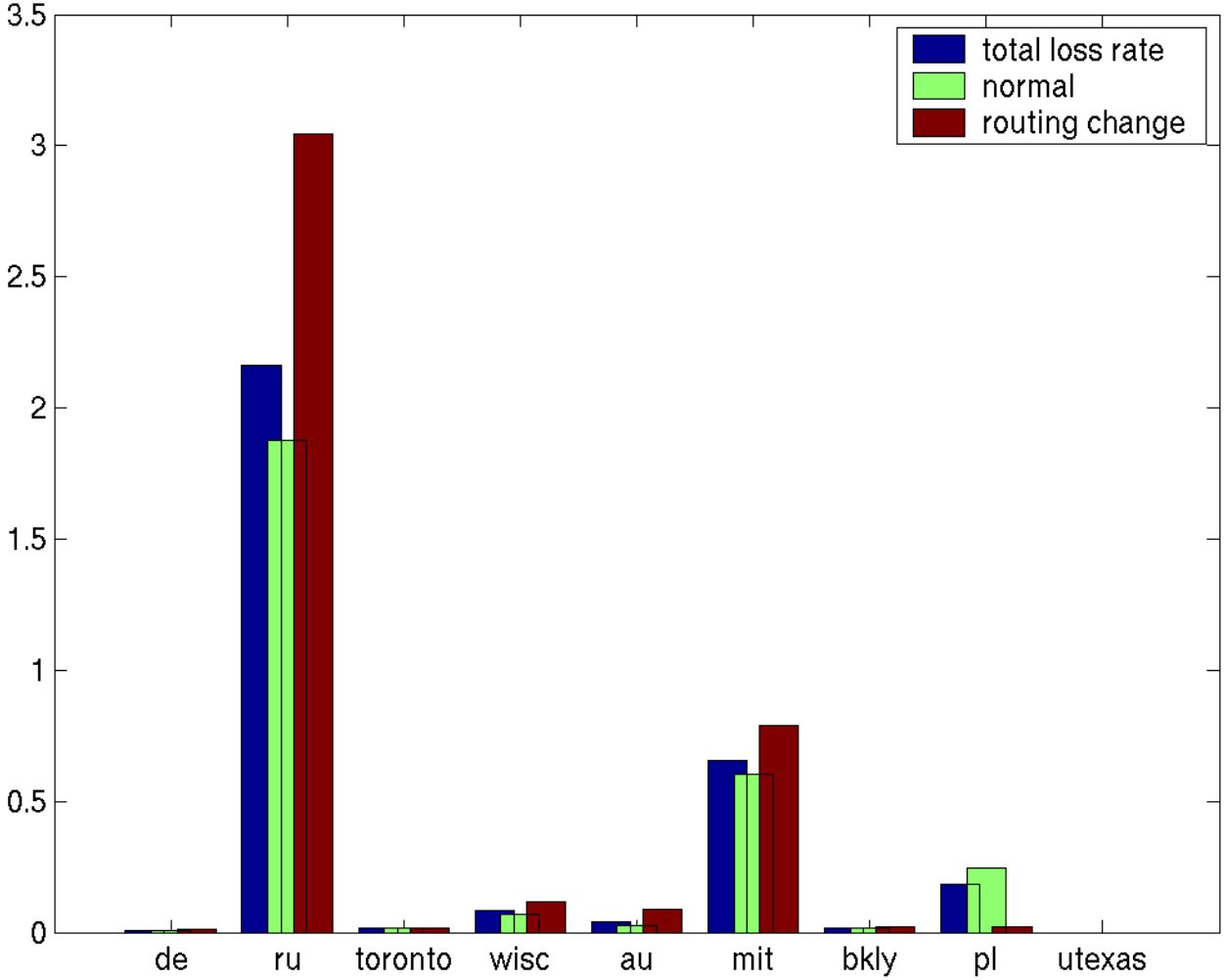

Transition: ispA,ispB -> ispB at 2004-1-3 12:00
from probe: lcs-bgp.vineyard.net (204.17.195.103)
Delay
Drops
Reorders
Delay (ms)
Drop/Reorders (%)
Time(s)
100
50
0
-50
-100
80
60
40
20
-600
-400
-200
200
400
600

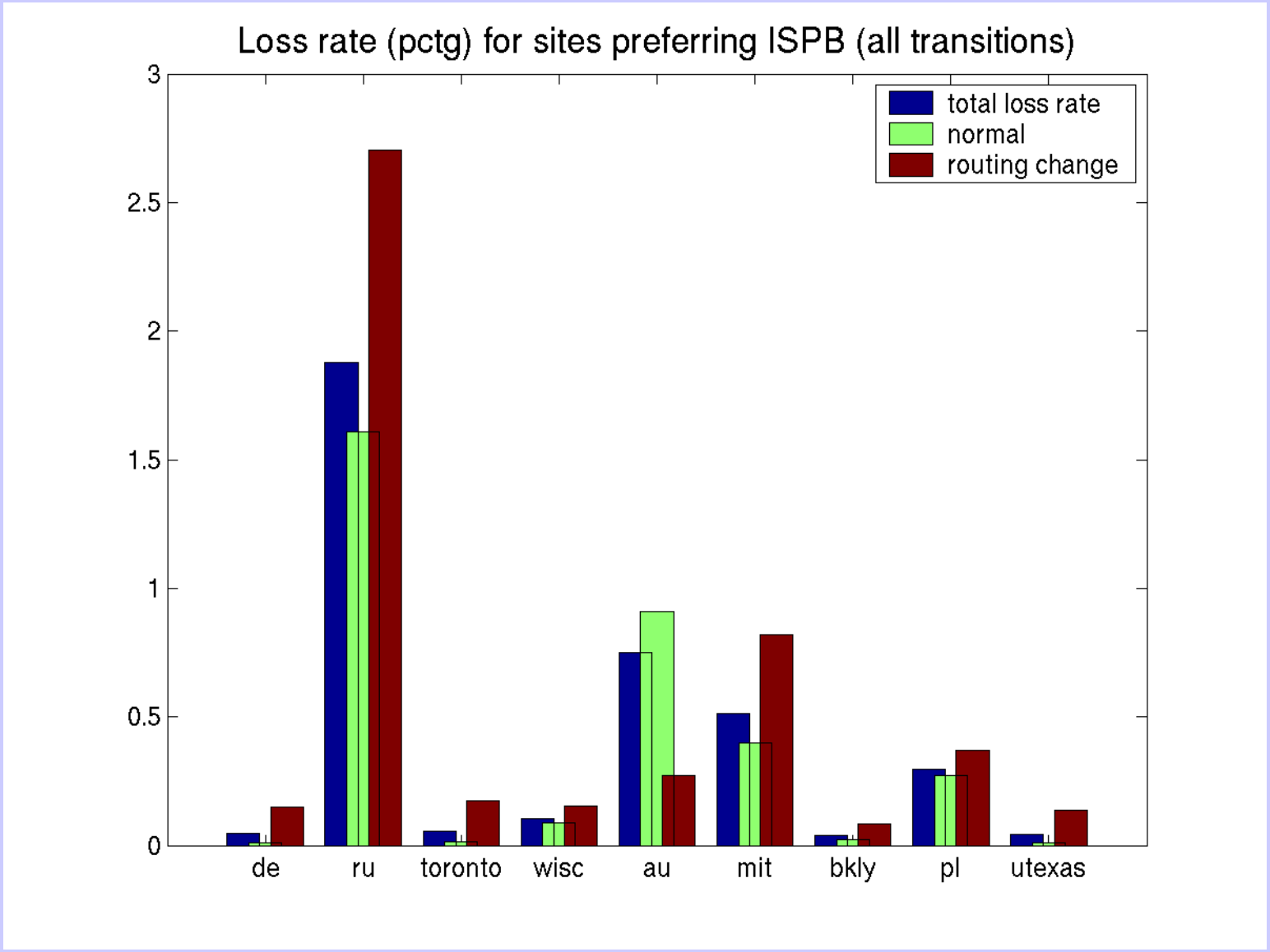

Transition: ispA,ispB -> ispB at 2004-1-3 12:00
from probe: lcs-bgp.vineyard.net (204.17.195.103)
Delay
Drops
Reorders
Delay (ms)
Drop/Reorders (%)
Time(s)
100
50
0
-50
-100
80
60
40
20
-20
-10
10
20
30
40
50
60

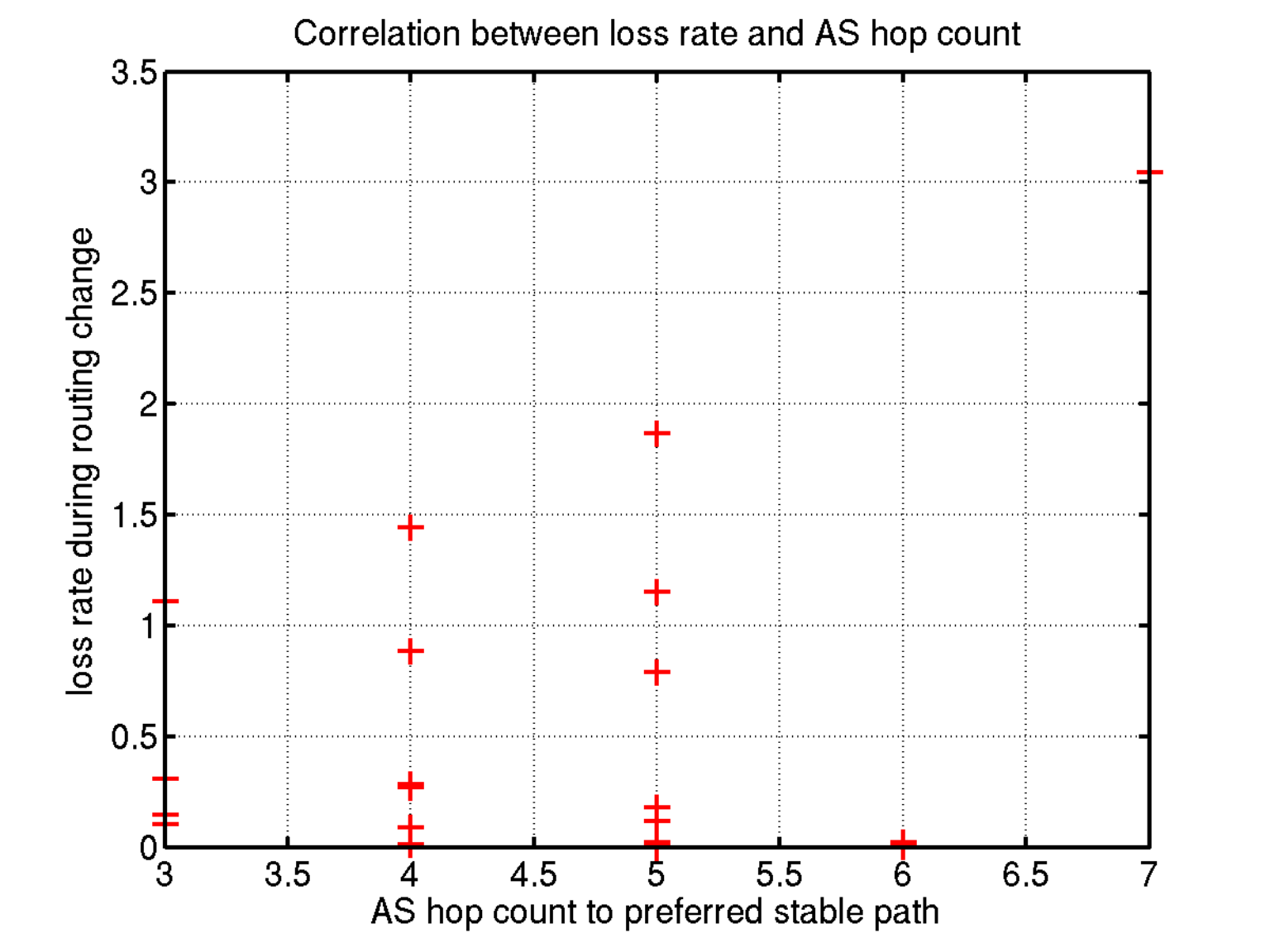

Loss rate for sites preferring ISPB [A,B]->[A]
total loss rate
normal
routing change
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
de
ru
toronto
wisc
au
mit
bkly
pl
utexas

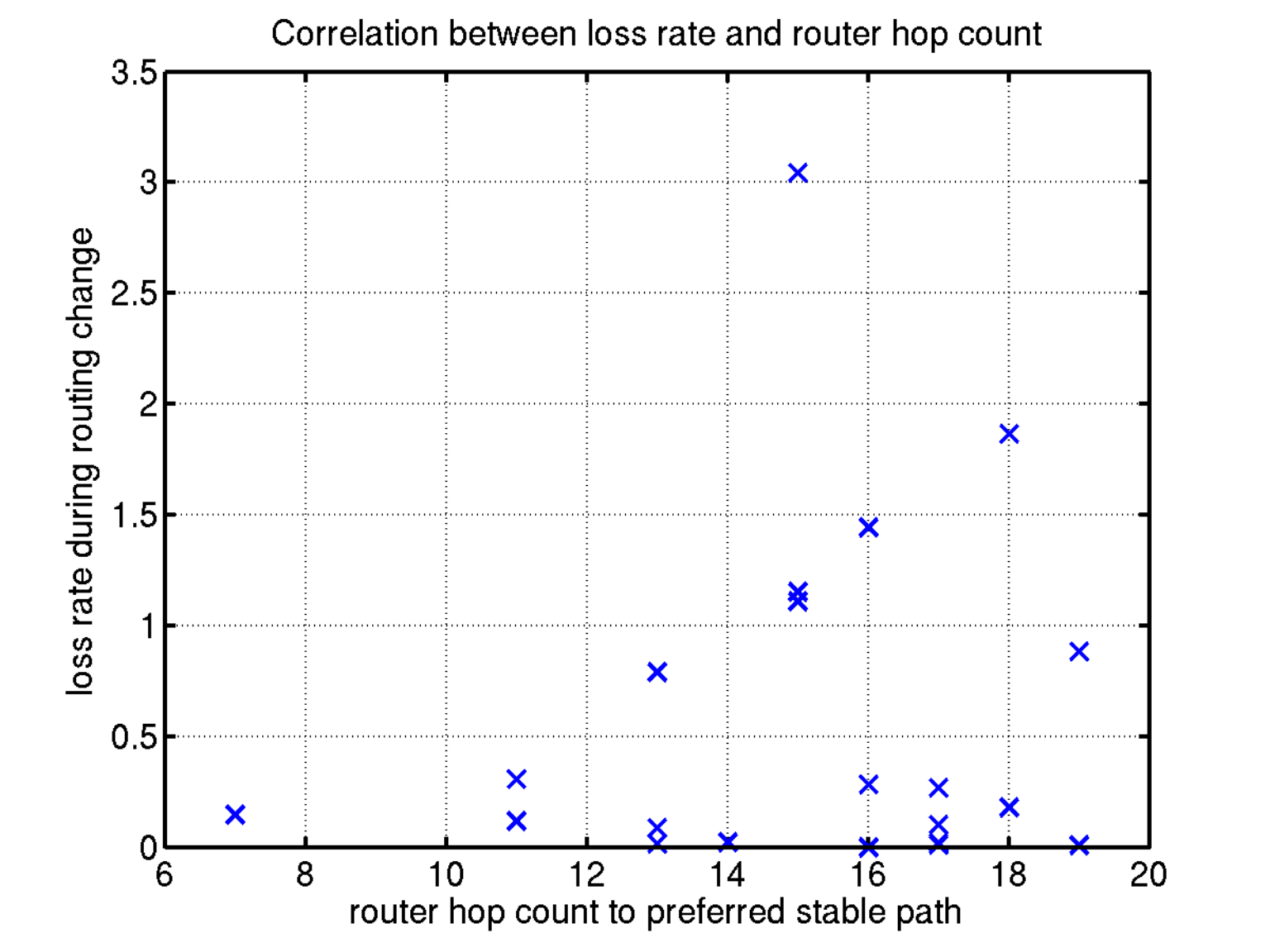

Loss rate for sites preferring ISPB [A]->[A,B]
total loss rate
normal
routing change
2.5
2
1.5
1
0.5
0
de
ru
toronto
wisc
au
mit
bkly
pl
utexas


Loss rate for sites preferring ISPB [A,B]->[B]
total loss rate
normal
routing change
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
de
ru
toronto
wisc
au
mit
bkly
pl
utexas


Loss rate for sites preferring ISPB [B]->[A,B]
total loss rate
normal
routing change
3.5
3
2.5
2
1.5
1
0.5
0
de
ru
toronto
wisc
au
mit
bkly
pl
utexas


Loss rate (pctg) for sites preferring ISPB (all transitions)
total loss rate
normal
routing change
3
2.5
2
1.5
1
0.5
0
de
ru
toronto
wisc
au
mit
bkly
pl
utexas


Correlation between loss rate and AS hop count
loss rate during routing change
AS hop count to preferred stable path
0
0.5
1
1.5
2
2.5
3
3.5
3
3.5
4
4.5
5
5.5
6
6.5
7


Correlation between loss rate and router hop count
loss rate during routing change
router hop count to preferred stable path
0
0.5
1
1.5
2
2.5
3
3.5
6
8
10
12
14
16
18
20

This Seems to Say

- Distant sites experience more loss.
- There is a correlation between a site's routing preference and the type of transition: sites preferring ISP A have more loss rate during AB->B than AB->A, similarly more loss rate during B->AB than A->AB.
- The correlation between loss rate and AS or router hop count is quite weak. (we need more data points here).
- At some sites, the loss rate during 'normal' periods (i.e., no injected routing change) is higher than that during 'routing change' periods. Maybe those paths' inherent loss may be due to congestion.

References

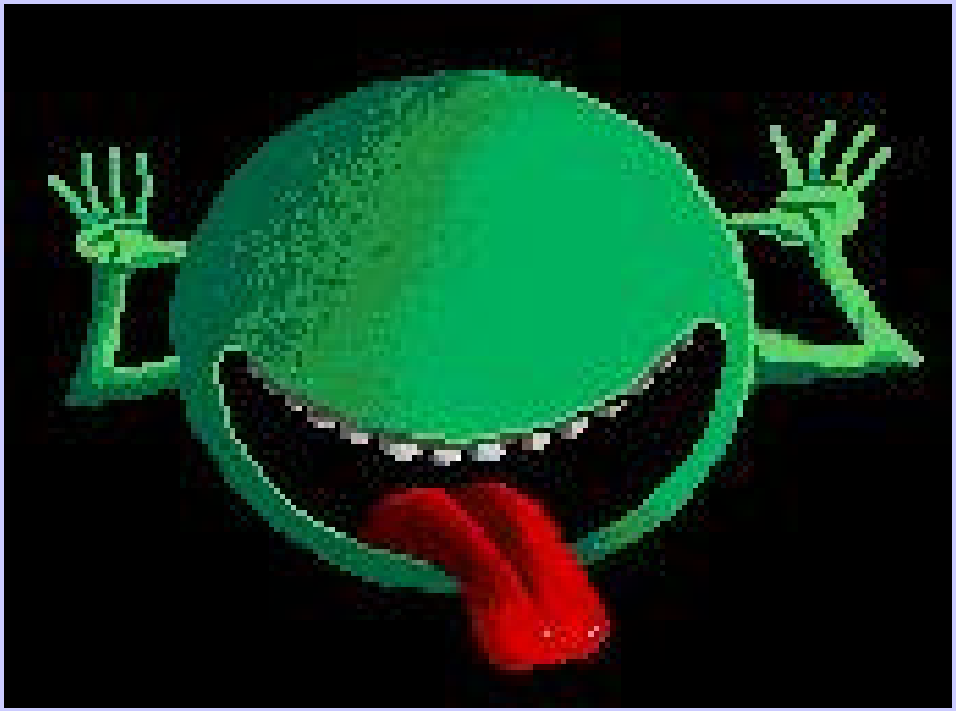
- Tim Griffin, “What is the sound of one route flapping” Dartmouth talk slides, June 2002
- "Global Routing Instabilities During Code Red II and Nimda Worm Propagation“ Jim Cowie and Andy Ogielski, Renesys. NANOG 23, October 2001
- C. Labovitz, G. R. Malan, F. Jahanian, “Internet Routing Instability”, TON 1998
- C. Labovitz, R. Malan, F. Jahanian, “Origins of Internet Routing Instability”, Infocom 1999

References (2)

- C. Labovitz, A. Ahuja, F. Jahanian, “Experimental Study of Internet Stability and Wide-Area Network Failures”, FTCS 1999
- C. Labovitz, A. Ahuja, A. Bose, F. Jahanian, “Delayed Internet Routing Convergence” Sigcomm 2000
- C. Labovtiz, A. Ahuja, R. Wattenhofer, S. Venkatachary, “The Impact of Internet Policy and Topology on Delayed Routing Convergence”, Infocom 2001

Thanks to Our Sponsors

- NSF via award ANI-0221435
- The University of Oregon (Dan & Eric)
- The University of Michigan (Morley)
- Internet Initiative Japan (Randy)
- Intel Corporation (Tim)
- Verio and Sprint (bandwidth)
- Juniper & Cisco (routers)



Don’t Panic,

Engineer Prudently