
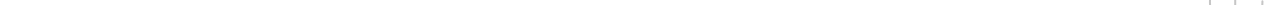




Interdomain routing reliability measurements



CS 8803 presentation
by Srinivasan Seetharaman
Fall 2003

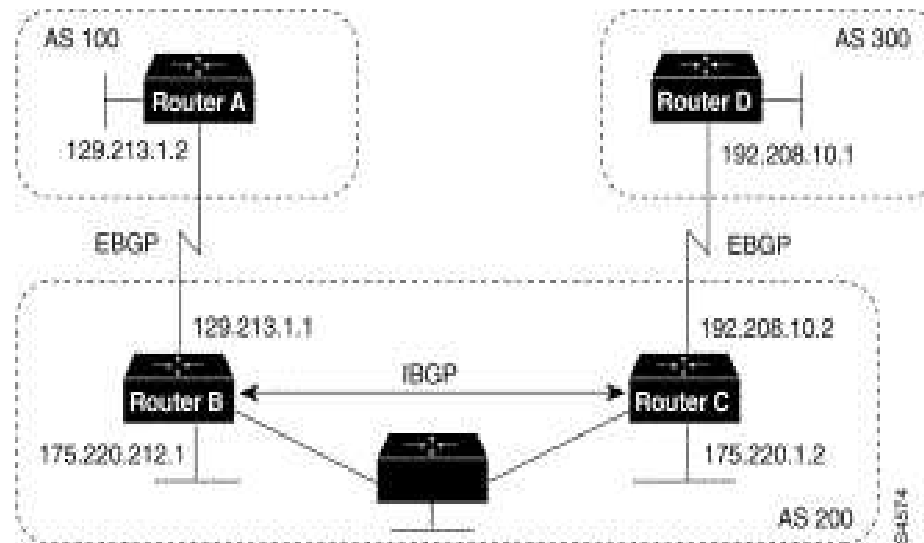


Outline..

- ◆ Introduction to Interdomain Routing
 - BGP
- ◆ Problems arising from backbone failures:
 - Delayed Internet Routing Convergence
 - Internet Routing Instability
 - Routing loops
- ◆ Tools available:
 - RouteViews
 - Zebra Listener
 - BGP Beacons
- ◆ Conclusion

Interdomain Routing

- ◆ Objective - Select the best path towards each destination that is compatible with the routing policies of the transit ASes without knowing the topology of the transit ASes



- ◆ External BGP (EBGP) – run between routers from different ASes.
- ◆ Internal BGP (IBGP) – run between routers within the same ASes.

Border Gateway Protocol

- ◆ Path vector protocol, Runs over TCP (port 179), Incremental, Use Classless Interdomain Routing (CIDR), Only best-path

- ◆ When a BGP speaker receives updates from multiple ASs that describe different paths to the same destination, it must choose the single best path for reaching that destination. Once chosen, BGP propagates the best path to its neighbors.

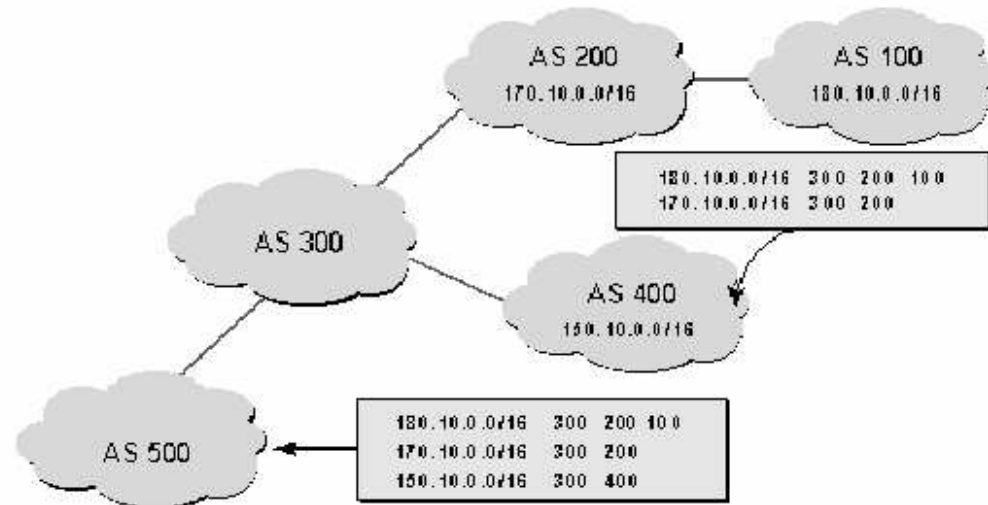
- ◆ The decision is based on the value of attributes:
 - AS Path
 - Next Hop
 - Local Preference
 - Multi-exit discriminator (MED)
 - Origin
 - Others...

Important BGP Attributes

- ◆ MinRouteAdver: Minimum interval between successive updates sent to a peer for a given prefix
 - Allow for greater efficiency/packing of updates
 - Announcement Rate throttle

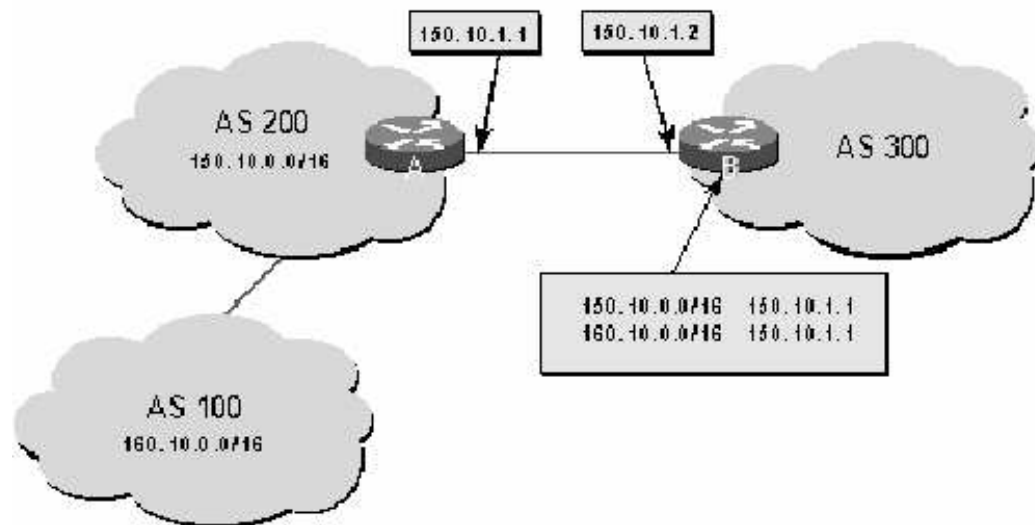
- ◆ AS Path

- Sequence of AS's a route has traversed
- Used for loop detection



Important BGP Attributes (contd)

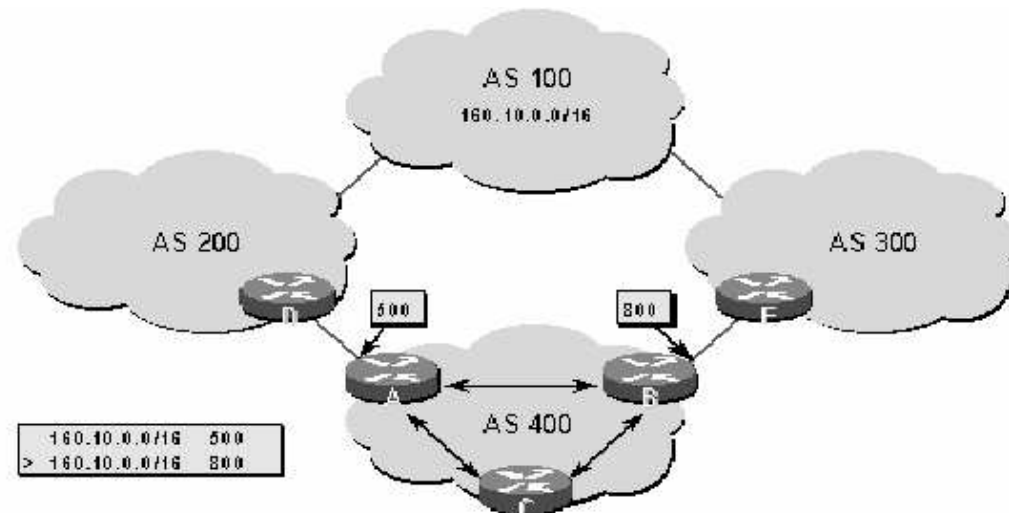
- ◆ eBGP & Next hop: To reach a certain destination network. For eBGP peers, the next-hop address is the IP address of the connection between the peers.



- ◆ Origin: IGP / EGP / Incomplete

Important BGP Attributes (contd)

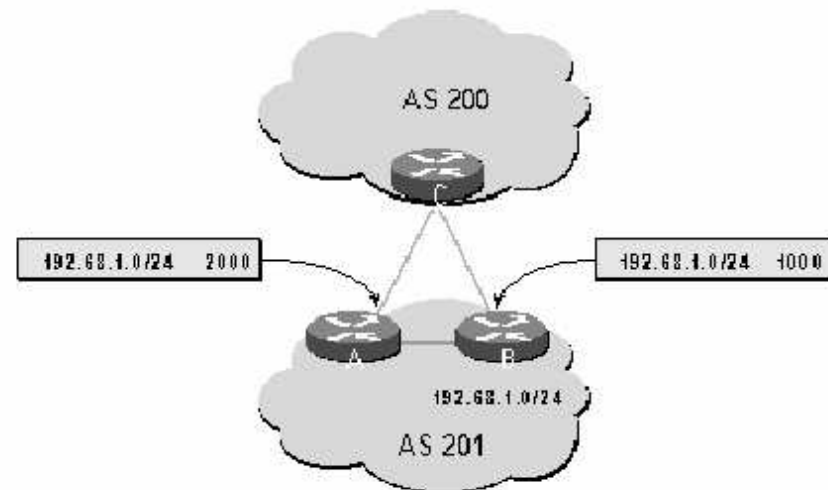
- ◆ Local Pref: When there are multiple paths to the same destination, used to influence BGP path selection
 - Local to AS
 - Used to prefer an exit point from the local AS
 - Path with highest local preference wins



Important BGP Attributes (contd)

◆ Multi-exit Discriminator (MED)

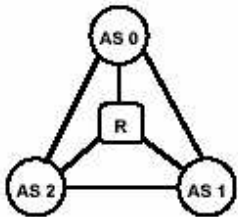
- When there are multiple entry points into the AS, used to convey the relative preference among entry points
- A suggestion to an external AS regarding the preferred route into the AS that is advertising the metric
- A lower MED value is preferred over a higher MED value



Useful Definitions

- ◆ Route Repair: A previously failed route to a network prefix is announced as reachable.
- ◆ Route Failover: A route is implicitly withdrawn and replaced by an alternative route with different next-hop or ASpath to a prefix destination.
- ◆ Policy Fluctuation: A route is implicitly withdrawn and replaced by an alternative route with different attributes, but the same next-hop and ASpath. (MED, etc).
- ◆ Pathological Routing: Repeated withdrawn or duplicate announcement the exact same route.

BGP bouncing problem



Stage	Time	Routing Tables	Messages Processing	Messages Queued in System
0	N/A	steady state 0(*R, 1R, 2R) 1(0R, *R, 2R) 2(0R, 1R, *R)		
1	N/A	R withdraws its route 0(-, *1R, 2R) 1(*0R, -, 2R) 2(*0R, 1R, -)	R -> 0 W R -> 1 W R -> 2 W	0 -> 1 01R 1 -> 0 10R 2 -> 0 20R 0 -> 2 01R 1 -> 2 10R 2 -> 1 20R
2	N/A	1 and 2 receive new announcement from 0 0(-, *1R, 2R) 1(-, -, *2R) 2(01R, *1R, -)	0 -> 1 01R 0 -> 2 01R	1 -> 0 10R 2 -> 0 20R 1 -> 0 12R 2 -> 0 21R 1 -> 2 10R 2 -> 1 20R 1 -> 2 12R 2 -> 1 21R
3	N/A	0 and 2 receive new announcement from 1 0(-, -, *2R) 1(-, -, *2R) 2(*01R, 10R, -)	1 -> 0 10R 1 -> 2 10R	2 -> 0 20R 1 -> 0 12R 2 -> 0 21R 0 -> 1 02R 2 -> 0 201R 2 -> 1 20R 1 -> 2 12R 2 -> 1 21R 0 -> 2 02R 2 -> 1 201R
4	N/A	0 and 1 receive new announcement from 2 0(-, -, -) 1(-, -, *20R) 2(*01R, 10R, -)	2 -> 0 20R 2 -> 1 20R	1 -> 0 12R 2 -> 0 21R 0 -> 1 02R 2 -> 0 201R 0 -> 1 W 1 -> 2 12R 2 -> 1 21R 0 -> 2 02R 2 -> 1 201R 0 -> 2 W
5	N/A	0 and 2 receive new announcement from 1 0(-, *12R, -) 1(-, -, *20R) 2(*01R, -, -)	1 -> 0 12R 1 -> 2 12R	2 -> 0 21R 0 -> 1 02R 2 -> 0 201R 0 -> 1 W 1 -> 0 120R 2 -> 1 21R 0 -> 2 02R 2 -> 1 201R 0 -> 2 W 1 -> 2 120R
6	N/A	0 and 1 receive new announcement from 2 0(-, *12R, 21R) 1(-, -, -) 2(*01R, -, -)	2 -> 0 21R 2 -> 1 21R	0 -> 1 02R 2 -> 0 201R 0 -> 1 W 1 -> 0 120R 0 -> 1 012R 0 -> 2 02R 2 -> 1 201R 0 -> 2 W 1 -> 2 120R 0 -> 2 012R
(steps omitted)				
48	N/A	steady state 0(-, -, -) 1(-, -, -) 2(-, -, -)		

- ◆ The above problem can be tempered by appropriate **MinRouteAdver**. An alternate method is to perform sender-side loop detection.

Stage	Time	Routing Tables	Messages Processing	Messages Queued in System
0	N/A	steady state 0(*R, 1R, 2R, 3R) 1(0R, *R, 2R, 3R) 2(0R, 1R, *R, 3R) 3(0R, 1R, 2R, *R)		steady state
1	N/A	R withdraws its route 0(-, *1R, 2R, 3R) 1(*0R, -, 2R, 3R) 2(*0R, 1R, -, 3R) 3(*0R, 1R, 2R, -)	R -> 0 W R -> 1 W R -> 2 W	R -> 3 W 0 -> 1 01R 1 -> 0 10R 2 -> 0 20R 3 -> 0 30R 0 -> 2 01R 1 -> 2 10R 2 -> 1 20R 3 -> 1 30R 0 -> 3 01R 1 -> 3 10R 2 -> 3 20R 3 -> 2 30R
2	N/A	announcement from 0 0(-, *1R, 2R, 3R) 1(-, -, *2R, 3R) 2(01R, *1R, -, 3R) 3(01R, *1R, 2R, -)	0 -> 1 01R 0 -> 2 01R 0 -> 3 01R	1 -> 0 10R 2 -> 0 20R 3 -> 0 30R 1 -> 2 10R 2 -> 1 20R 3 -> 1 30R 1 -> 3 10R 2 -> 3 20R 3 -> 2 30R
3	N/A	announcement from 1 0(-, -, *2R, 3R) 1(-, -, *2R, 3R) 2(*01R, 10R, -, 3R) 3(*01R, 10R, 2R, -)	1 -> 0 10R 1 -> 2 10R 1 -> 3 10R	2 -> 0 20R 3 -> 0 30R 2 -> 1 20R 3 -> 1 30R 2 -> 3 20R 3 -> 2 30R
4	N/A	announcement from 2 0(-, -, -, *3R) 1(-, -, 20R, *3R) 2(01R, 10R, -, *3R) 3(*01R, 10R, 20R, -)	2 -> 0 20R 2 -> 1 20R 2 -> 3 20R	3 -> 0 30R 3 -> 1 30R 3 -> 2 30R
Min Route Timer expires 5	30	announcement from 3 0(-, -, -, -) 1(-, -, *20R, 30R) 2(*01R, 10R, -, 30R) 3(*01R, 10R, 20R, -)	3 -> 0 30R 3 -> 1 30R 3 -> 2 30R	0 -> 1 W 1 -> 0 120R 2 -> 0 201R 3 -> 0 301R 0 -> 2 W 1 -> 2 120R 2 -> 1 201R 3 -> 1 301R 0 -> 3 W 1 -> 3 120R 2 -> 3 201R 3 -> 2 301R
6	N/A	withdrawal from 0 0(-, -, -, -) 1(-, -, *20R, 30R) 2(-, *10R, -, 30R) 3(-, *10R, 20R, -)	0 -> 1 W 0 -> 2 W 0 -> 3 W	1 -> 0 120R 2 -> 0 201R 3 -> 0 301R 1 -> 2 120R 2 -> 1 201R 3 -> 1 301R 1 -> 3 120R 2 -> 3 201R 3 -> 2 301R
7	N/A	announcement from 1 0(-, -, -, -) 1(-, -, *20R, 30R) 2(-, -, -, *30R) 3(-, 120R, *20R, -)	1 -> 0 120R 1 -> 2 120R 1 -> 3 120R	2 -> 0 201R 3 -> 0 301R 2 -> 1 201R 3 -> 1 301R 2 -> 3 201R 3 -> 2 301R
8	N/A	announcement from 2 0(-, -, -, -) 1(-, -, -, *30R) 2(-, -, -, *30R) 3(-, 120R, *201R, -)	2 -> 0 201R 2 -> 1 201R 2 -> 3 201R	3 -> 0 301R 3 -> 1 301R 3 -> 2 301R
Min Route Timer expires 9	60	announcement from 3 0(-, -, -, -) 1(-, -, -, -) 2(-, -, -, *301R) 3(-, *120R, 201R, -)	3 -> 0 301R 3 -> 1 301R 3 -> 2 301R	1 -> 0 W 2 -> 0 2301R 3 -> 0 3120R 1 -> 2 W 2 -> 1 2301R 3 -> 1 3120R 1 -> 3 W 2 -> 3 2301R 3 -> 2 3120R
10	N/A	withdrawal from 1 0(-, -, -, -) 1(-, -, -, -) 2(-, -, -, *301R) 3(-, -, *201R, -)	1 -> 0 W 1 -> 2 W 1 -> 3 W	2 -> 0 2301R 3 -> 0 3120R 2 -> 1 2301R 3 -> 1 3120R 2 -> 3 2301R 3 -> 2 3120R
11	N/A	announcement from 2 0(-, -, -, -) 1(-, -, -, -) 2(-, -, -, *301R) 3(-, -, -, -)	2 -> 0 2301R 2 -> 1 2301R 2 -> 3 2301R	3 -> 0 3120R 3 -> 1 3120R 3 -> 2 3120R
Min Route Timer expires 12	90	announcement from 3 0(-, -, -, -) 1(-, -, -, -) 2(-, -, -, -) 3(-, -, -, -)	3 -> 0 3120R 3 -> 1 3120R 3 -> 2 3120R	2 -> 0 W 3 -> 0 W 2 -> 1 W 3 -> 1 W 2 -> 3 W 3 -> 2 W
13	N/A	process withdrawals 0(-, -, -, -) 1(-, -, -, -) 2(-, -, -, -) 3(-, -, -, -)	2 -> 0 W 2 -> 1 W 2 -> 3 W 3 -> 0 W 3 -> 1 W 3 -> 2 W	

BGP bouncing problem - MinRouteAdver

- ◆ Applied only to announcements (at least according to BGP RFC)
- ◆ $30*(N-3)$ delay due to creation mutual dependencies. Provide proof that $N-3$ rounds necessarily created during bounded BGP MinRouteAdver convergence
- ◆ Rounds due to
 - Ambiguity in the BGP RFC and lack senderside loop detection
 - Inclusion of BGP withdrawals with MinRouteAdver (in violation of RFC) – Cisco bug solved in IOS 2000

BGP bouncing problem – Theorems

- ◆ For a complete graph of 'N' nodes, $O((N-1)!)$ distinct paths exist to reach a particular dest.
- ◆ With adoption of MinRouteAdver, the lower bounds on convergence for BGP requires at least (N-3) rounds.
- ◆ For complete graphs of size $N \leq 3$, BGP converges within a single MinRouteAdver period for a route withdrawal
- ◆ If loop detection is performed on both the sender and receiver side, all dependencies will be discovered within a single round

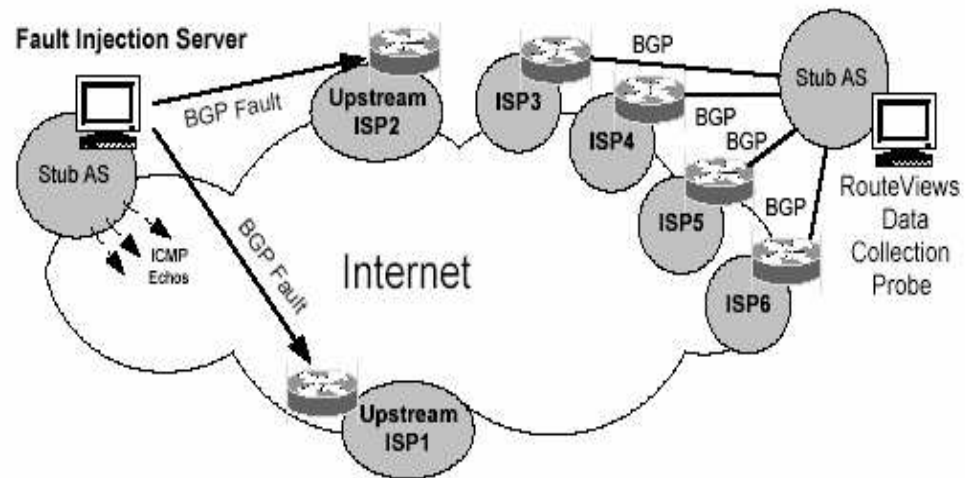
Nodes	Time	States	Messages	Nodes	Time	States	Messages	Nodes	Time	States	Messages
4	N/A	12	41	4	30	11	26	4	30	11	26
5	N/A	60	306	5	60	26	54	5	30	23	54
6	N/A	320	2571	6	90	50	92	6	30	39	92
7	N/A	1955	23823	7	120	85	140	7	30	59	140

(a) Unbounded

(b) MinRouteAdver

(c) Modified

Experiment Methodology



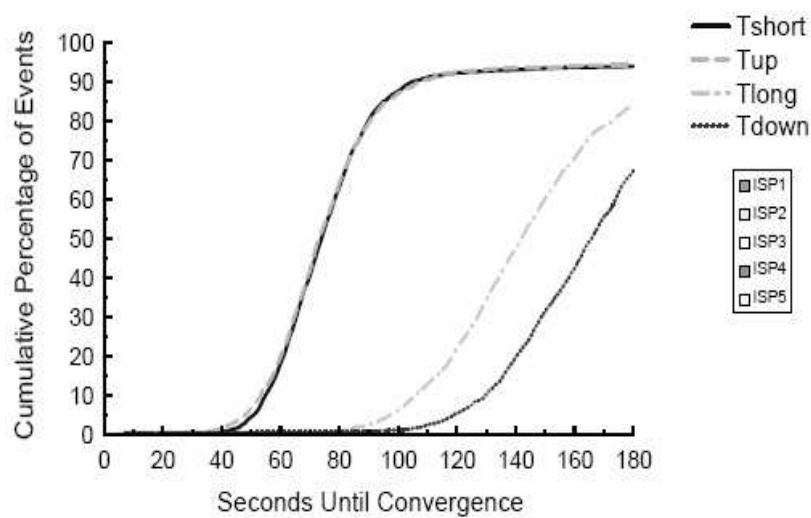
fault injection and measurement infrastructure.

- ◆ Injected over 250,000 routing faults into geographically and topologically distributed peering sessions over a two year period
- ◆ Monitor impact in two ways:
 - Active – Monitored end-to-end performance
 - Passive – RouteViews probe which peered with over 25 ISPs
- ◆ Establish primary path and longer backup path

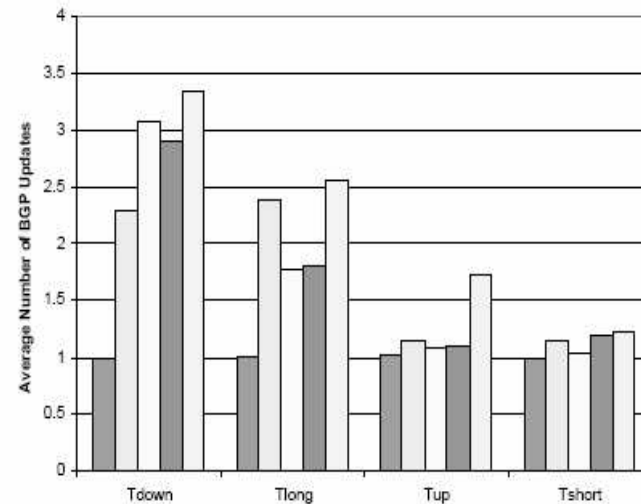
Routing events injected

- ◆ Tup: A new route is advertised (Represents route repair too)
- ◆ Tdown: A route is withdrawn (i.e. single-homed failure)
- ◆ Tshort: Advertise a shorter/better ASPath (i.e. primary path repaired)
- ◆ Tlong: Advertise a longer/worse ASPath (i.e. primary route failure and failover)

Passive - Convergence for Tup, Tshort, Tlong and Tdown events



(a) Latency



(b) Messages

◆ Observations:

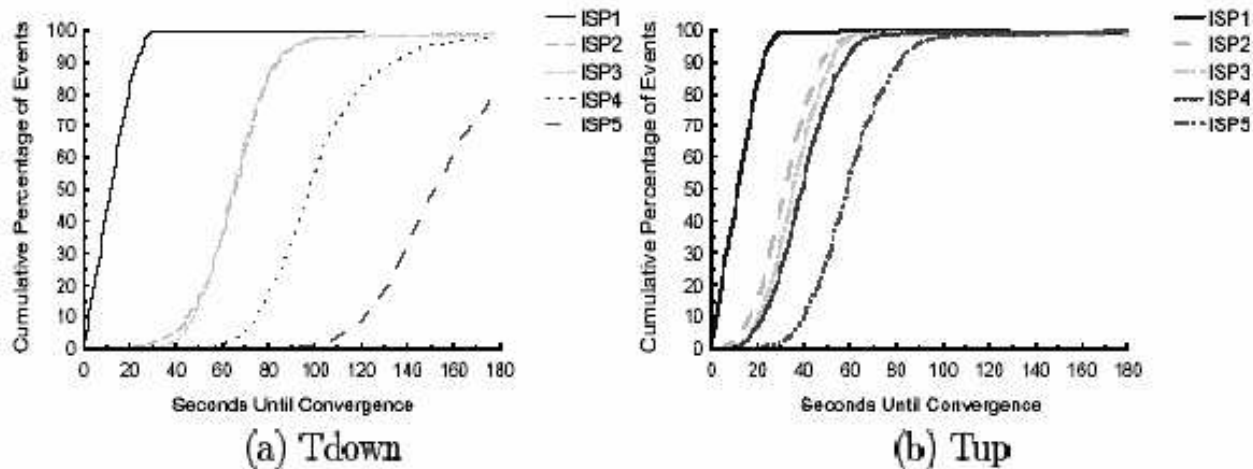
- Less than half of Tdown events converge within two minutes
- Tup/Tshort and Tdown/Tlong form equivalence classes
- Long tailed distribution (up to 15 minutes)
- ISP1 always had only 1 BGP update

Passive - Convergence for Tup, Tshort, Tlong and Tdown events

- ◆ Routing convergence requires an order of magnitude longer than expected (10s of minutes)
- ◆ Routes converge more quickly following Tup/Repair than Tdown/Failure events (“bad news travels more slowly”)
- ◆ Curiously, withdrawals (Tdown) generate several times the number of updates than announcements (Tup)

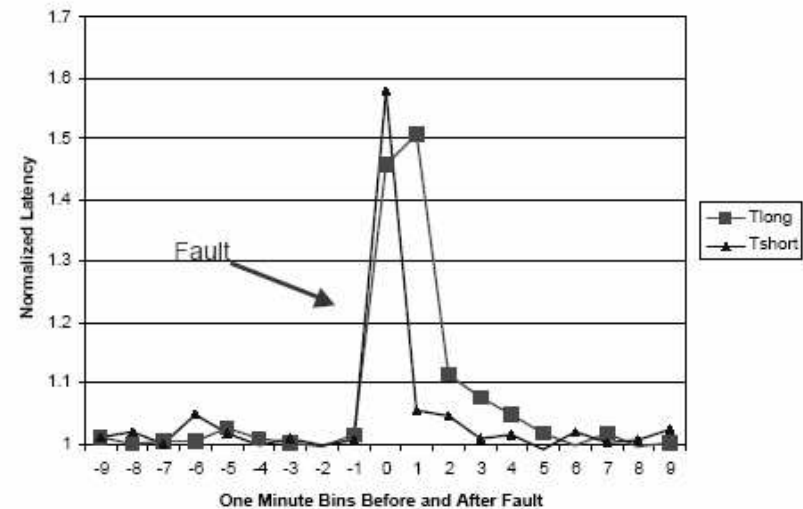
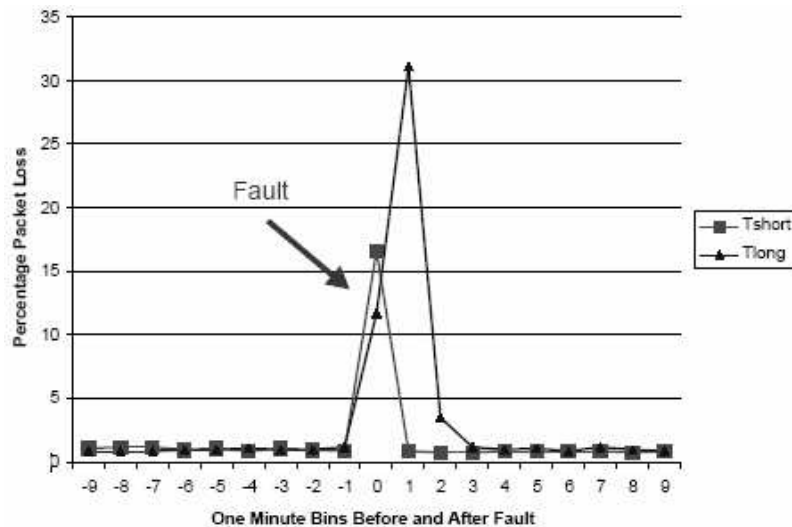
<u>TIME</u>	<u>BGP Message/Event</u>
10:40:30	Route Fails/Withdrawn by AS2129
10:41:08	2117 announce 5696 2129
10:41:32	2117 announce 1 5696 2129
10:41:50	2117 announce 2041 3508 3508 4540 7037 1239 5696 2129
10:42:17	2117 announce 1 2041 3508 3508 4540 7037 1239 5696 2129
10:43:05	2117announce 2041 3508 3508 4540 7037 1239 6113 5696 2129
10:43:35	2117 announce 1 2041 3508 3508 4540 7037 1239 6113 5696 2129
10:43:59	2117 sends withdraw

Passive - Convergence for Tup, Tshort, Tlong and Tdown events



- ◆ 3 min gap separates 80% of ISP1 converged events from ISP5
- ◆ Data showed no correlation between convergence latency and geographical/network distance.
- ◆ ISP3 in Japan converged faster than ISP5 in Canada.
- ◆ ISP1 – Hardly any EBGP oscillations because it has shortest ASPath
- ◆ Also: No temporal relationship with failover latency (Convergence delay and time of day). Hence, dependent on network load and congestion.

Active – End-to-end measurements



- ◆ Based on 512 byte ICMP echoes sent to 100 randomly chosen websites every second during the 10 minutes before and after the fault injection
- ◆ Tlong/Tshort exhibit similar relationship as before. In both, latencies more than tripled for the 3 mins following the fault.
- ◆ For T_{up}, 80% websites replied in 30 secs (100% before 1 min)
- ◆ Delayed convergence explains Paxson's observations.

Summary of Delayed Routing Convergence

- ◆ Path vectors remove the count-to-infinity problem, but routing table oscillations are exponentially exacerbated.
- ◆ The delay in inter-domain path failovers averaged 3 mins during the two years of study.
- ◆ The theoretical upper bound on computational states is $O(N!)$ with N being number of AS. The bound is very theoretical
- ◆ Lower bound is $\Omega((n-3)*30)$ secs
- ◆ Packet loss grew by a factor of 30 and latency by factor of 4 (during path restoral)
- ◆ Minor changes to vendor implementations could reduce lower bound to $\Omega(30)$ secs

Experimental Study of Internet Stability - Methodology

Inter-domain BGP data collection

- ◆ RouteView probe : participate in remote BGP peering session. Collected 9GB complete routing tables of 3 major ISPs in US.
- ◆ About 55,000 route entries
- ◆ Each of the three ISPs under test have varying size, architecture and transmission technology.



Map of major U.S. Internet exchange points.

Experimental Study of Internet Stability – Methodology (contd)

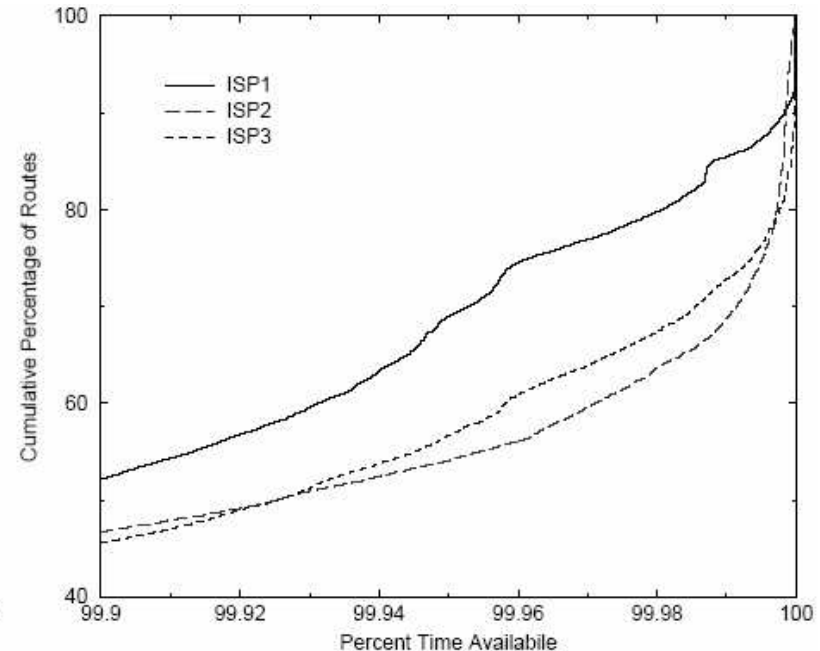
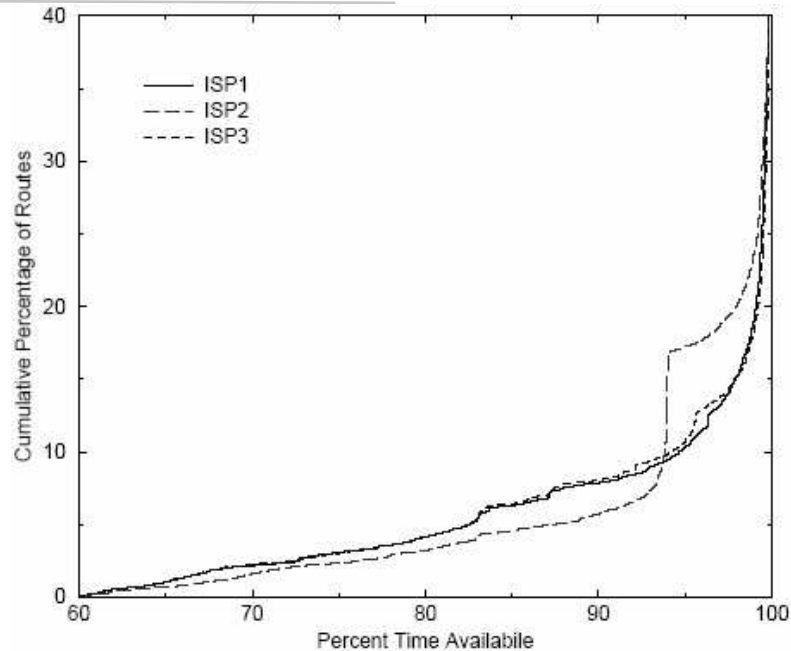
Intra-domain BGP data collection

- ◆ Medium size regional network --- MichNet Backbone.
- ◆ Contains 33 backbone routers with several hundred customer routers.

Data from:

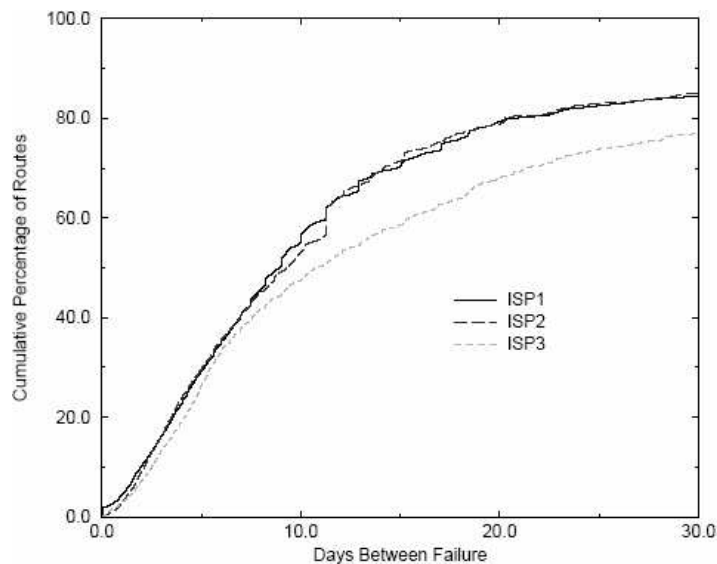
- ◆ A centralized network management station (CNMS) log data
 - Ping every router interfaces every 10 minutes.
 - Used to study frequency and duration of failures.
- ◆ Network Operations Center (NOC) log data.
 - CNMS alerts lasting more than several minutes.
 - Prolonged degradation of QoS to customer sites.
 - Used to study network failure category.

Inter-domain Route Availability

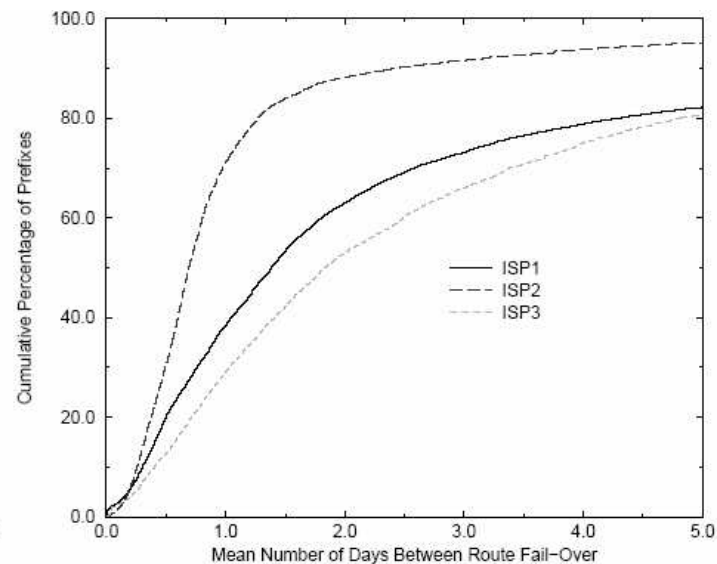


- ◆ Default-free route availability in the 10 month study period
- ◆ Only between 30-35% of ISP2 & ISP3 and 25% of routes from ISP1 had availability > 99.99% of study period
- ◆ ISP1 had significantly less availability above 99.9% than others.

Inter-domain failure analysis



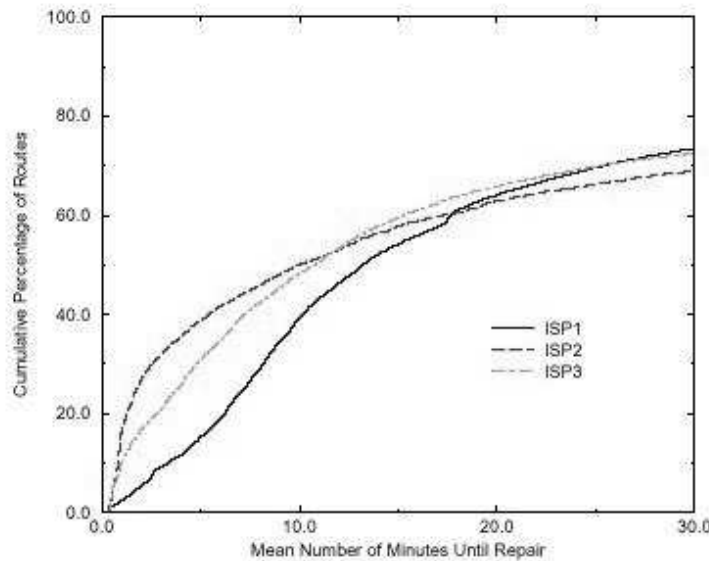
(a) Failure



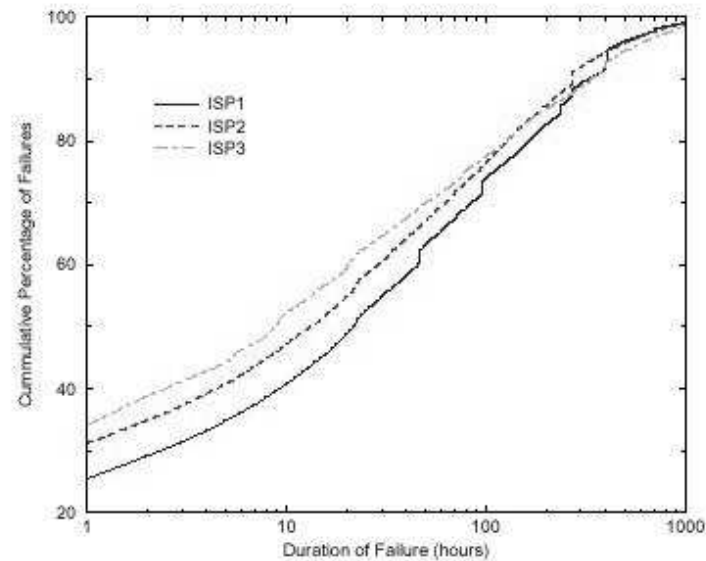
(b) Fail-Over

- ◆ More than 50% routes exhibit a Mean-Time-to-Fail (MTTF) ≥ 15 days
- ◆ By end of 30 days, more than 75% of routes had failed atleast once
- ◆ All ISPs are similar. Results diverge only after 10days.
- ◆ (b) focuses only on paths with backup
- ◆ 20% routes from ISP1 and ISP3, 5% from ISP2 do not failover in 5 days

Inter-domain failure analysis (contd)



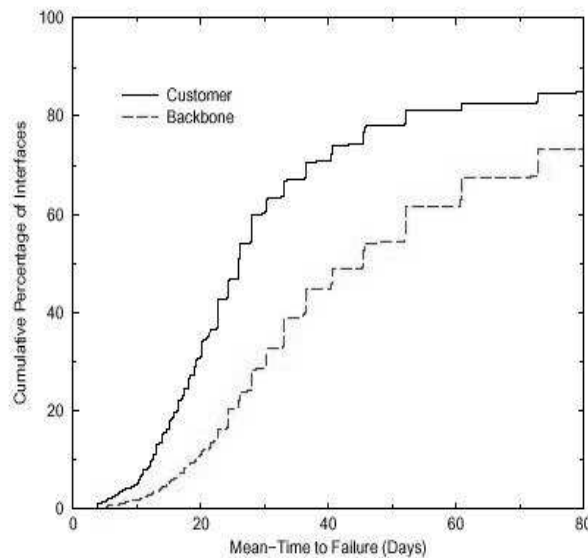
(a) MTTR



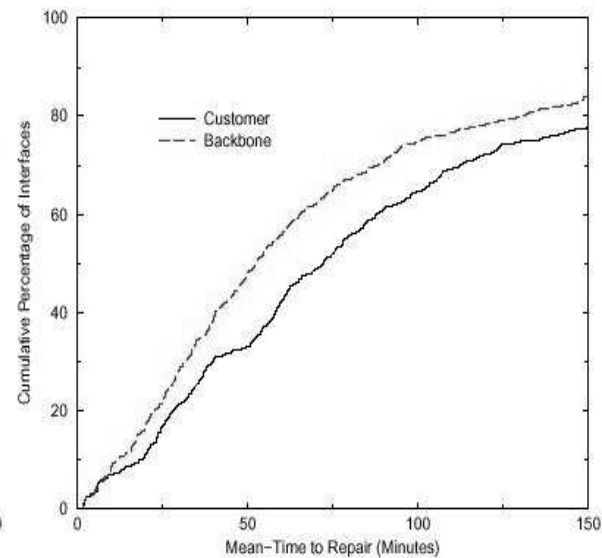
(b) Duration

- ◆ 40% of failures repaired in under 10 mins. 60% repaired within 30 mins.
- ◆ Demonstrates heavy-tailed distributions after 30mins, with slow asymptotic growth towards 100%.
- ◆ Duration curve for ISP1 rises at a slower rate than MTTR curve
- ◆ ISP1's lower avg MTTF and slower MTTR contribute to lower availability

Intra-domain failure analysis



(a) MTTF



(b) MTTR

- ◆ 40% of all interfaces experienced a failure within average of 40 days
- ◆ 5% failed within a mean time of 5 days.
- ◆ Contrastingly, BGP inter-domain failures occur within 30 days. Why?
- ◆ The steps are caused mainly because of simultaneous failures.

Failure analysis

- ◆ The data is taken from MichNet NOC log data.
- ◆ Most outages were not related to IP backbone infrastructure.
- ◆ Most outages were from customer sites than backbone nodes.

Outage Category	Number of Occurrences	Percentage
Maintenance	272	16.2
Power Outage	273	16.0
Fiber Cut/Circuit/Carrier Problem	261	15.3
Unreachable	215	12.6
Hardware problem	154	9.0
Interface down	105	6.2
Routing Problems	104	6.1
Miscellaneous	86	5.9
Unknown/Undetermined/No problem	32	5.6
Congestion/Sluggish	65	4.6
Malicious Attack	26	1.5
Software problem	23	1.3

Table1: Category and number of recorded outages Internet in MichNet. (11/97~11/98)

Observation of availability of backbone

- ◆ Data is taken from CNMS monitor logs.
- ◆ Overall up time is 99.0% for the year.
- ◆ Failure logs reveal a number of persistent circuit or hardware faults repeatedly happened.
- ◆ Operation staffs said: (NOC log data has no duration statistics)
 - Most backbone outages tend be on order of several minutes.
 - Customer outages persist longer on order of several hours.
 - Power outages and hardware failure tend to be resolved within 4 hours.
 - Routing problem last within 2 hours.

Summary of Experimental Study

- ◆ Internet backbone has less availability and a lower meantime to failure than the Public Switched Telephone Network (PSTN).
- ◆ Majority of Internet backbone paths have MTTF \leq 25 days, and a MTTR \leq 20 mins.
- ◆ Internet backbones are rerouted (either due to failure or policy changes) on the average of once every three days or less.
- ◆ Routing instability inside of an autonomous network does not exhibit the same daily and weekly cyclic trends as previously reported for routing between Inter provider backbones, suggesting that most inter-provider path failures stem from congestion collapse.
- ◆ A small fraction of network paths in the Internet contribute disproportionately to the number of long-term outages and backbone unavailability.

Correlation between Routing Loops & BGP Updates

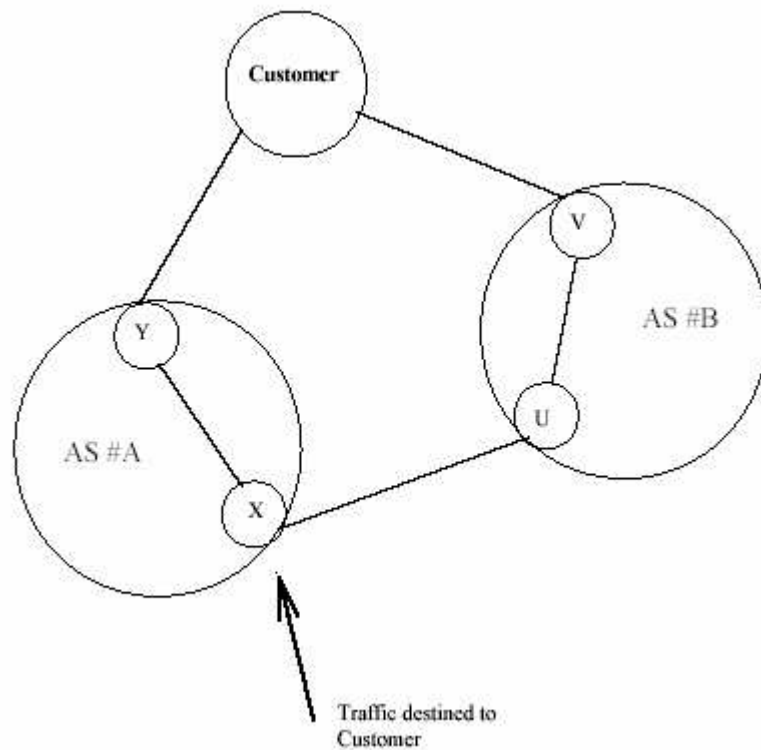
- ◆ Loops are caused by inconsistencies in state
 - Transient – Because of normal convergence. Short lived.
 - Persistent – Because of misconfiguration. Last longer.

- ◆ Hypothesis:
 - Strong correlation exists between routing loops and BGP updates / ISIS update

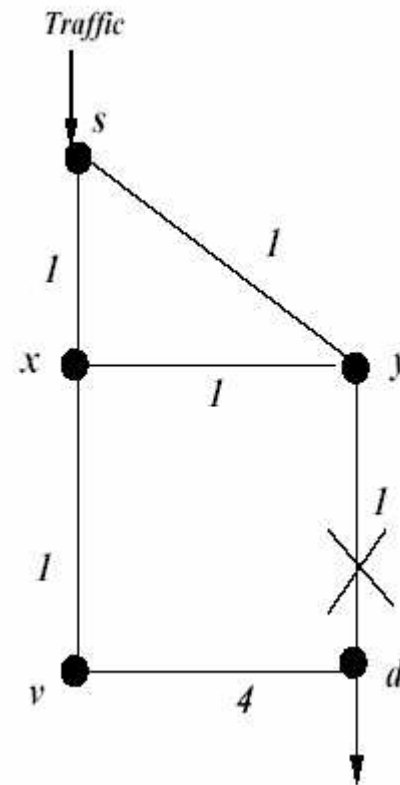
- ◆ Technique:
 - Robust loop detection scheme
 - Matching technique to associate loops with updates
 - Apply and confirm in Sprint backbone

- ◆ Discusses the factors influencing loop path lengths

Illustration of Routing Loops



BGP Update



ISIS Failure

Identifying routing loops

- ◆ Packet loop is unique for each packet
- ◆ Step 1: Packet traces collected in Sprint backbone through IPMON machines in various PoPs
- ◆ Step 2: Hash packets into different hash buckets and examine each bucket to see if there are identical (Ignore TTL and IP checksum)
- ◆ Step 3: So as not to ignore loops that span hash bucket boundaries, they maintain a history of all loops that are within 500 ms of the current hash bucket boundary and check against that first.

Trace Name	No. of Packet Loops	Duration (hrs)
NYC-20	2476	1
NYC-21	3838	1
NYC-23	1895	1
NYC-22	8672	12
NYC-24	719	12
NYC-25	1691	12

Number of Packet Loops in Each Trace

Matching BGP Update

1. We determine if any packet loop is potentially impacted by a BGP update through a longest prefix match for the destination address of the packet loop on the set of advertised and/or withdrawn routes in the update.
2. Next we determine if the BGP update lies in the temporal vicinity of the loop. This was set to a value of 2 minutes.
3. If both previous conditions are satisfied, then we examined any change in the current next hop or AS path of the destination prefix by feeding the update to a Zebra router which emulates the BGP decision process.
4. If the first 2 conditions are satisfied and a change in next hop or AS Path is detected we conjecture that the loop was caused by this update.

Conditions for ISIS Loops

- ◆ Need to define necessary and sufficient conditions
- ◆ Condition 1: A necessary condition is the change in the forwarding path of the ingress node of the observed link possessing loop.
- ◆ Condition 2: Either case must hold
 - Case 1: The observation node has updated its path but a set of nodes on the new path, that were originally pointing to the observation node at time t , have not yet updated their paths in response to the change
 - Case 2: Similar to Case 1, but here the observation node is yet to update.
- ◆ A BGP update can change the egress router. Hence, that must not be confused with a ISIS event.

Experiments & Results - ISIS

- ◆ None of the loops could be correlated with an ISIS event
- ◆ Reason:
 - Multiple forwarding paths supported by ISIS causes immediate switchover
 - Also ISIS uses complete topology to compute paths.
- ◆ Consequence: Convergence time of ISIS is immaterial

Experiments & Results - BGP

Trace	% Transient & BGP Updates	% Persistent BGP Updates	% Persistent No Updates	Total
NYC-20	40.1	0	50.8	90.8
NYC-21	80.2	0	7.5	87.9
NYC-23	3.3	0	0	3.3
NYC-22	18.8	0	80.6	99.4
NYC-24	70.0	0	0	70.0
NYC-25	43.7	15.5	0	59.2

BGP Update Matches for Loops using Sprint Link Information

- ◆ As can be seen from the table, for most traces we were able to account for more than half the loops, as either identifiable with a BGP event or persistent (& unidentifiable with a BGP event)
- ◆ Note:
 - The first potential factor is the presence of loops that are persistent in nature and originate before the trace collection.
 - NYC-25 associated all persistent loops with BGP, while NYC-24 and NYC-23 did not have any persistent loops

Experiments & Results – BGP (contd)

- It is possible that some of the changes happened external to the Sprint network. (particularly for NYC-23). This brings in the geographical significance. Wider destination distribution may lead to poorer matching ratios.

Trace	Avg. No. Of ASes traversed
NYC-20	1.34
NYC-21	1.04
NYC-23	1.74
NYC-22	0.513
NYC-24	1.61
NYC-25	1.63

Trace	% Sprint Matches	% RouteViews Matches
NYC-20	40.1	43.1
NYC-21	80.2	82
NYC-23	3.3	10.6

BGP Update Matches for Loops using RouteViews Information

BGP Beacons

- ◆ BGP Beacons refer to a publicly documented prefix having global visibility and a published schedule for announcements and withdrawals.
- ◆ There are currently two groups of Beacons:

PSG
Uses 198.133.206.0/24,
192.135.183.0/24, 203.10.63.0/24
with period of two hours and
198.32.7.0/24 of variable period

Have timestamps, sequence
numbers and anchor prefixes

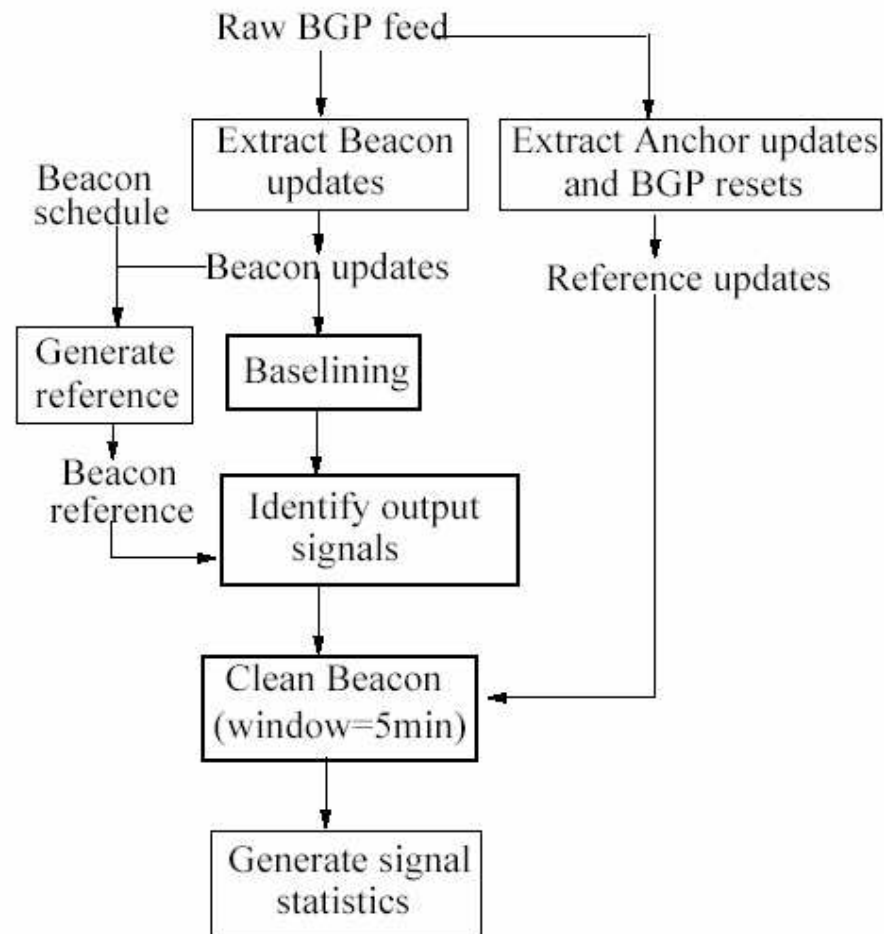
Is not

RIPE
Uses 195.80.224.0/24 through
195.80.232.0/24 with period of 2
hrs.

Does not

Associated with BGP routing
monitors

Data cleaning & Signal Identification



Data cleaning (contd)

Table 2: Effect of cleaning on observed *announcement* signals (Route Views): signal count, average duration, delay, and length

Beacon	Before cleaning				After cleaning			
	count	avgDur (sec)	avgDelay (sec)	avgSigLen	count	avgDur (sec)	avgDelay (sec)	avgSigLen
1	33536	27.13	50.60	1.47	33318 (99.35%)	19.36	41.89	1.47
2	34522	9.13	29.56	1.20	33726 (97.69%)	6.75	25.21	1.17
3	32504	10.82	34.99	1.22	32188 (99.03%)	5.77	28.40	1.21
4	39044	41.95	63.66	1.52	37970 (97.25%)	22.79	43.16	1.46

Table 3: Effect of cleaning on observed *withdrawal* signals (Route Views): signal count, average duration, delay, and length

Beacon	Before cleaning				After cleaning			
	count	avgDur (sec)	avgDelay (sec)	avgSigLen	count	avgDur (sec)	avgDelay (sec)	avgSigLen
1	33443	37.88	100	2.07	33261 (99.46%)	32.98	90.09	2.07
2	33860	45.24	109.23	2.19	33344 (98.48%)	42.94	94.38	2.19
3	32379	59.16	120.64	2.55	31182 (96.30%)	56.36	114.40	2.55
4	36633	96.33	139.63	3.43	35776 (97.66%)	75.65	115.90	3.41

- ◆ Less than 5% of the signals have been deleted after cleaning. The signal length seems to be almost the same.

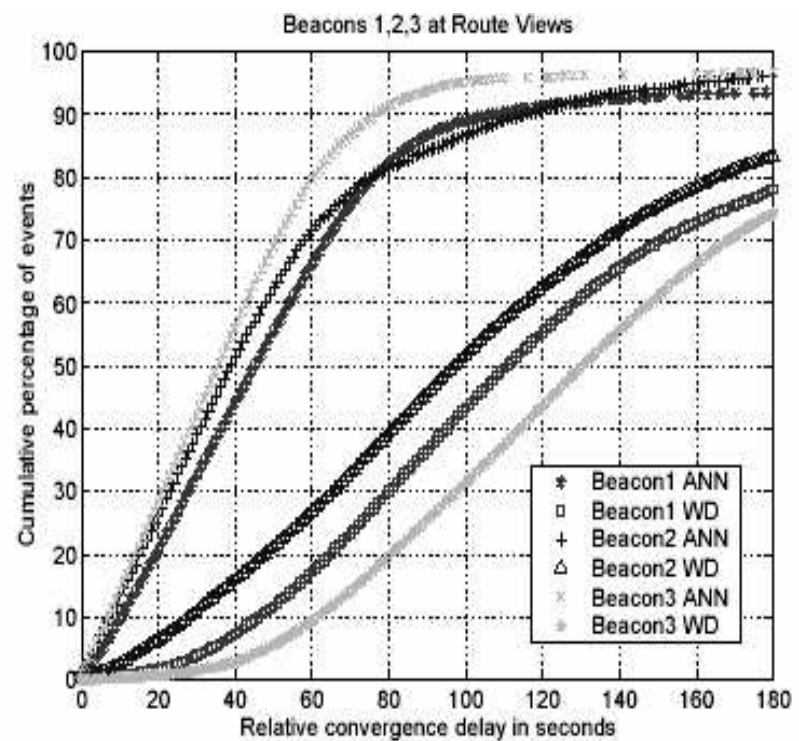
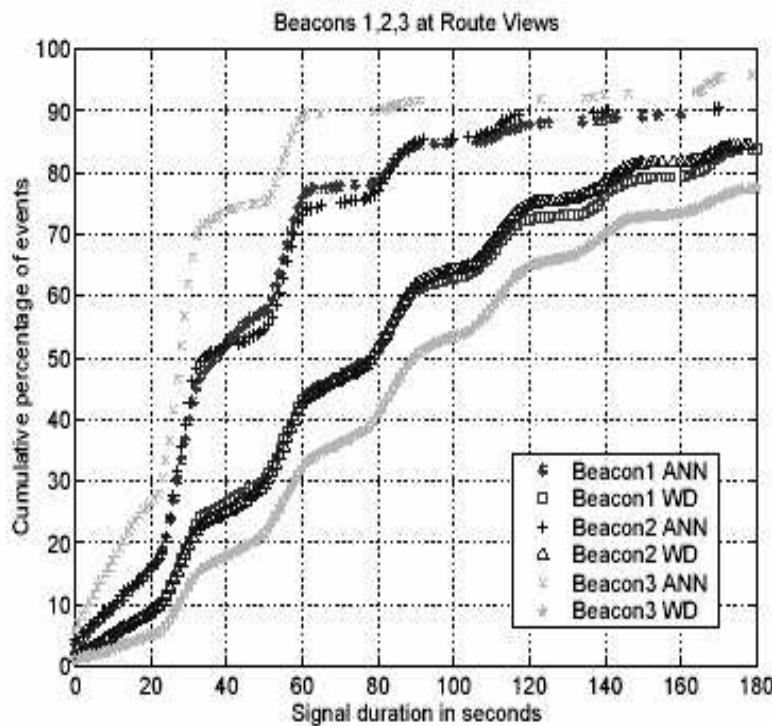
Cisco Vs Juniper

- ◆ Juniper routers send 25% more updates, has a similar update duration and a smaller average inter-arrival time for update.
- ◆ The number of short inter-arrival times is higher for Juniper.
- ◆ Signal length isn't correlated to Signal duration for Juniper as the rate limiting algorithm allows updates sent in burst.
- ◆ Cisco signal duration are multiples of 30 secs. While Juniper has a spread out. (Note: Announcement duration < Withdrawal duration)
- ◆ Cisco is more aggressive than Juniper in route suppression

Peer	Type	signal length		duration		inter-arrival		% of short inter-arrivals	
		A	W	A	W	A	W	A	W
147.28.255.1	Cisco	1.20	2.07	6.79	48.4	34.8	45.4	1.56	0.44
147.28.255.2	Juniper	1.50	2.49	7.13	44.3	14.2	29.6	12.76	4.37

Summary of BGP Beacons

- ◆ Using the Beacons for BGP convergence analysis:



- ◆ A model could be constructed to make use of all BGP attributes.

References

1. Craig Labovitz, Abha Ahuja, Farnam Jahanian, "Experimental Study of Internet Stability and Backbone Failures," FTCS 1999.
2. Craig Labovitz, Abha Ahuja, Abhijit Bose, Farnam Jahanian "Delayed Internet routing convergence," IEEE/ACM Transactions on Networking, pp 293-306 (2001)
3. Craig Labovitz, G. Robert Malan, Farnam Jahanian, "Origins of Internet Routing Instability," INFOCOM 1999, pp 218-226.
4. Craig Labovitz, G. Robert Malan, Farnam Jahanian, "Internet routing instability," IEEE/ACM Transactions on Networking, pp 515-528 (1998)
5. Ashwin Sridharan, Sue. B. Moon, C. Diot, "On the Correlation between Route Dynamics and Routing Loops," Proceedings of IMC 2003.
6. Zhuoqing Mao, Randy Bush, Timothy Griffin, Matthew Roughan, "BGP Beacons," Proceedings of IMC 2003.