Experimental Measurement of Delayed Convergence

Abha Ahuja

Internap/Merit Network, Inc.

<ahuja@umich.edu>

Craig Labovitz

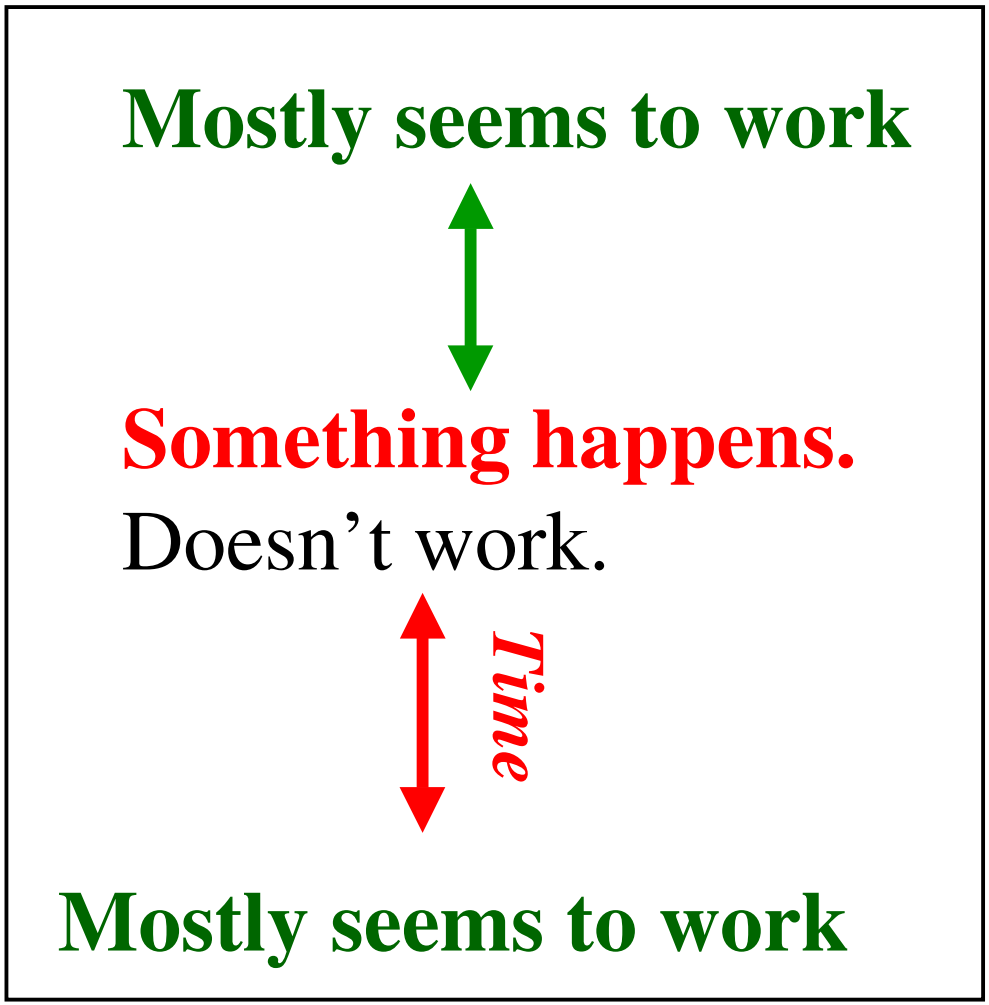
Microsoft Research/Merit Network, Inc.

Farnam Jahanian, Abhijit Bose

University of Michigan

Apricot - 2000

The Internet: Failure Analysis



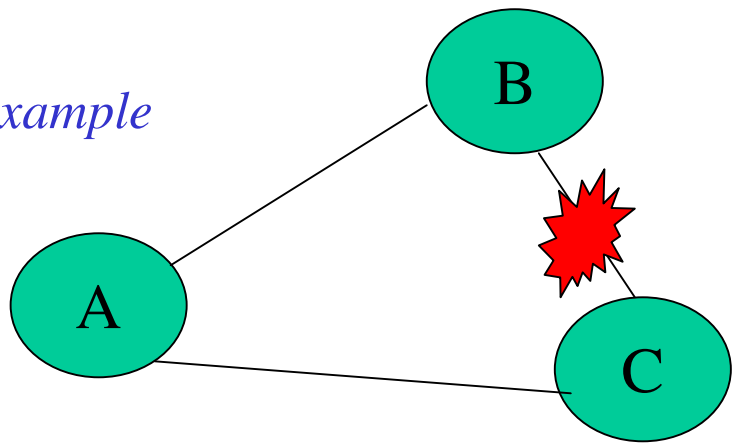
Routing Protocol Convergence

- Unlike connection oriented PSTN (~30 ms), Internet does not have fail-over.
- Instead, each node recalculates on a hop-per-hop basis (i.e. no flooding of changes)
- Distance-vector algorithms (e.g. RIP, BGP) exhibit slower convergence than link state protocols
- During convergence
  - Latency, loss, out of order
  - Additional update messages (CPU processing)

Distance Vector (BF) Protocols

- Suffer from counting to infinity problem
- Solutions
  - Poison reverse
  - Split horizon
  - Path vectors

Example



Conventional Wisdom

- "Restoral is not an issue in the IP world"
  - Just reroute around in a few milliseconds or whatever
- BGP convergence takes only a few _____
- "Bad news travels fast"
  - Fast withdraw propagation valid goal
  - Announcements slower because bundled
- BGP has great convergence properties
  - ASPath solved the convergence and counting to infinity problems
- All my customers are multi-homed, triple-homed
  - Convergence -- *what, me worry?*

More Conventional Wisdom

- Enough bandwidth will solve anything

"It will all be one big network one day soon anyways"

(Especially after yesterday)

Internet Failures

- Replication, round-robin DNS, etc. helps reliability of inter-domain content oriented services
- Inter-domain **transaction oriented services** (e.g. VoIP, EBay, database commits, etc.) still pose a challenge
- Important model how long does it take for the Internet to converge
  - After Failure
  - After Fail-Over
  - After Repair

BGP: Bad news

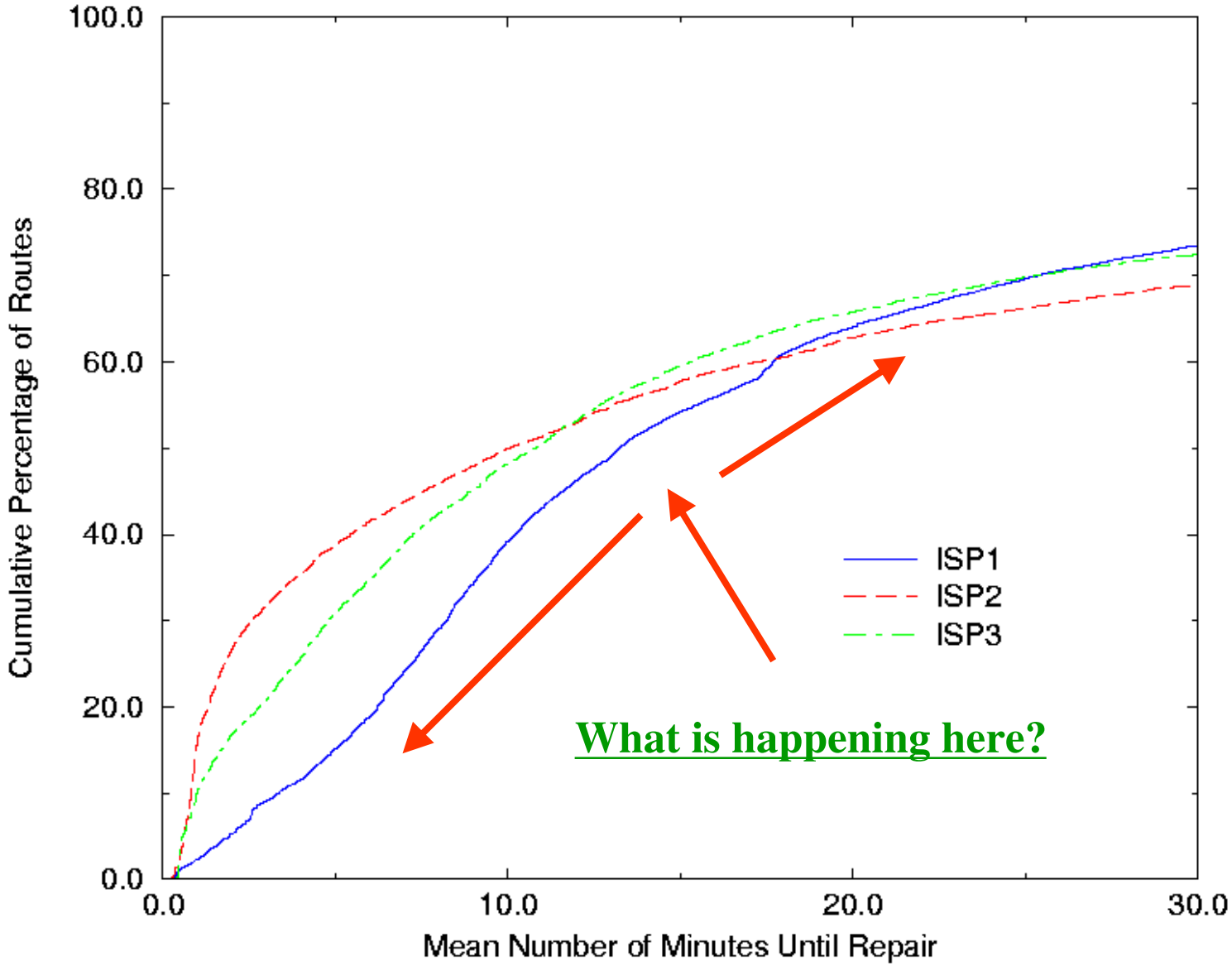
- With unconstrained policies (Griffin99, Varadhan96)
  - Divergence
  - Possible create mutually unsatisfiable policies
  - NP-complete to identify these policies in IRR
  - Happening today?
- With constrained policies (e.g. shortest path first)
  - Transient oscillations
  - BGP usually converges
  - It might just take a very long time….
- This talk is about constrained policies

Some Observations

- How do we study convergence?
  - From BGP logs (e.g. debug ip bgp), difficult to determine causal relationships
  - Earlier work studied BGP pathologies and failures
  - Still lots of BGP duplicates and oscillations
- Failure/repair data (next slide) for default-free routes shows 30 minute curve
  - Examined long-lived default-free routes from 24 providers for a year
  - Restoral time for given provider after failure (i.e. route withdrawn)

How long until routes return?

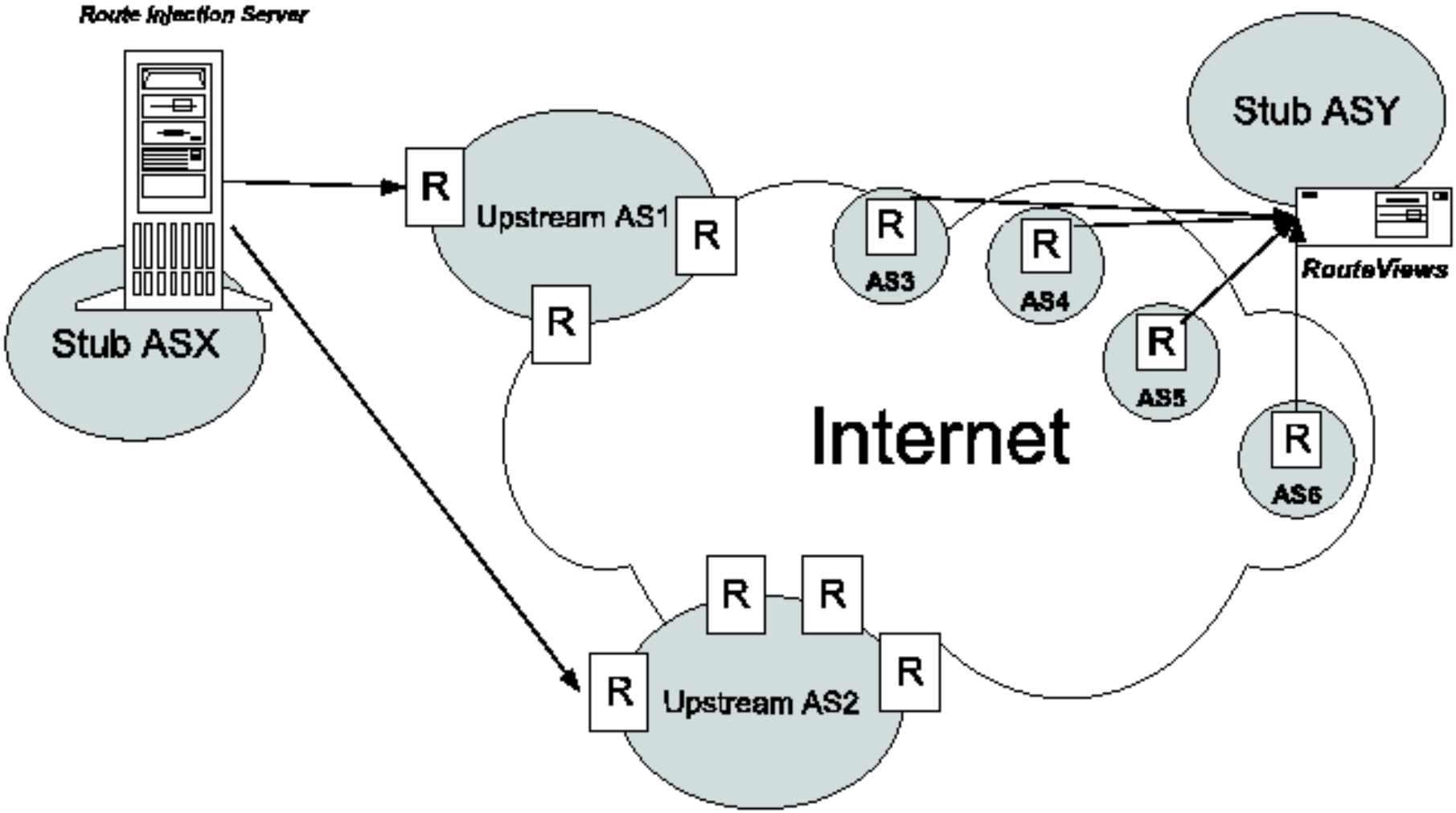
(From A Study of Internet Failures)



16 Month Study of Convergence

- Instrument the Internet
  - Inject routes into geographically and topologically diverse provider BGP peering sessions (Mae-West, Japan, Michigan, London)
  - Periodically fail and change these routes (i.e. send withdraws or new attributes)
  - Time events using ICMP echos and NTP synchronized BGP “routeviews” monitoring machines (also http gets)
  - Write lots of Perl scripts
  - Wait a sixteen months… (45,000 routing events)

Setup


Route Injection Server
Stub ASX
R
Upstream AS1
R
R
R
AS3
R
AS4
Stub ASY
RouteViews
R
AS5
R
AS6
Internet
R
R
R
R
Upstream AS2

How Many Announcements Does it Take For an AS to Withdraw a Route?

7/5	19:33:25	Route **R** is **withdrawn**	
7/5	19:34:15	AS6543 **announce** **R**	6543 66665 8918 1 5696 999
7/5	19:35:00	AS6543 **announce** **R**	6543 66665 8918 67455 6461 5696 999
7/5	19:35:37	AS6543 **announce** **R**	6543 66665 4332 6461 5696 999
7/5	19:35:39	AS6543 **announce** **R**	6543 66665 5378 6660 67455 6461 5696 999
7/5	19:35:39	AS6543 **announce** **R**	6543 66665 65 6461 5696 999
7/5	19:35:52	AS6543 **announce** **R**	6543 66665 6461 5696 999
7/5	19:36:00	AS6543 **announce** **R**	6543 66665 5378 6765 6660 67455 6461 5696 999
		…	
7/5	19:38:22	AS6543 **withdraw** **R**	

(AS6543 chosen as an example – all AS'es exhibit similar behavior)

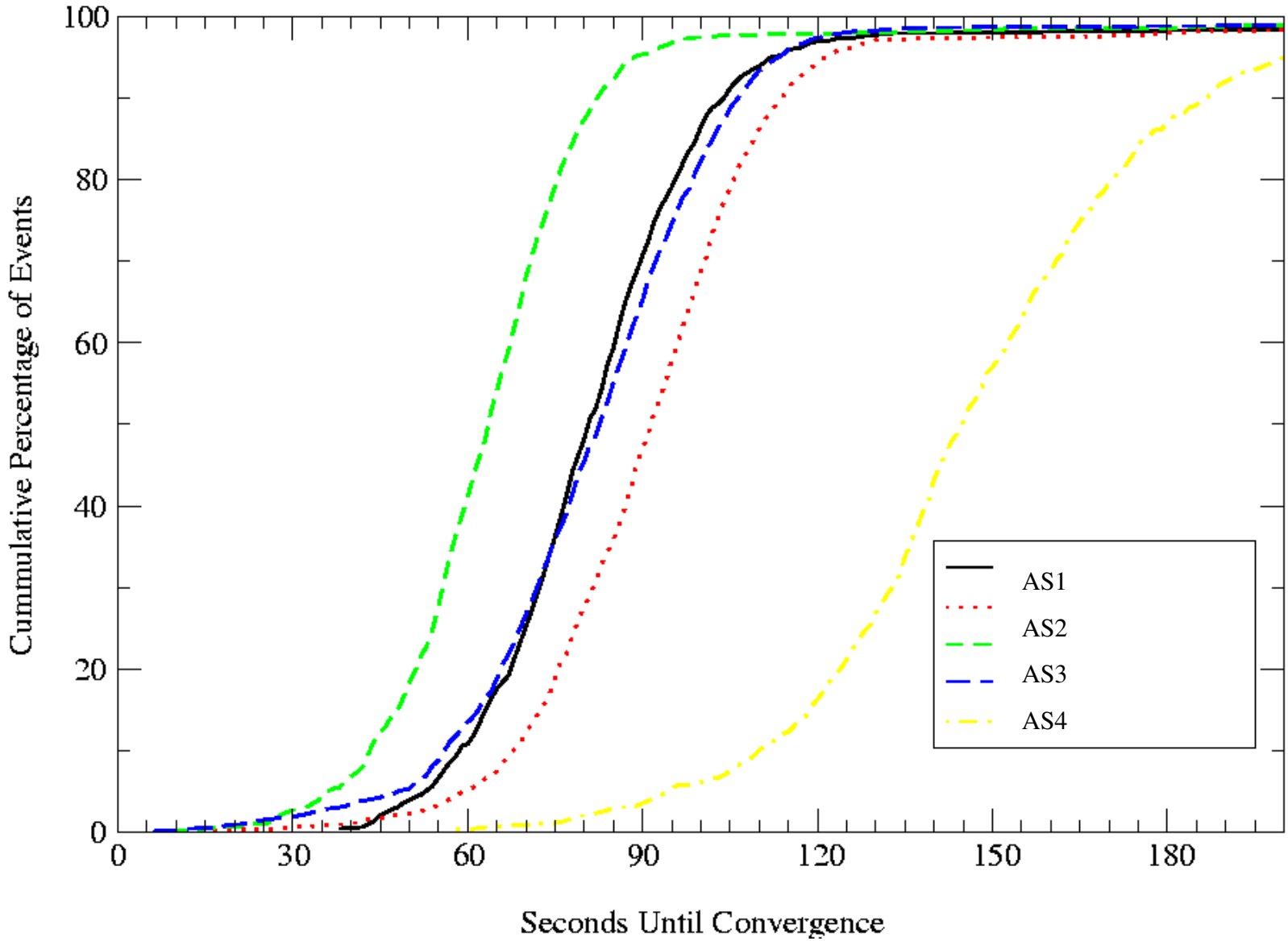
Abha made me change the AS numbers

Answer: Up to 19

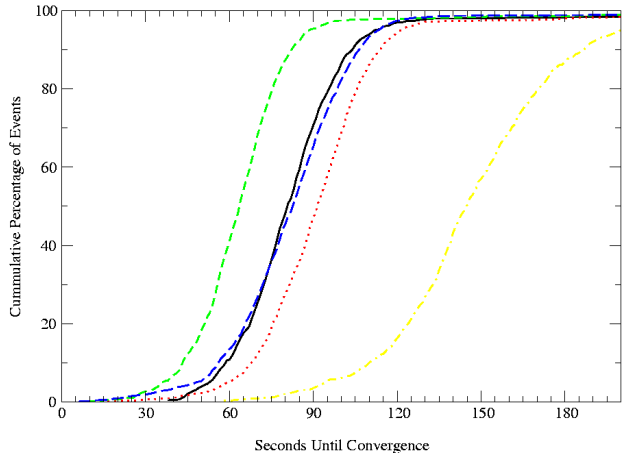
Withdraw Convergence

After a BGP route is withdrawn, barring other failures, how long does it take Internet routing tables to reach steady-state?

Withdraw Convergence



Withdraw Convergence



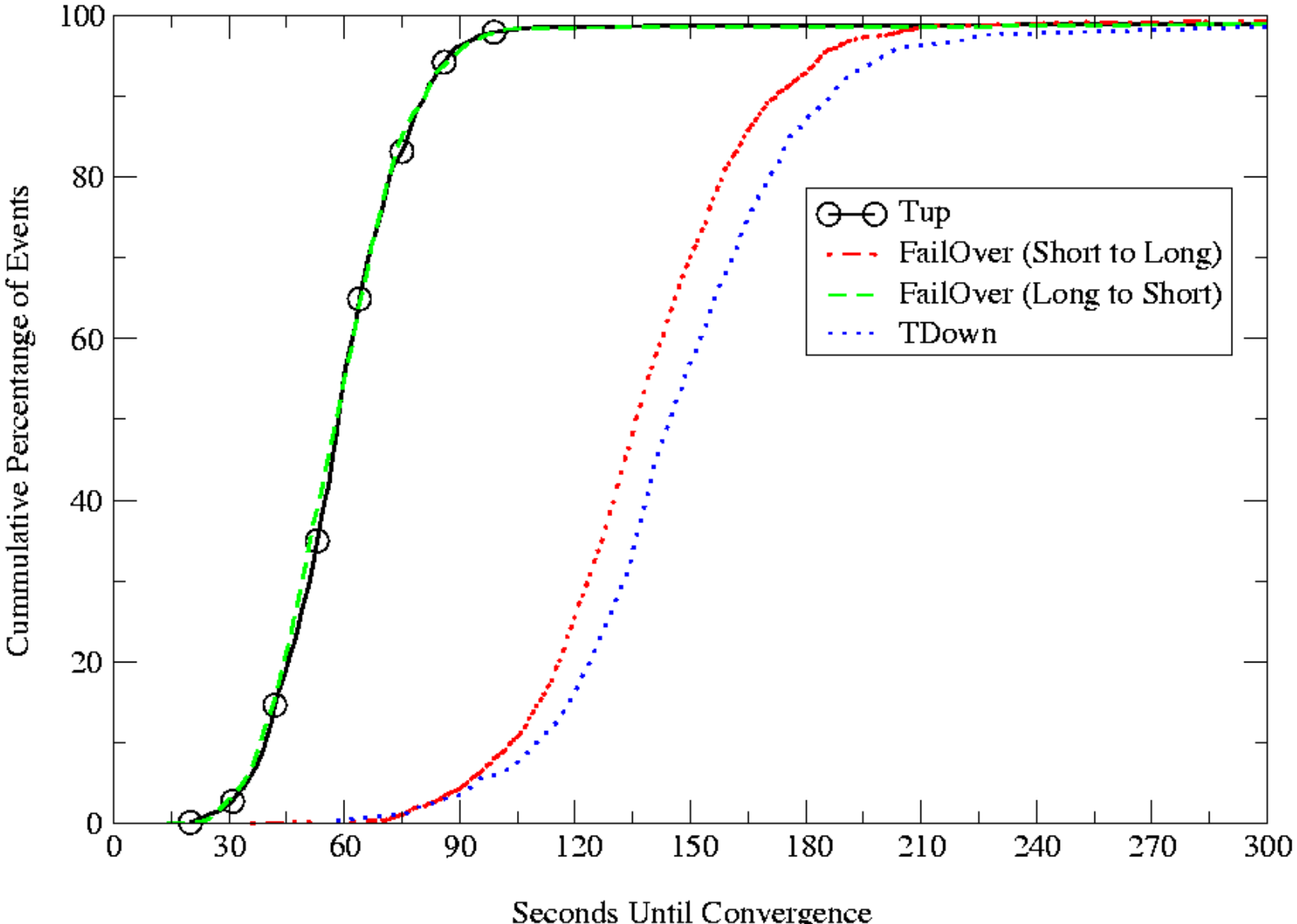
- Probability distribution
- Providers exhibit different, but related convergence behaviors
- 80% of withdraws from all ISPs take more than a minute
- For ISP4, 20% withdraws took more than three minutes to converge

Fail-Overs and Repairs

What are the relative convergence latencies for fail-overs and repairs?

Does bad news (withdraws) travel faster?

Failures, Fail-overs and Repairs



Failures, Fail-overs and Repairs

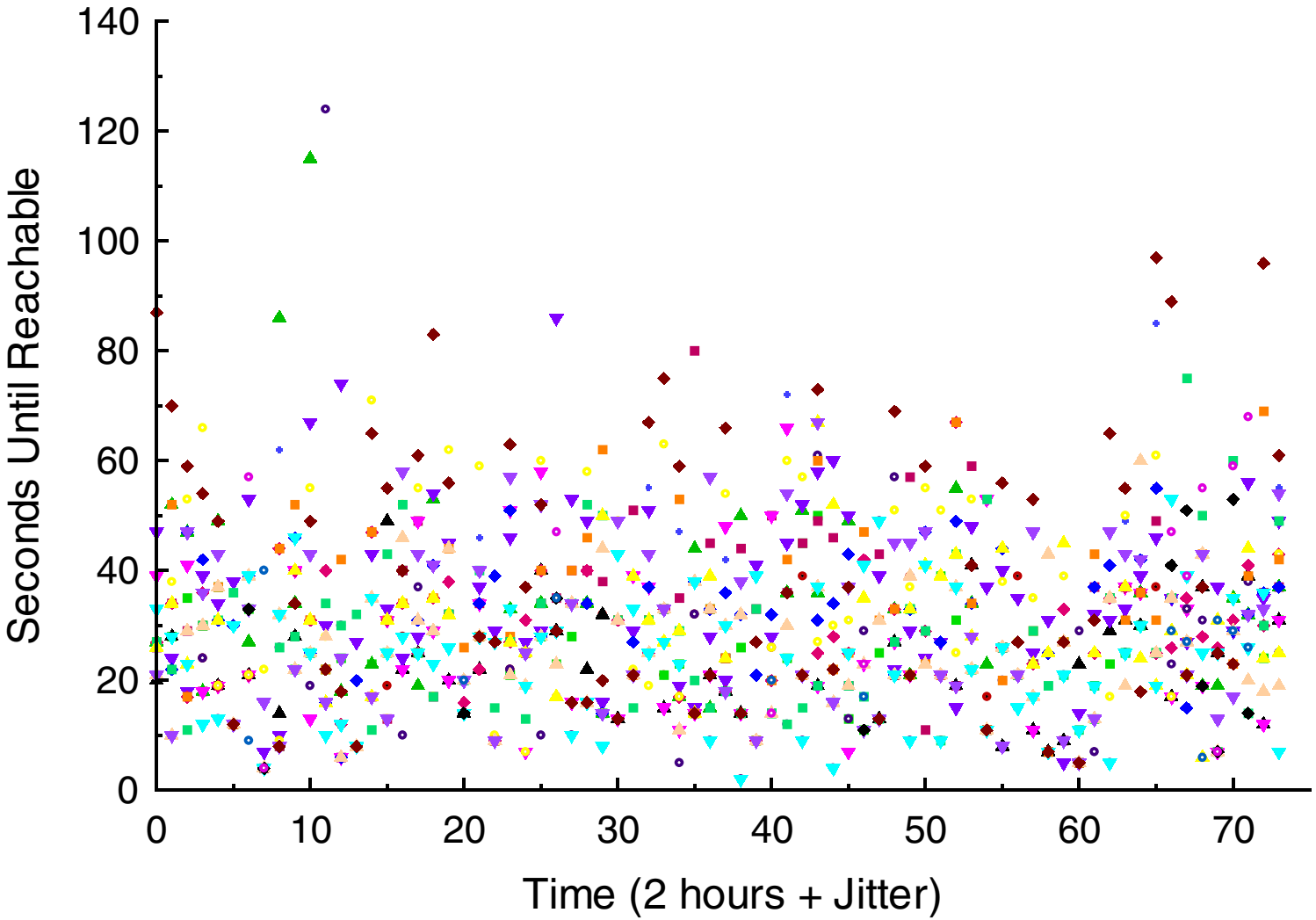
- Bad news does not travel fast…
- Repairs (Tup) exhibit similar convergence properties as long-short ASPath fail-over
- Failures (Tdown) and short-long fail-overs (e.g. primary to secondary path) also similar
  - Slower than Tup (e.g. a repair)
  - 60% take longer than two minutes
  - Fail-over times degrade the **greater** the degree of multi-homing!

End2End Connectivity

After a repair, how long before my site is reachable?

- Modified ICMP pings and HTTP sent once a second
- Source IP address block of pseudo-AS
- 100 randomly chosen web sites from parent cache logs

ICMP Response after Repairs



What is Happening?

- Non-deterministic ordering of BGP update messages leads to
  - Transient oscillations
  - Each change in FIB adds delay (CPU, BGP bundling timer)
  - At extreme, convergence triggers BGP dampening

BGP Bad News

Given best current routing practices, inter-domain BGP convergence times degrade **exponentially** with increase in the degree of interconnectivity for a given route

… and the degree of inter-connectivity (multi-homing, transit, etc) is increasing