Towards Harmless Maintenance Operations in IP Networks

Pierre Francois

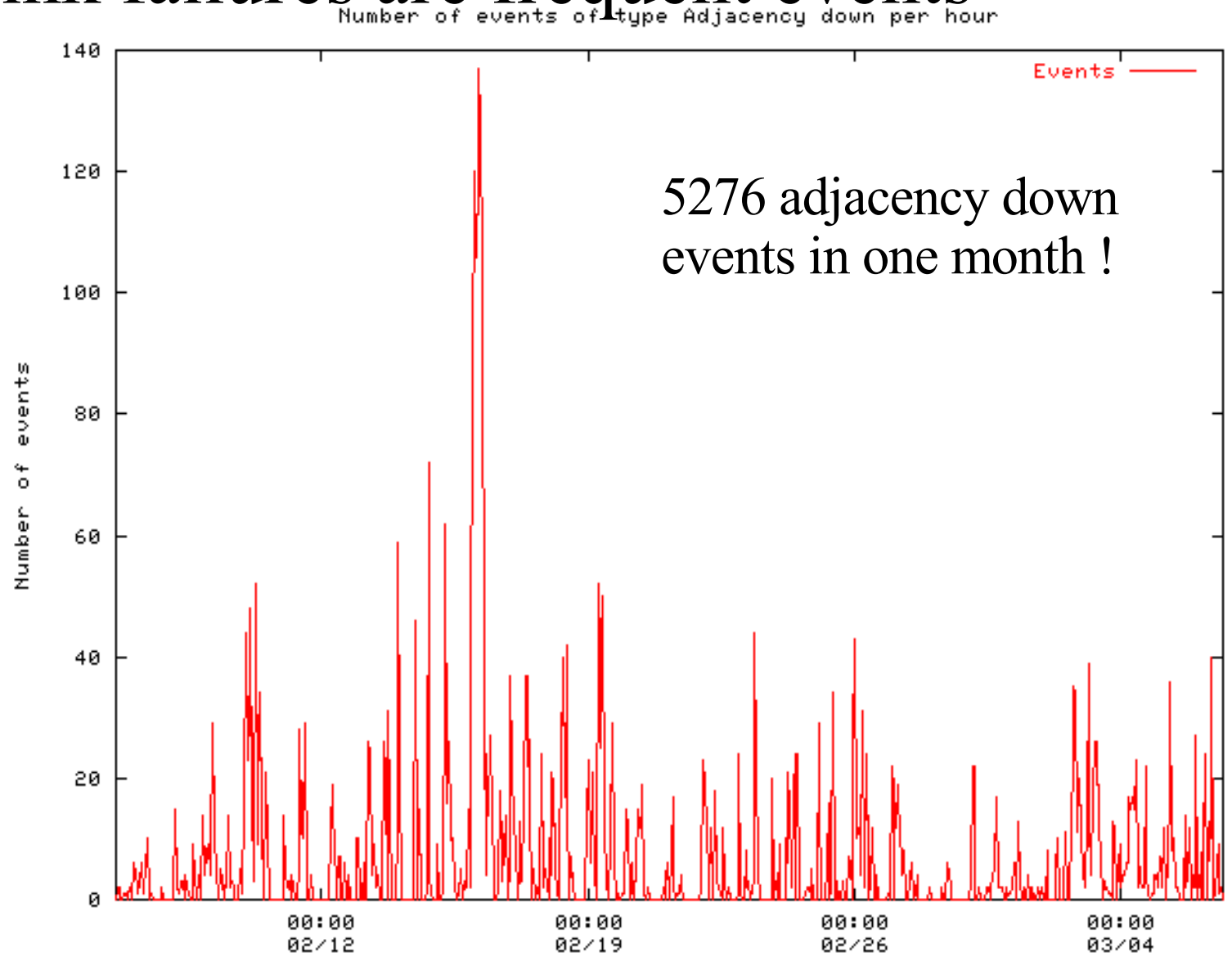
francois@info.ucl.ac.be

Agenda

- Link failures/maintenance operations in IP Networks
- Packet loss during predictable IGP convergence
  - Solutions
- Packet loss / routing failures during predictable eBGP peering down operation
  - Towards solutions

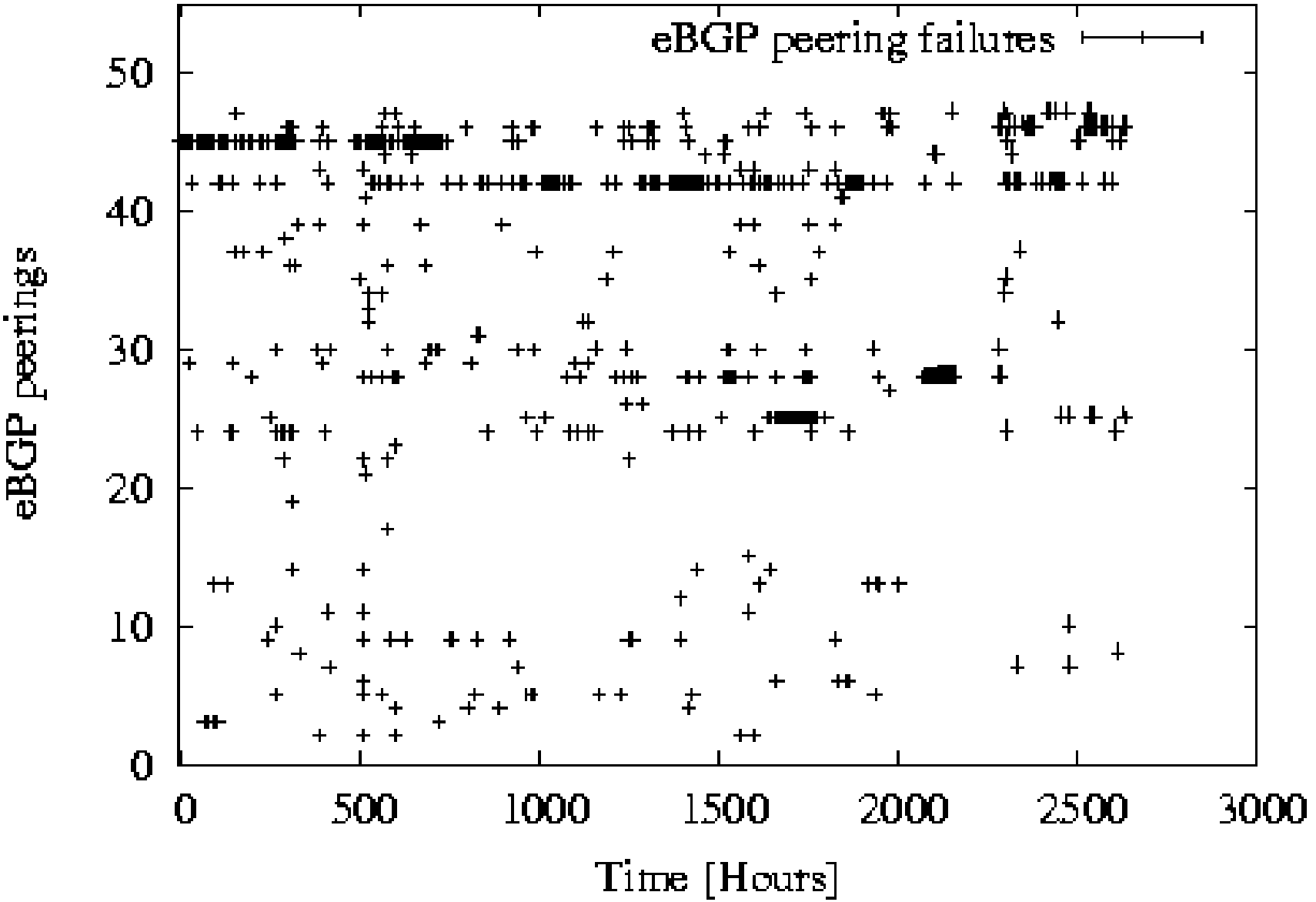
Link failures in IP Networks

- Link failures are frequent events



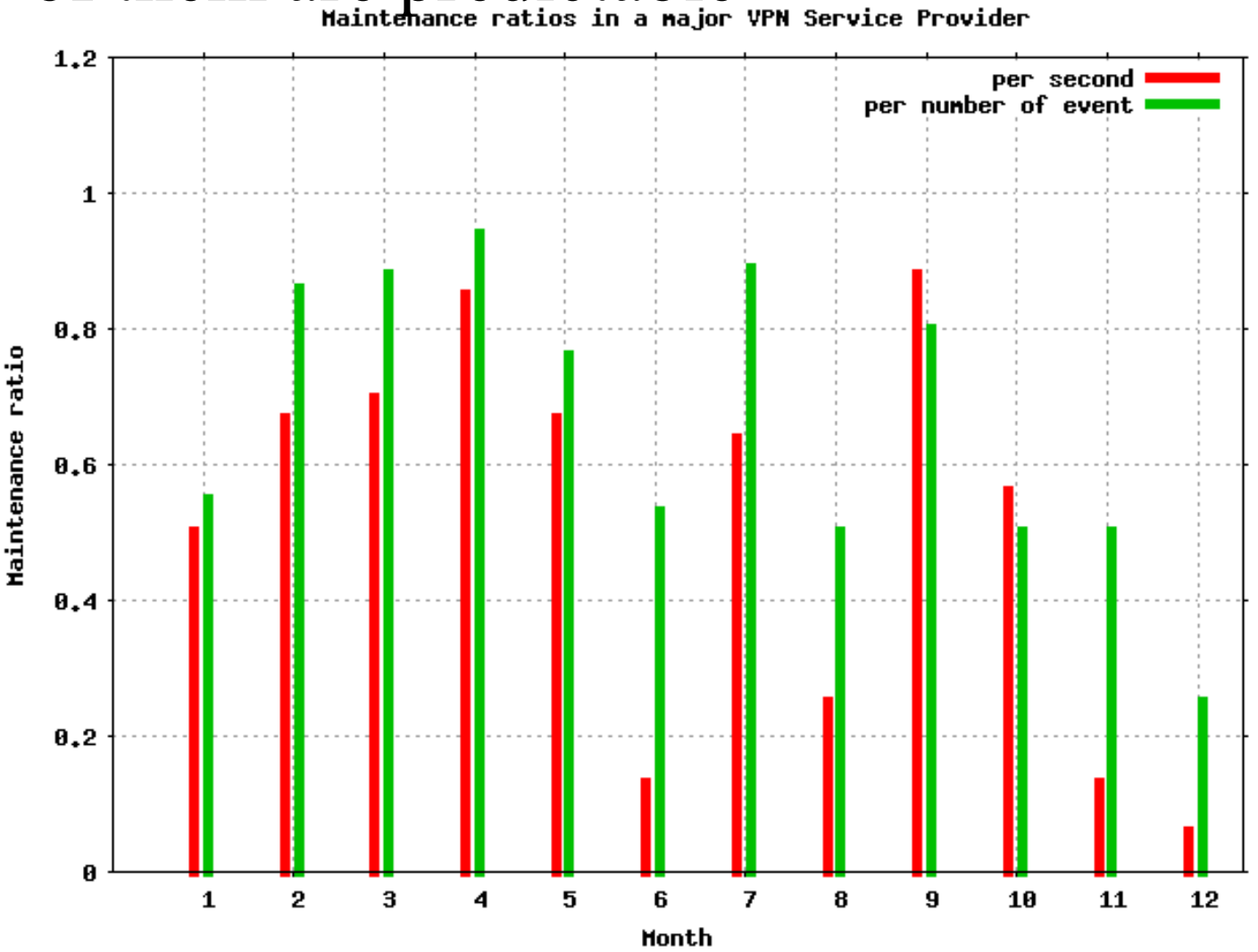
Failures of eBGP peering links

- Failures of eBGP links are also frequent events



Failures of eBGP peering links

- Many of them are predictable



See also : A. Markopoulou, G. Iannaconne, S. Bhattacharrya, C-N. Chuah, and C. Diot, "Characterization of Failures in an IP Backbone," IEEE INFOCOM, March 2004.

Sudden failures

- Sudden failures can be turned into non urgent failures
  - IP-FRR
  - MPLS-FRR
  - BGP-FRR[1]
- Reachability is recovered once
  - failure is detected
  - protection is activated

[1]Achieving Sub-50 Milliseconds Recovery Upon BGP Peering Link Failures, Olivier Bonaventure , Clarence Filsfils, Pierre François , In Proceedings of ACM CoNext, 2005

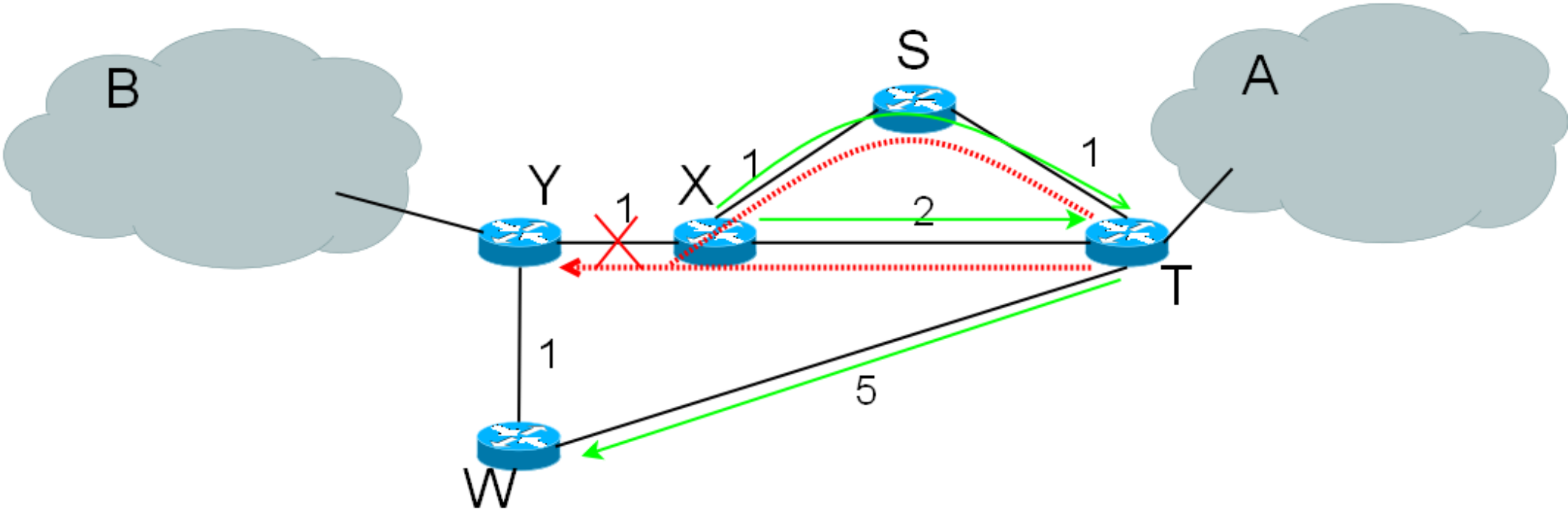
So what's the problem ?

- IGP and iBGP convergence leads to packet losses/loops.
  - EVEN IF
    - a FRR recovery has been established around the failure or
    - the event is predictable (link manual shutdown)

- Bringing up a new link in the IGP can make you loose packets !!!

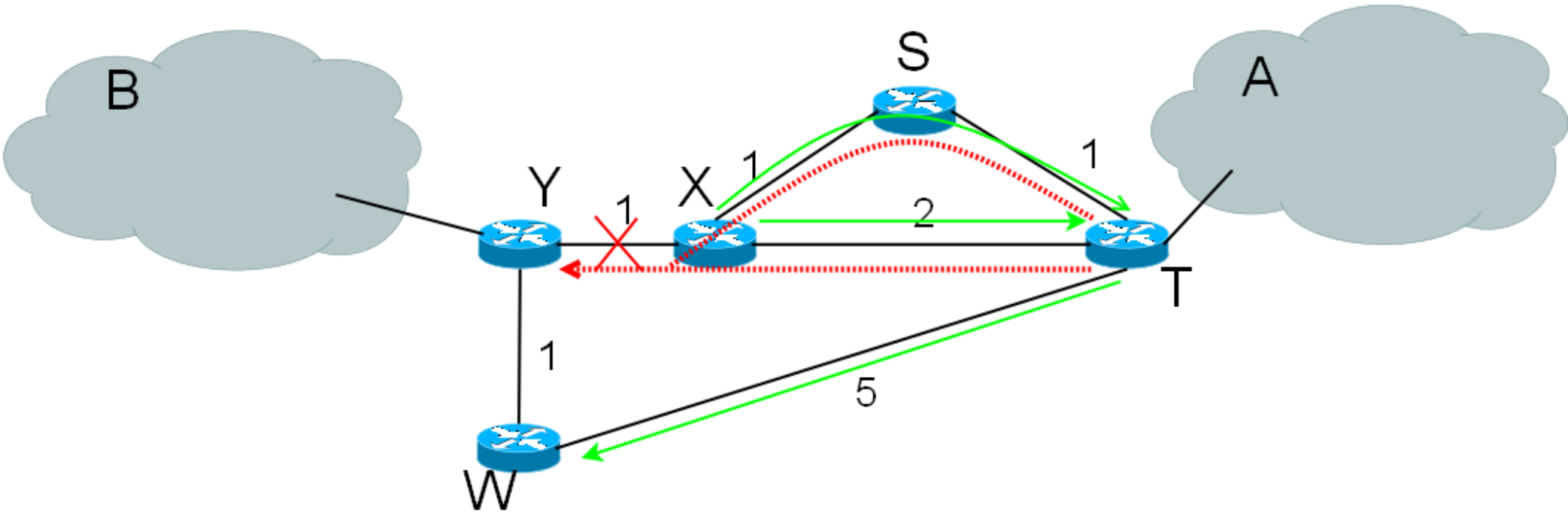
Loosing packets in the IGP

- Let's manually shut X—Y down...
  ...or set its metric to MAX_METRIC-1
- The closer to the failure, the sooner the FIB update (in general)...



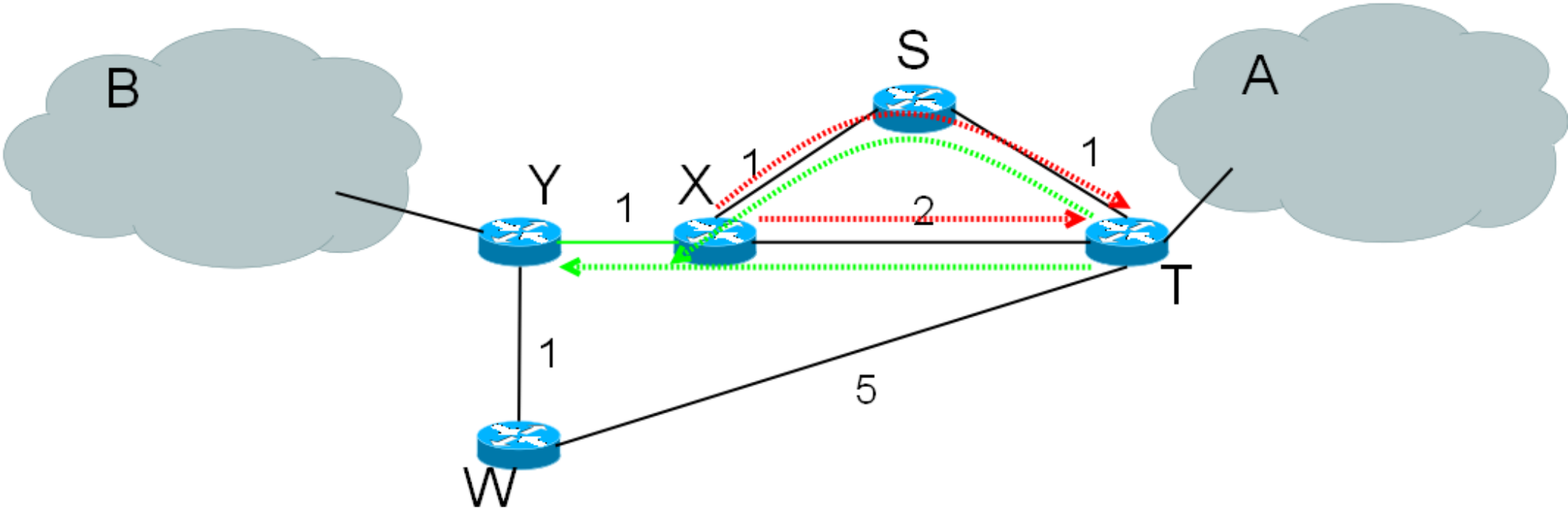
Loosing packets in the IGP(2)

- Potential forwarding loops along
  - X—S, S—T, X—T
  - for all destination prefixes lying in cloud B...



Loosing packets in the IGP(3)

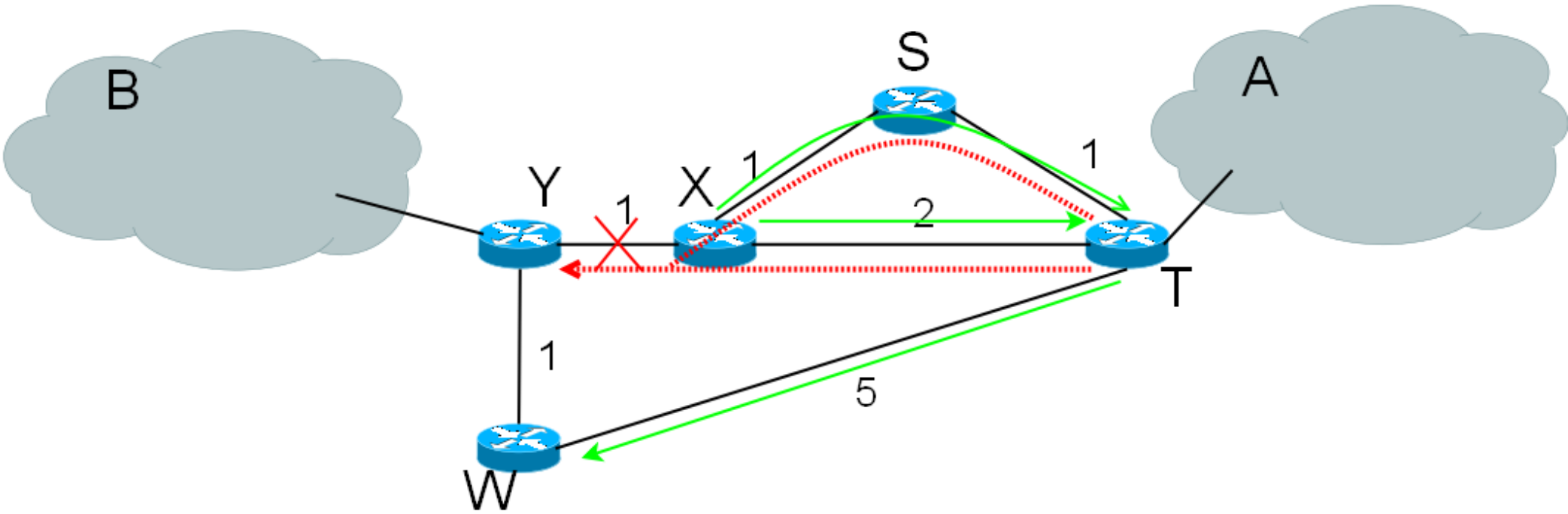
- The link is brought back up... same potential loops



- Though, less packets are lost (in general)

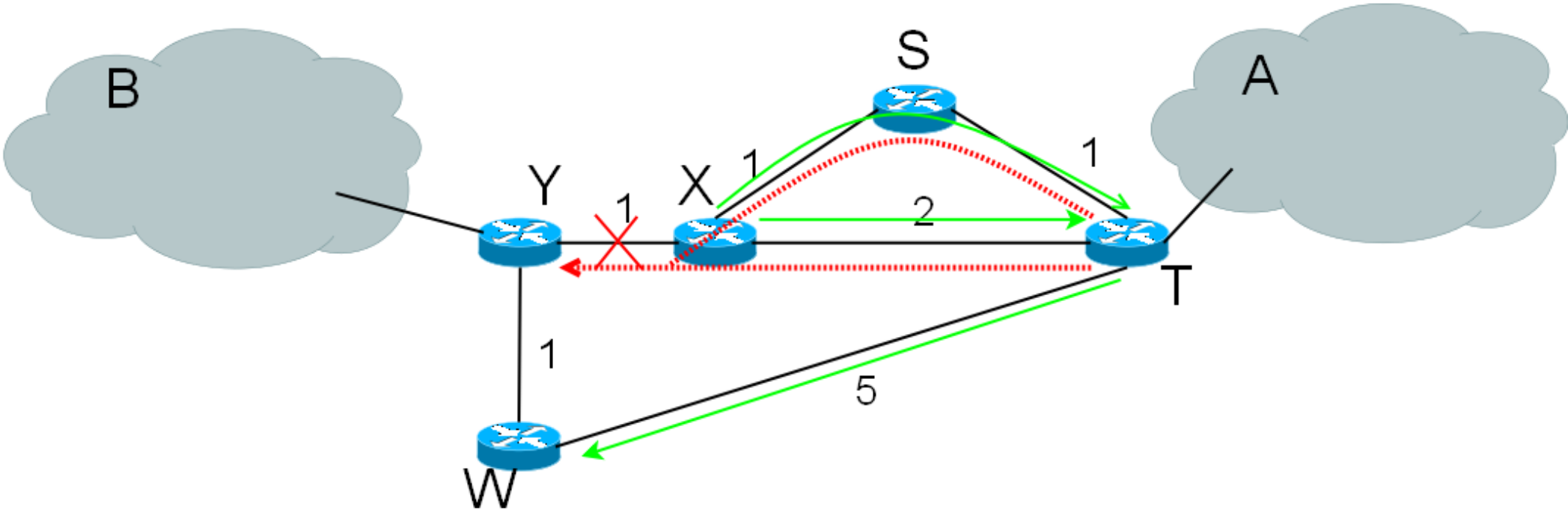
Solutions

- Ensure forwarding consistency during convergence process
  - **PLSN, OFIB**, Metric Increments, FIR



PLSN

- Very simple idea...
- For each FIB entry to be updated
  - Update FIB to safe neighbors with no delay
  - Wait for a (fixed) while if neighbor is not safe
  - May temporarily reroute to a non primary safe neighbor



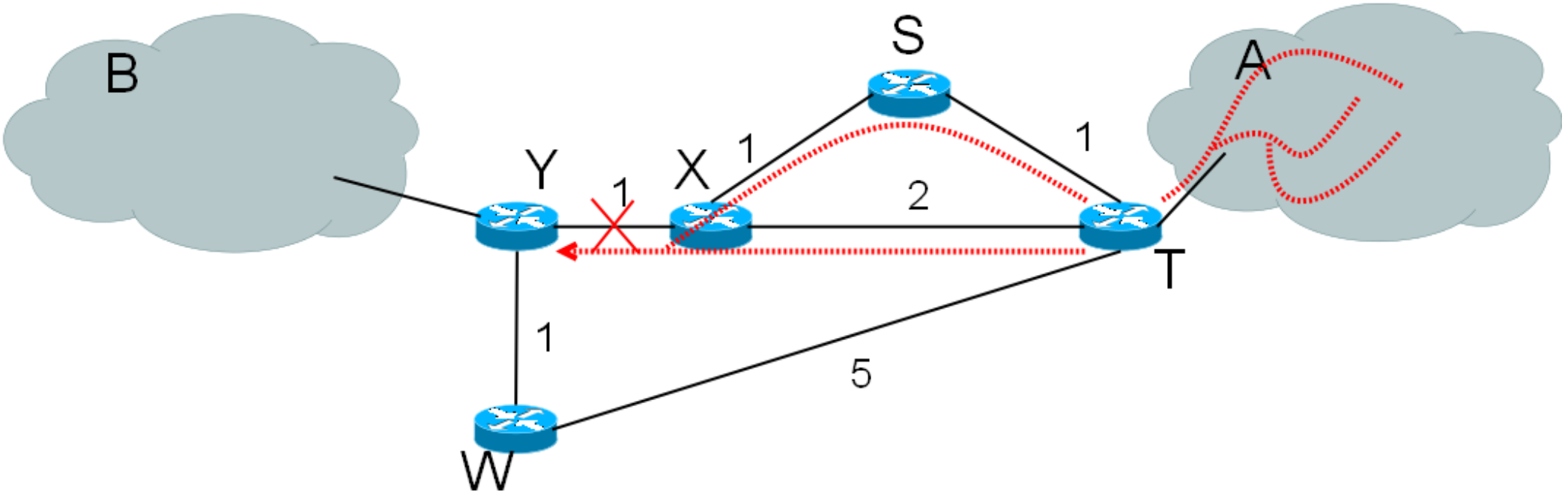
http://www1.tools.ietf.org/html/draft-ietf-rtgwg-microloop-analysis-01

PLSN

Coverage depends on topology

oFIB

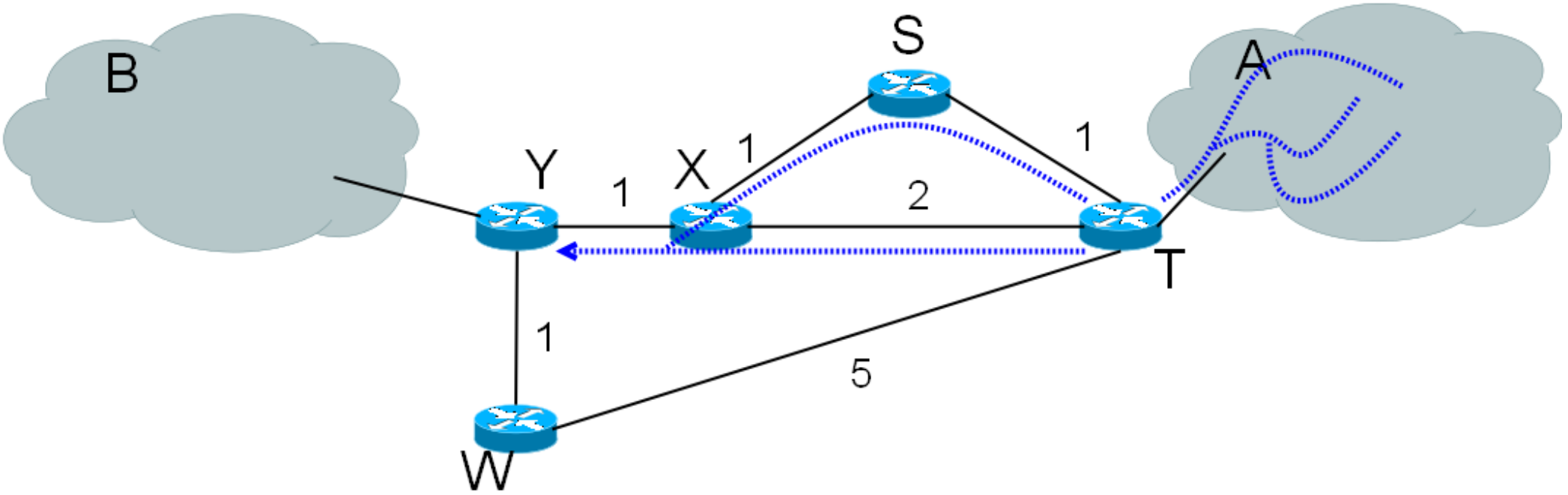
- Router R updates its FIB after the routers that use R to reach a failing link.
  - based on rSPT(X→Y)
  - Using a timer and completion messages



  - http://www.ietf.org/internet-drafts/draft-francois-ordered-fib-02.txt

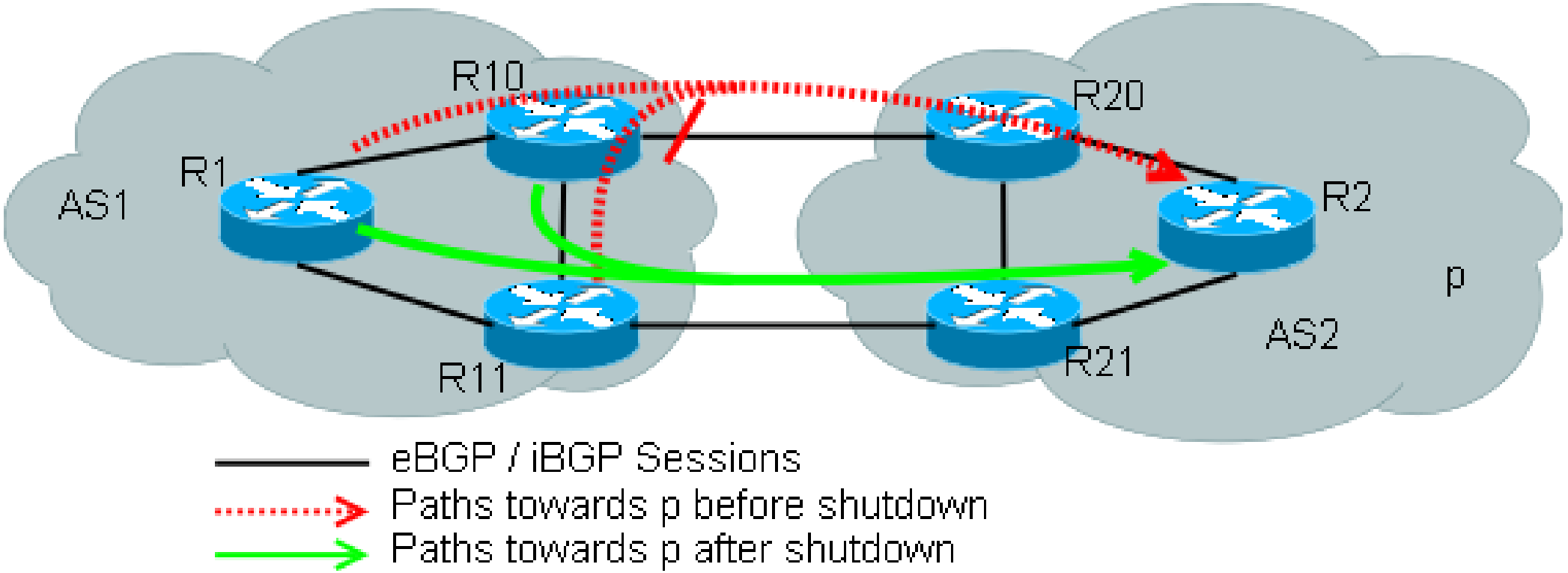
oFIB(2)

- Router R updates its FIB after the routers that R will use to reach an upcoming link.
  - Based on renewed SPT of R
  - Using a timer and completion messages



Maintenance of eBGP peering links

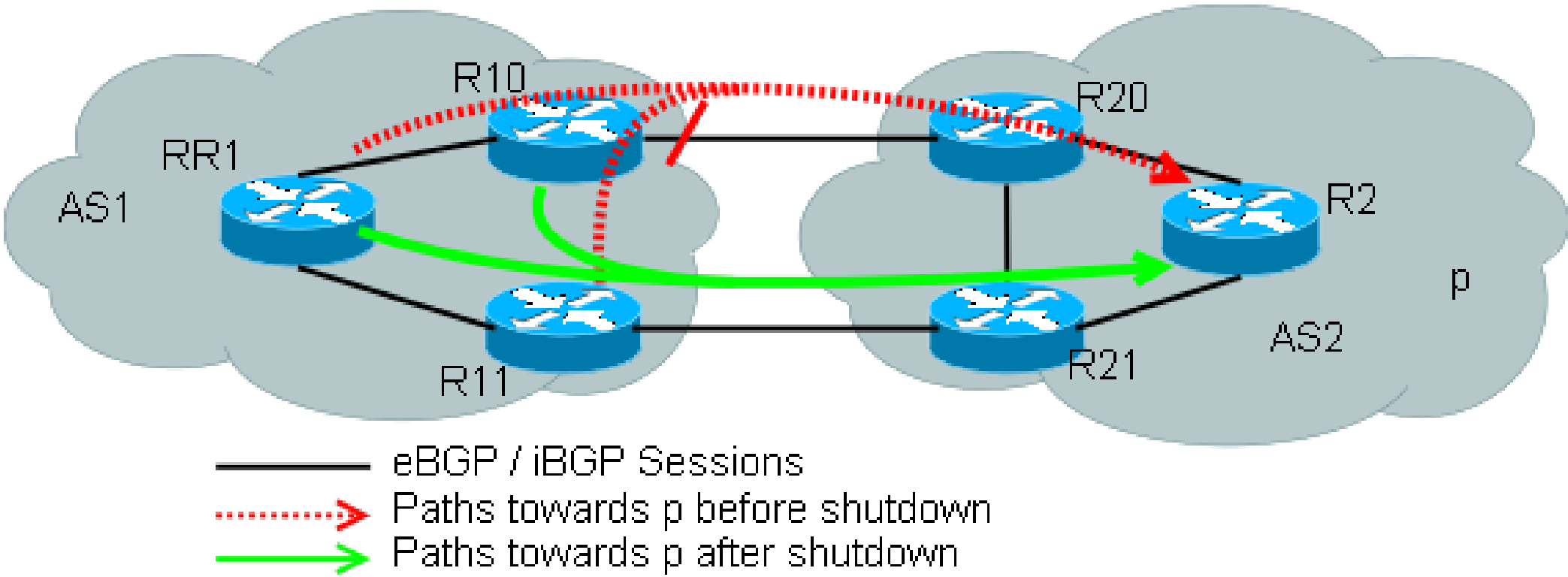
- BGP speakers can lack of information on alternate paths



- R1/R10 do not know about path to p via R11—R21

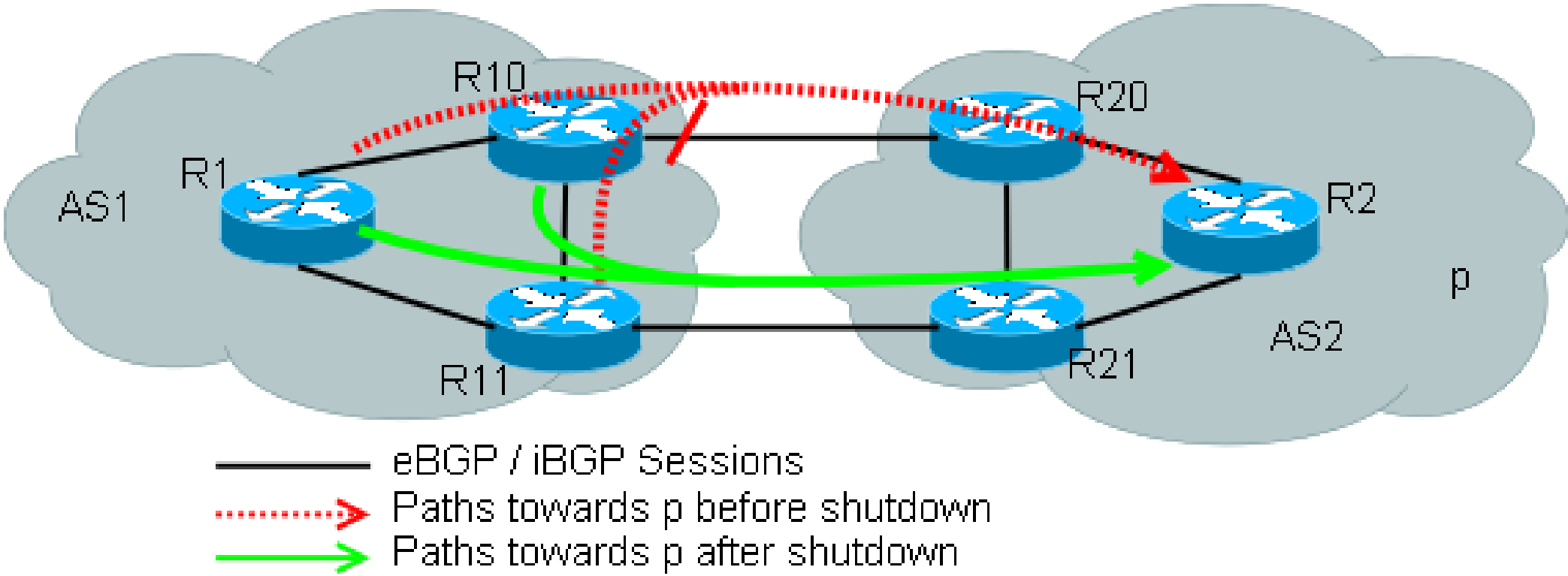
Maintenance of eBGP peering links : RR

- Route Reflectors worsen the problem...
  - Memory load reduction at the cost of less diversity.
  - If all your RR use the same nexthop to reach p...



Maintenance of eBGP peering links

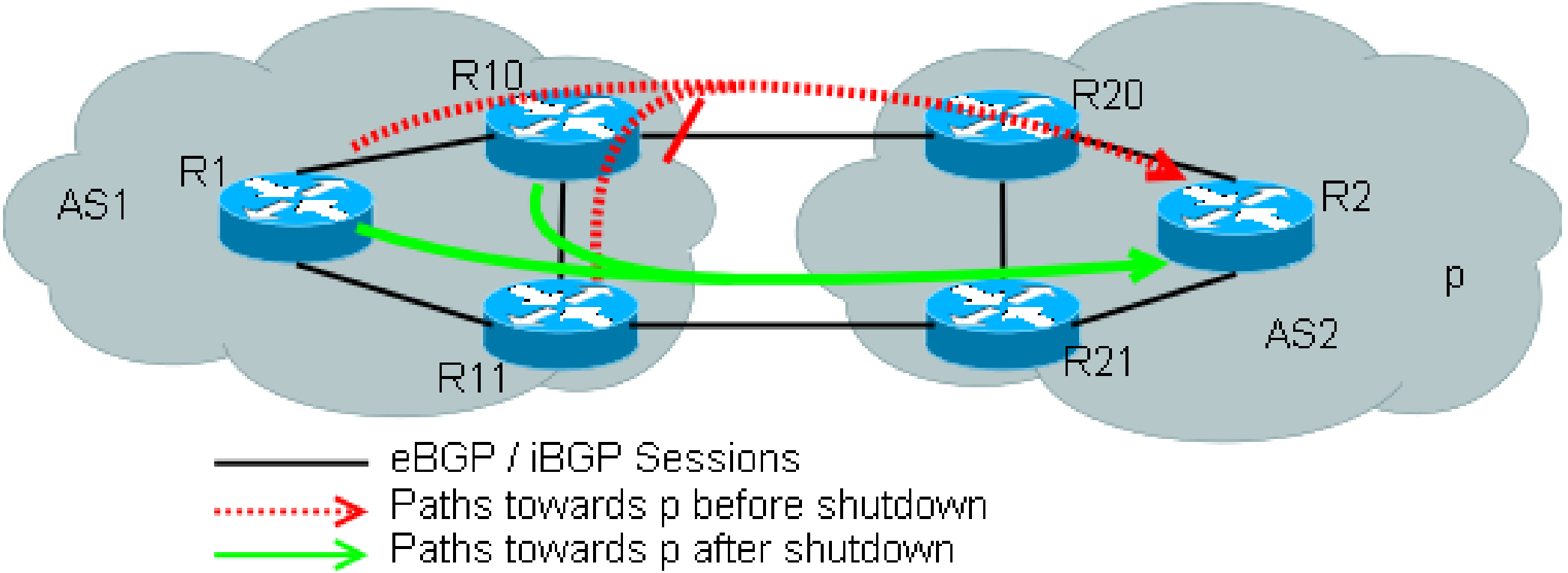
- Sending path withdrawal first is not sufficient...
- R1 has no alternate path towards p, drops packets...
  - R11 must process the withdraw, run DP, propagate its own path
  - Before the others start knowing about this path
  - Number of affected prefixes can be large...



Second attempt

- Propagate Local-Pref Update to 0 first
  - The outdated path will survive...
  - And be replaced when alternate paths are propagated
- R11 receives a LP update to 0
  - prefers its own route,
  - propagates it to R1 and R10...

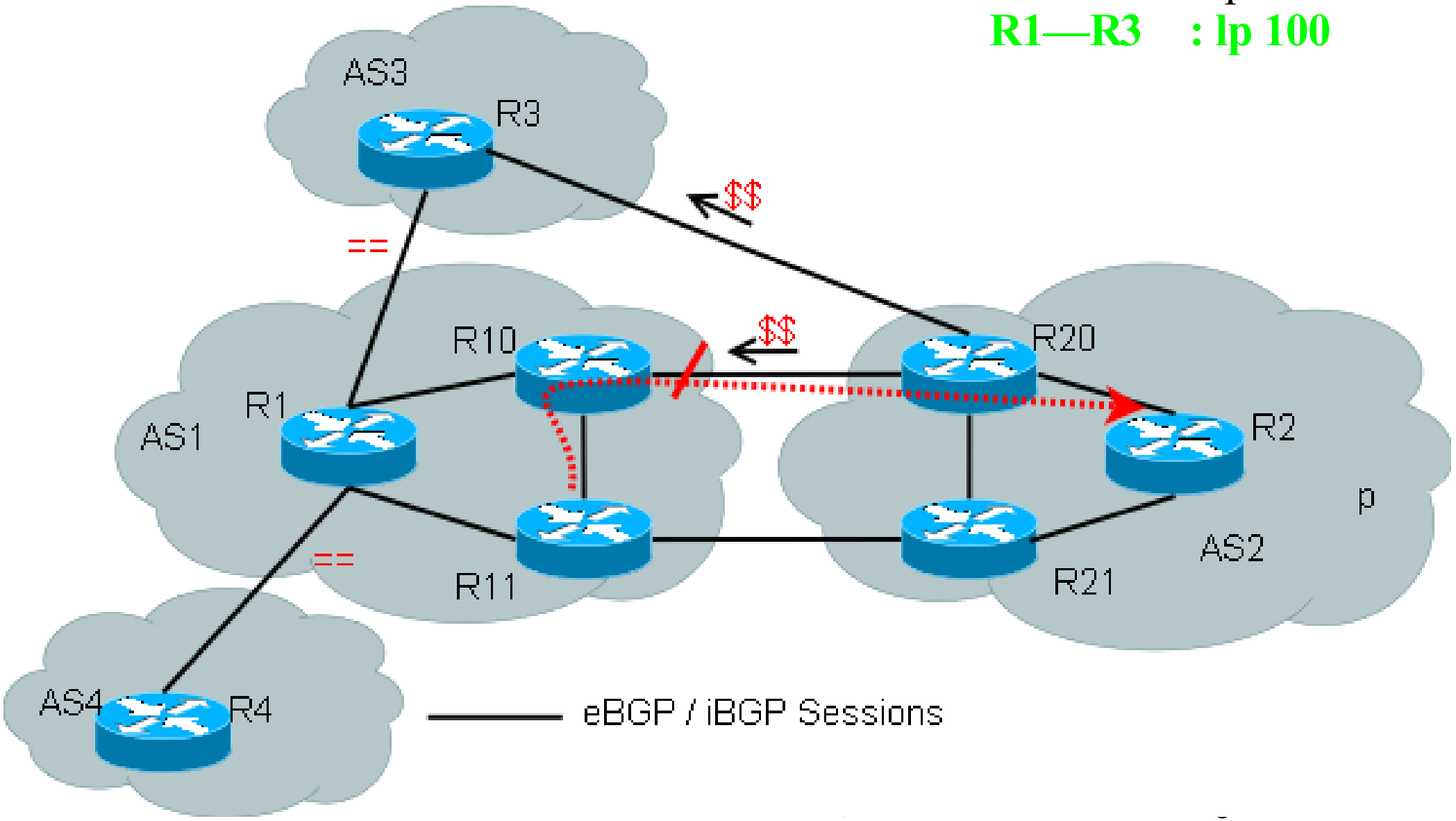
R11 Rib-IN : R10—R20 : lp ~~160~~ **0**
R11—R21: lp 150



Second attempt : we still loose packets !

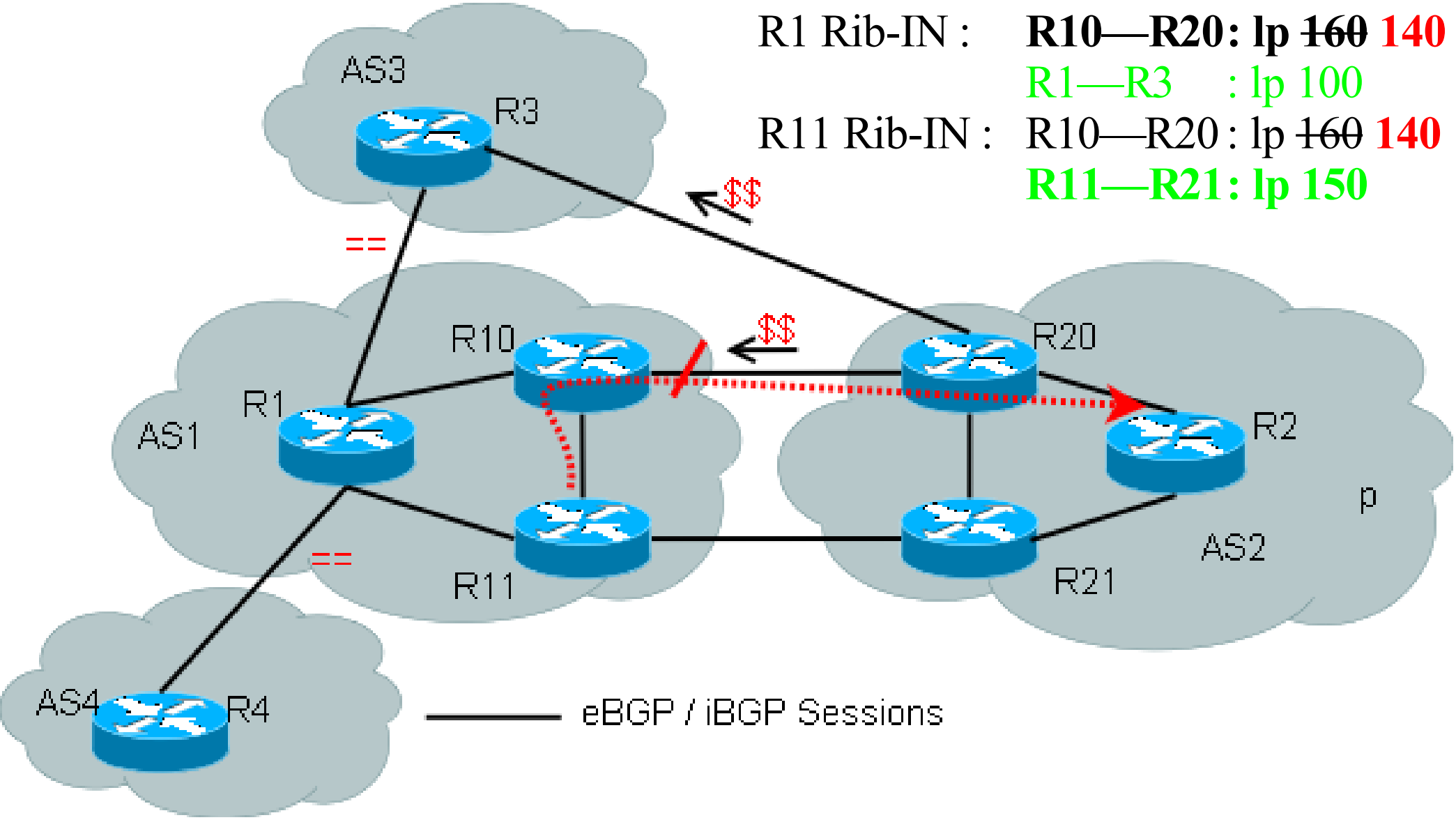
- Propagate Local-Pref Update to 0 first

R1 Rib-IN : R10—R20 : lp ~~160~~ **0**
R1—R3 : lp 100



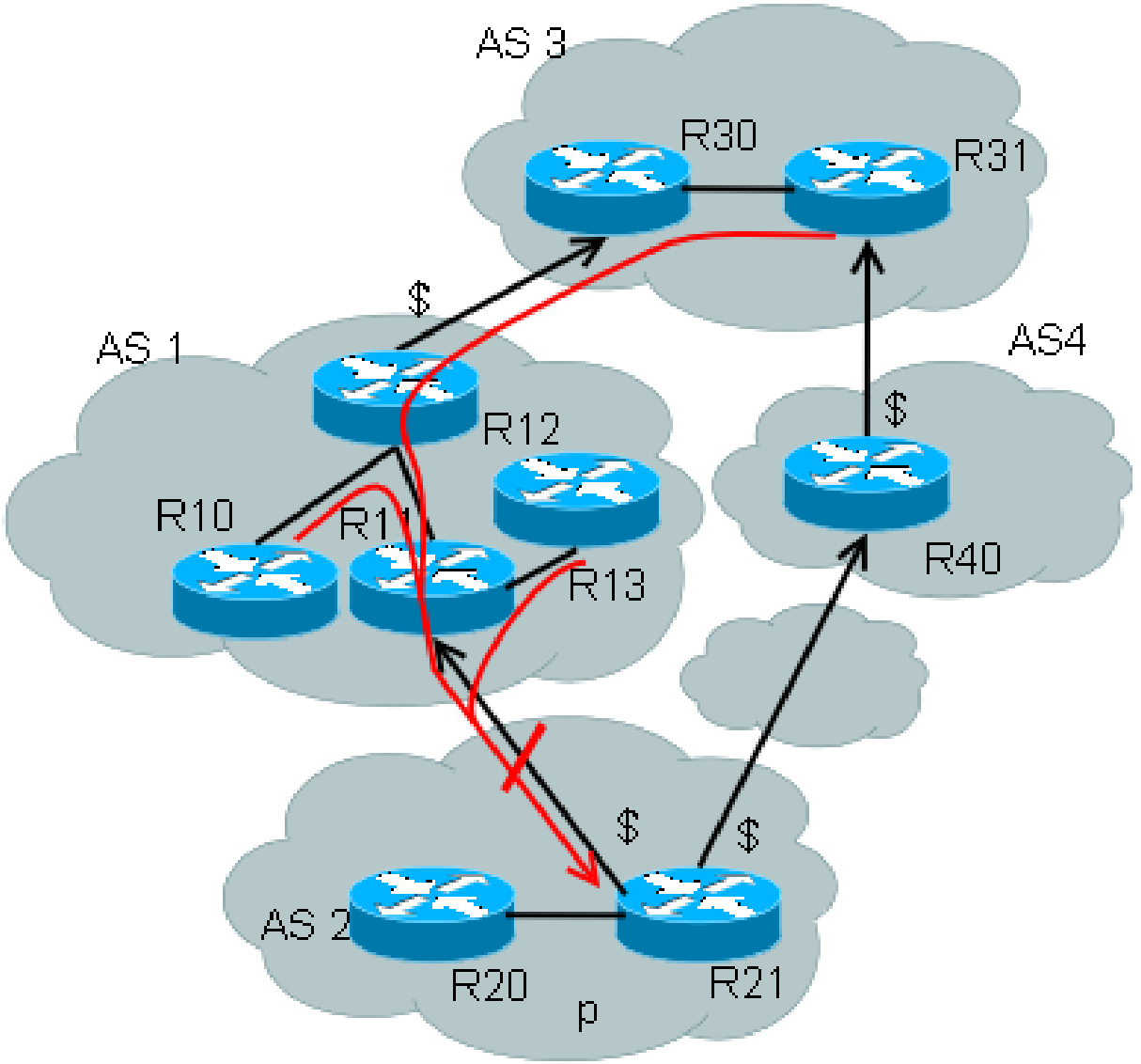
Third attempt

- LP range : [50-75] [100-125] [150-175]
  - Propagate Local-Pref Update to 140 first, then 0 (if necessary) !



Third attempt : limitations

- Solution works okay when AS-local recovery is doable
- What if re-convergence requires neighboring AS participation ?



eBGP shutdown : other solutions

- Avoid lack of alternate paths in the routers
  - BGP “External Best”
  - Propagate multiple paths for each prefix on iBGP sessions (Walton et al.)
  - Tradeoff memory load/path diversity
- Implementations not there yet
- LP tuning technique can be applied now !

Questions / Comments ?