UCL

INFOCOM'06 routing panel

Fast convergence

Olivier Bonaventure

Department of Computing Science and Engineering
Université catholique de Louvain (UCL)
Place Sainte-Barbe, 2, B-1348, Louvain-la-Neuve (Belgium)

URL : http://www.info.ucl.ac.be/people/OBO

INGI
Département d'ingénierie informatique

What should be the goal of routing protocols ?

- Main goals
  - Discover network and reachable destinations
  - Allow routers to build correct forwarding tables to forward ***ALL*** packets towards reachable destinations ...
    even if the network topology frequently changes

- Secondary goals
  - Quality of Service routing
  - Security
  - Multicast
  - Traffic engineering

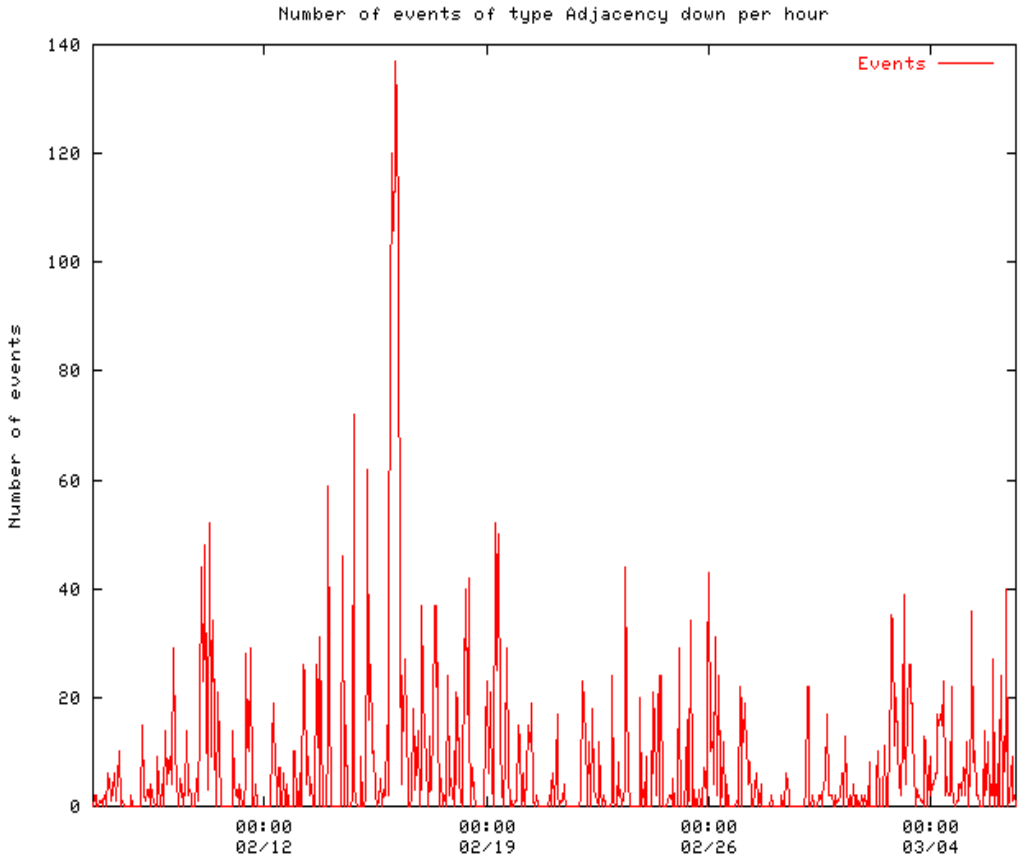
IS-IS in a tier-1 ISP

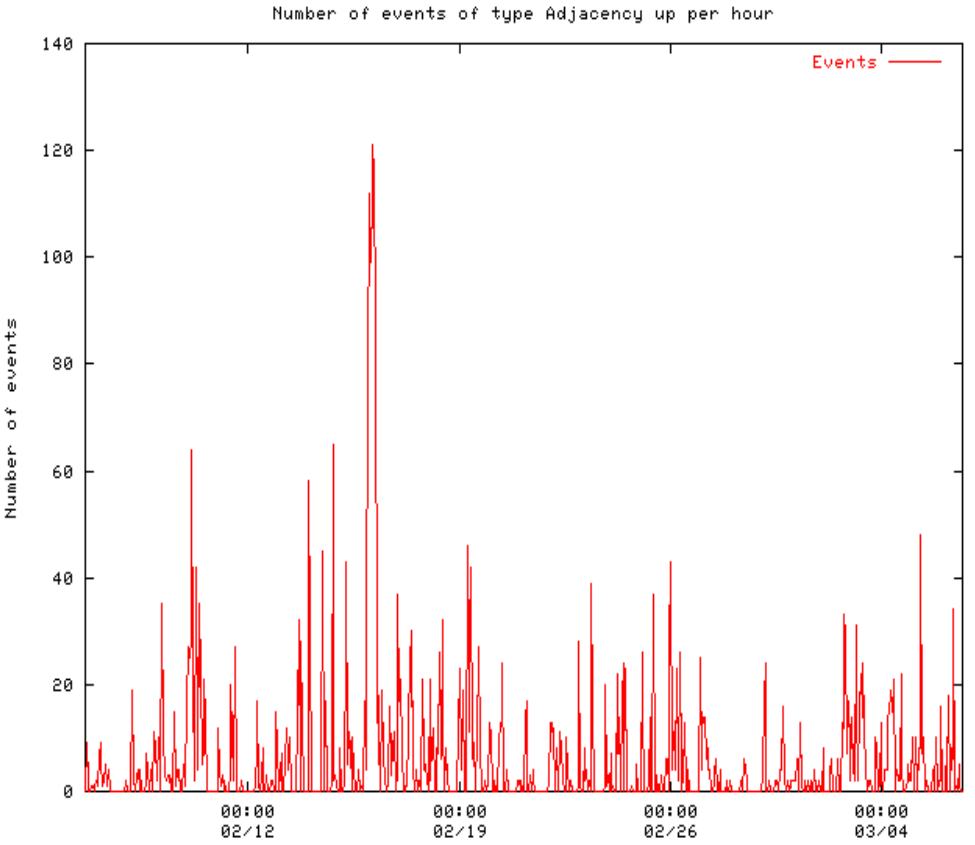
- The Network
  - Large tier-1 transit ISP
  - 400 routers in studied ISIS area
  - IS-IS wide metrics and TE extensions are used in the network

- The trace
  - IS-IS adjacency between a PC running a modified tcpdump and a router
  - all IS-IS packets logged in libpcap format during one month
    - analysed with scripts and `lisis`
    - http://totem.info.ucl.ac.be/tools.html

The adjacency changes per hour

- 5276 adjacency down LSPs (left)
- 4487 adjacency up LSPs (right)





- Maintenance operations and sudden failures

How fast can link state routing converge ?

- In the past
  - Routers used their CPU to forward packets and support link state routing
  - To protect CPU, routers waited *five* seconds after a topology change to update forwarding table

- Today
  - Faster convergence is possible
    - Sub-second convergence in large ISP networks
  - Key bottlenecks in large networks are
    - Link propagation delays
    - Time to update a prefix in forwarding table
      - 100 microsecond on Cisco 12k
    - Number of prefixes advertised inside by link-state routing protocol

How to avoid loosing packets when links fail?

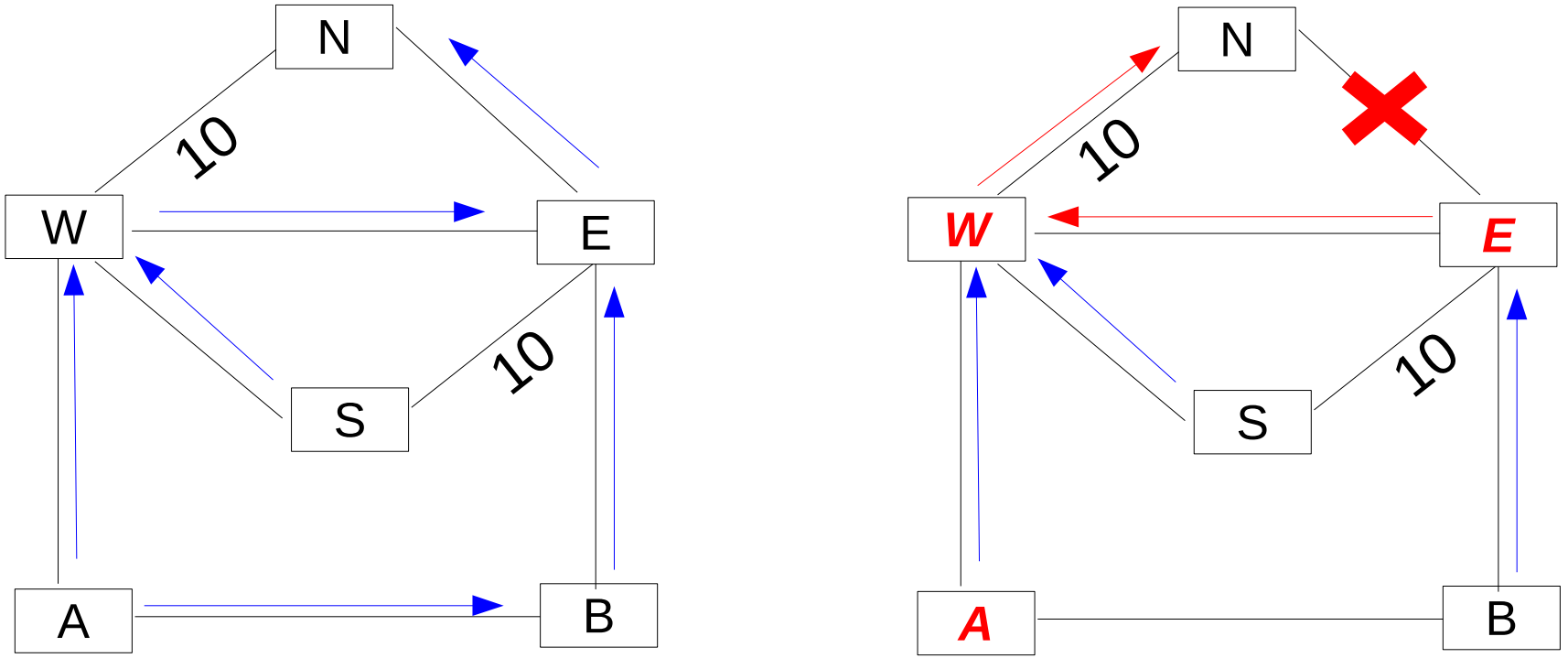
- First step
  - Quickly detect the failure
    - Physical layer aid for Packet over SONET
    - BFD protocol for other technologies
- Second step
  - Reroute the packets *at the router that detects the failure to an alternate router*
    - MPLS fast-reroute and bypass tunnels
    - IP-based techniques (loop-free alternates, tunnels, not-via addresses, ...)
- Is it sufficient ?
  - Unfortunately not, transient loops can occur during the update of the forwarding table

How to avoid loosing packets when links fail ? (2)

- First step
  - Quickly detect the failure
    - Physical layer aid for Packet over SONET
    - BFD protocol for other technologies
- Second step
  - Reroute the packets *at the router that detects the failure to an alternate router*
    - MPLS fast-reroute and bypass tunnels
    - IP-based techniques (loop-free alternate, tunnels, ...)
- Third step
  - Orderly update the forwarding tables of all affected routers to avoid all transient loops

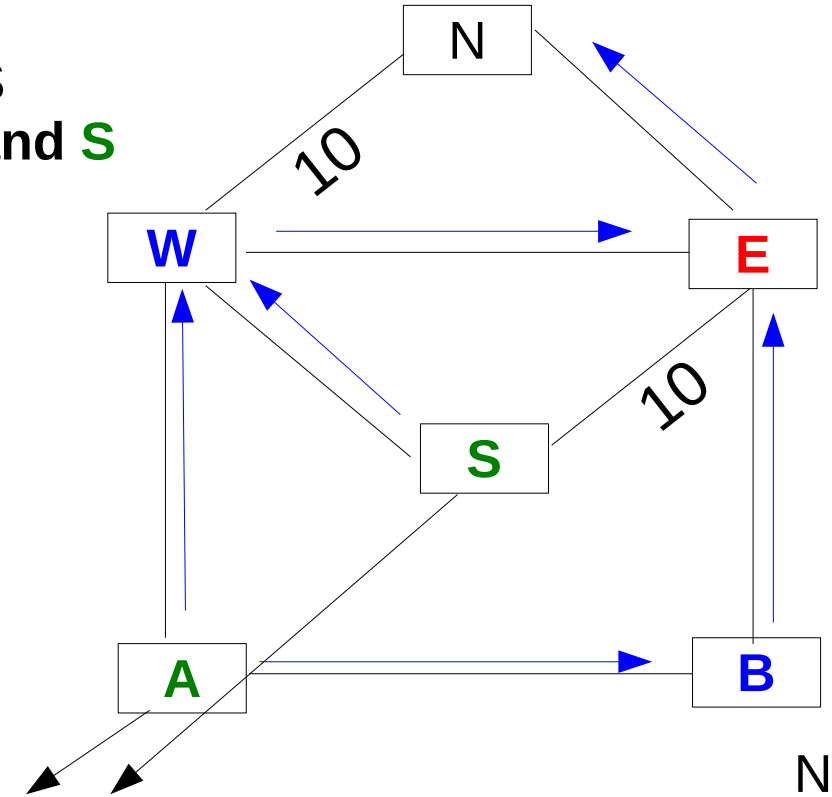
Ordering of forwarding table updates

- Principle
  - When a link fails, routers far away from the failure must update their FIB before routers close to the link failure



Ordering of forwarding table updates (2)

Node W :
•1 hop from A and S
• **Updates after A and S**



Node E :
•2 hops from A
•**Updates after W and B**

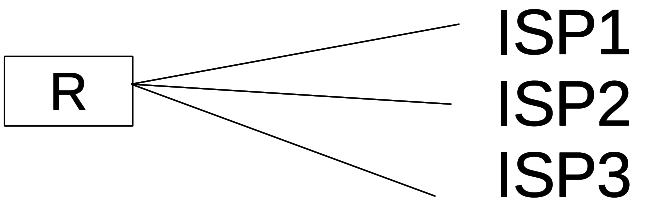
Farthest nodes from failure, **first** to update

Node B :
•1 hop from A
• **Updates after A**

Paths towards N

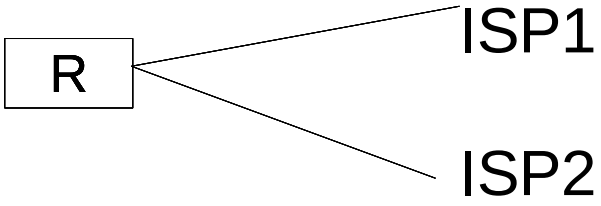
Fast convergence and interdomain routing

- Current BGP convergence times in global Internet
  - Craig Labovitz's measurements
    - Several tens of seconds or more



R starts to advertise p to all
R stops to advertise p to all

  - Happy packets from dual-homed beacon presented at PAM
    - BGP convergence time
      - Up to a few 100s seconds
    - Packet convergence time
      - A few tens of seconds



R advertises p via ISP1 and ISP2
R only advertises p via ISP2
...

Can we achieve sub-second interdomain routing convergence ?

- Is a three step approach possible ?
  - First step
    - Quickly detect the failure
    - **Possible**, same techniques as for link-state routing
  - Second step
    - Reroute the packets *at the router that detects the failure to a loop-free alternate router*
    - **Possible**, but not yet implemented
  - Third step
    - Orderly update the forwarding tables of all affected routers to avoid all transient loops
    - **More difficult**

Can we achieve sub-second interdomain routing convergence ? (2)

- Issues for ordered updates of interdomain forwarding tables
  - Routers do not know entire network topology
    - BGP is a path vector protocol
  - Some routers do not know an alternate path to reach failed destination
    - Route reflectors, non-preferred routes
  - Entire AS's may not know an alternate path