

Improvements to IP routing convergence

Pierre Francois
Virginie van den Schrieck
Olivier Bonaventure

IGP Fast convergence

- Pushing the IGP to the limits
 - Design
 - Configuration
- IP Fast Reroute

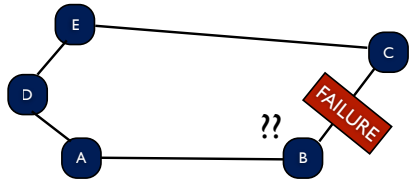
Agenda

- IGP
 - IGP Fast Convergence / Fast Reroute
 - Hitless maintenance operations
- BGP
 - Observation on path diversity in transit ISPs
 - Hitless maintenance operations

Components of the IGP convergence

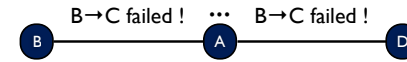
- Failure detection
- Link-State Packet generation / propagation
- Processing of the LSP, Paths recomputation
- FIB updates

Failure detection



BFD, L2 alarms,..
Fast detection, no control plane stress

Link-State generation/propagation



- Generation
 - No more fixed “wait” time 5s, 5s, 5s, ...
 - Exponential back off 0ms, 500ms, 1s
- Propagation
 - “Fast-flood” vs. fixed Pacing timer
 - No more artificial delaying under normal operation

Paths recomputation time

- more and more negligible
 - iSPF
 - Full SPF is a few msec now...

FIB Updates

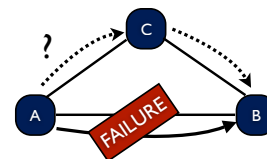
- bottleneck component of the convergence time
- 10k prefixes at 33μsec / prefix...

IGP Fast Convergence

- convergence time much below 1s now...
- but convergence time scaling factors exist
 - number of prefixes in the IGP
- **recovery** mechanisms can help

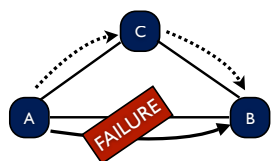
Loop-free Alternates

C is a loop free alternate of A, for the failure of A→B, if C would not forward the traffic sent along A→B back to A, when A deviates it to C



FIB design to allow direct deviation of traffic when A→B is flagged down

Loop-free alternates FIB design



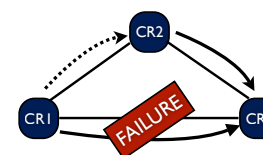
FIXED recovery time

prefix	oif
p/P	--B

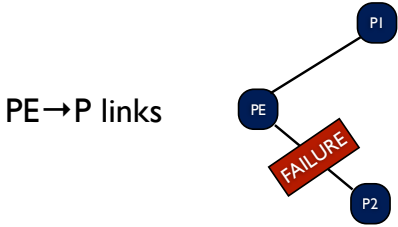
--B	DOWN	--C
--C	UP	--B

LFAs : where does it apply

core links of meshed cores



LFAs : where does it apply

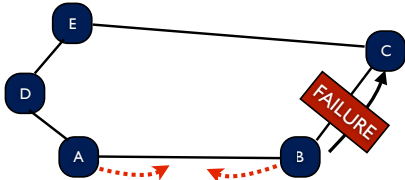


All the links of a non transit node can protect each other

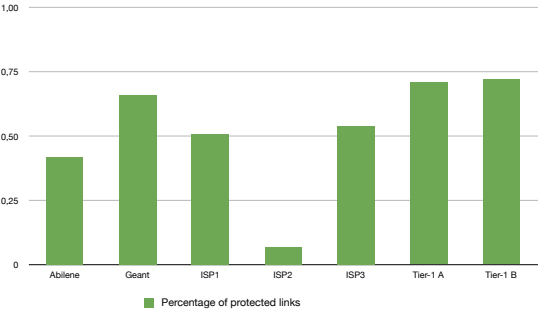
These links carry many prefixes

LFAs : where it does not apply

“Ring-ish” parts of topologies



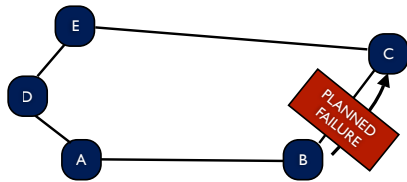
Links protected by LFAs



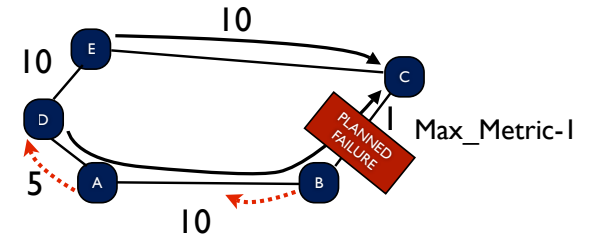
How to reach 100% coverage ?

- Using more complex techniques
- Tunnels, NotVia, MPLS FRR

IGP Hitless maintenance operations

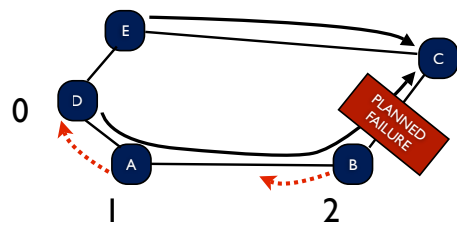


IGP Hitless maintenance operations



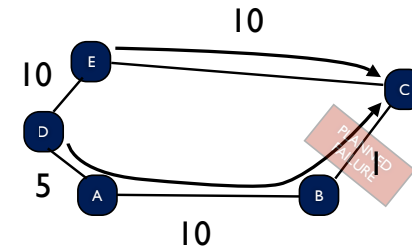
ordered FIB updates

Modifications to OSPF / IS-IS to enforce a loop-free ordering of the FIB updates upon planned maintenance

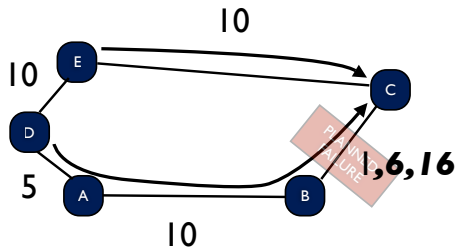


Metric-increments

Reconfiguration of OSPF / IS-IS to enforce a loop-free transition to post-convergence state



Metric-increments



Pierre François, Mike Shand and Olivier Bonaventure. Disruption-free topology reconfiguration in OSPF Networks. IEEE INFOCOM, Anchorage, USA, May 2007.

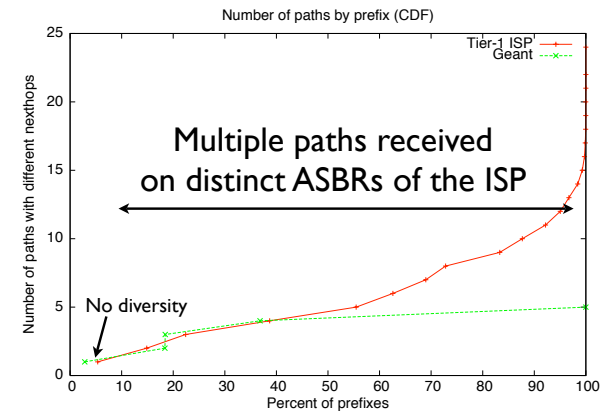
Conclusion

- Sub-second convergence is very conservative now
- Local FRR mechanisms make it close to Failure Detection time
- Hitless maintenance / reconfig is feasible

BGP

- Preliminary observation on path diversity
- BGP Graceful shutdown

Diversity is at the borders



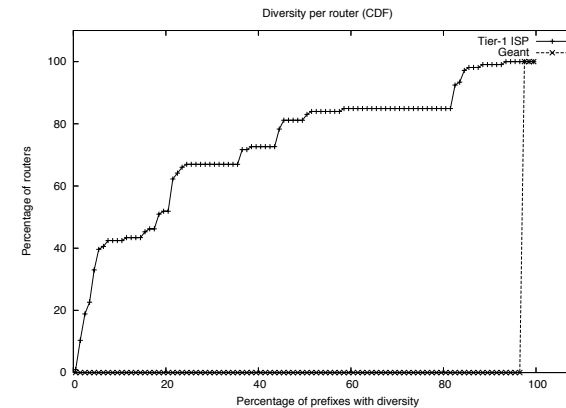
Diversity is at the borders...

Is it correctly propagated ?

- cBGP model of the ISPs
- iBGP configuration
- IGP configuration (links, link metrics)
- Injection of the paths in the cBGP model
- per router RIB-IN Analysis
- Cross check with “show ip bgp”

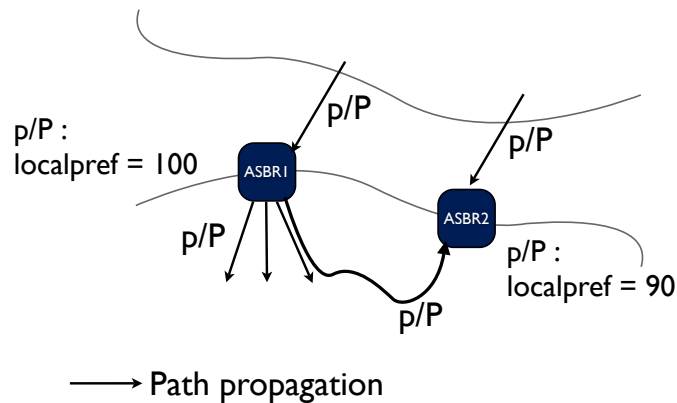
<http://cbgp.info.ucl.ac.be/>

Diversity does not propagate well



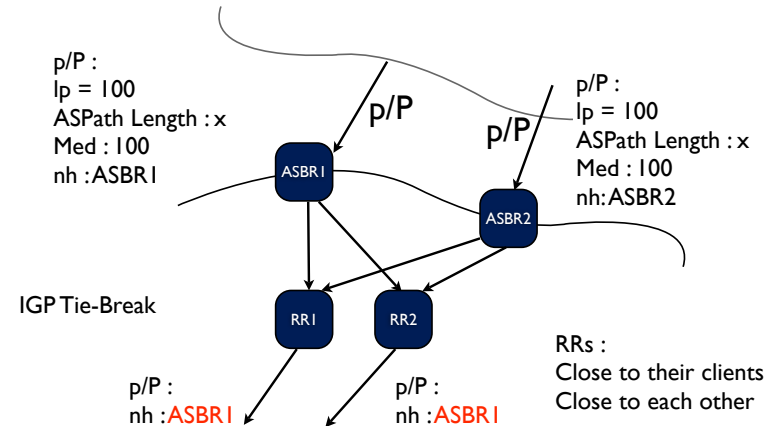
Diversity does not propagate well

Policies : One winner problem

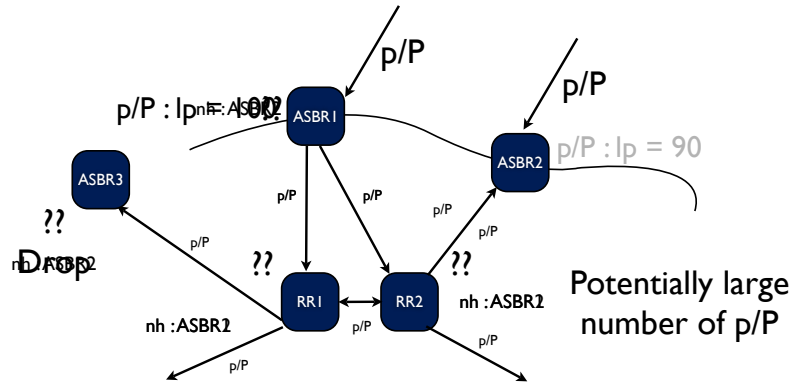


Diversity does not propagate well

Route Reflection



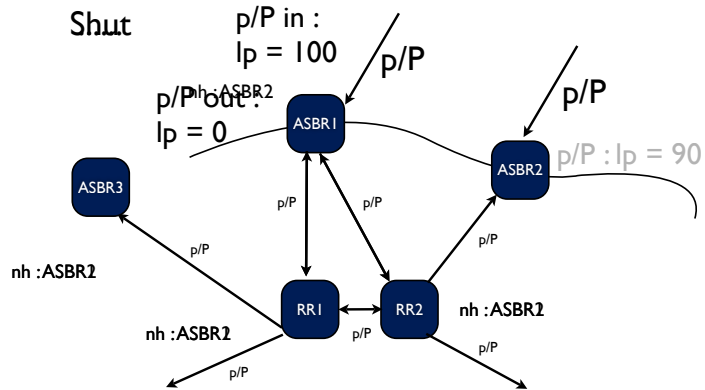
LoC during planned maintenance



G-Shut Procedures

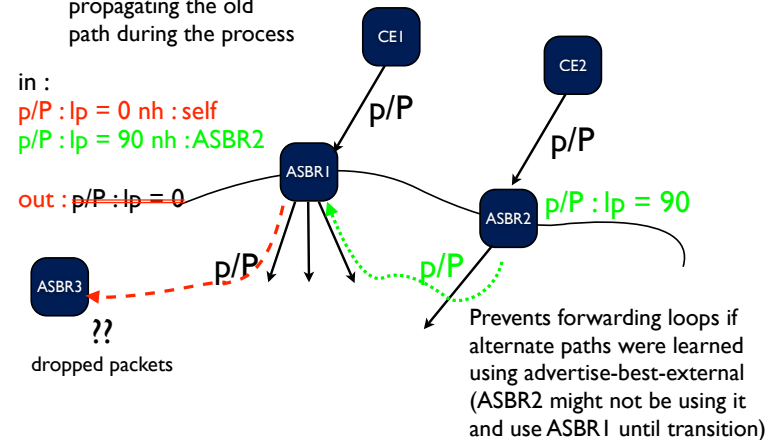
- Outbound traffic
 - Reconfiguration of iBGP **out**-filter at maintenance time
- Inbound traffic
 - Pick up the phone, or
 - Agreement on a G-Shut community value
 - Pre-configuration of iBGP **out**-filters related to the community
 - Reconfiguration of eBGP out-filter at maintenance time

Outbound traffic



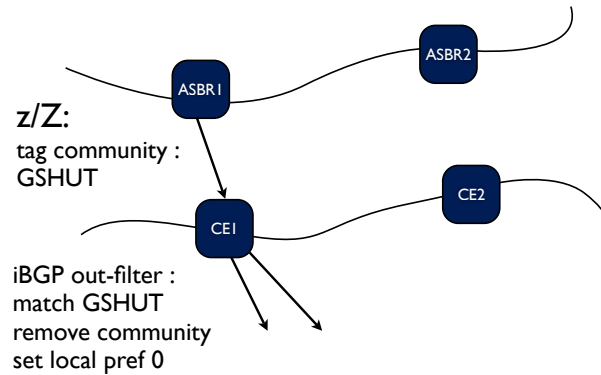
Why out-filters ?

ASBR1 should keep propagating the old path during the process

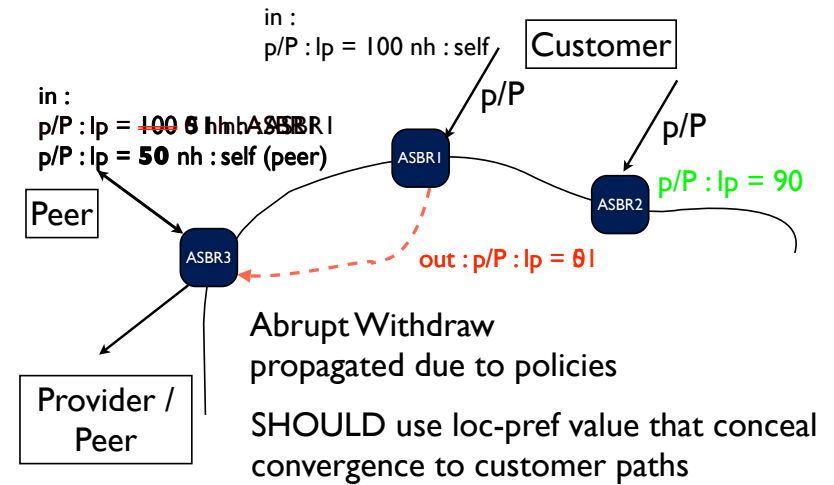


In-Bound traffic

Trigger out-bound g-shut at the other side of the peering link



BGP G-Shut convergence concealment



do this or solve the diversity problem

- BGP/MPLSVPN
 - Advertise best external
 - Use different RDs
 - cisco **"rd auto"**
 - juniper **"route-distinguisher-id"**
- Internet traffic
 - Advertise best external, establish more sessions
 - Wait for add-paths implementation / deployment
 - Consider the Internet as aVPN

Conclusion

- Policies and route reflection kills diversity
- solutions are on their way
- G-Shut procedures can be applied for the maintenance case

Thanks !