Fast convergence project

Nicolas DUBOIS	FTR&D/DAC	nicolas.dubois@francetelecom.com
Benoît FONDEVIOLE	FTR&D/DAC	benoit.fondeviole@francetelecom.com
Nicolas MICHEL	FTR&D/DAC	nicolas.michel@francetelecom.com

Agenda

- What is the network context ?
- How can we improve the scheduled maintenance ?
- How can we improve the IS-IS convergence ?
- Our roadmap

Network context

Services

- Migration of a lot of new services on IP networks:
  - Voice over IP, video over IP
  - VPN with SLA (*Service Level Agreement*)
- New applications are :
  - Sensitive in term of performance
  - Critical for the business

Equipments

- Many of them with IP functions :
  - Switch (ATM, GE, etc.)
  - BAS et NAS (mobility) …
- A lot of different vendors within France Telecom networks
  - Routers : Cisco, Juniper, Redback, Cosine, etc.

Networks

- 4 major Ases (3215, 5511, 25186, Equant IGN)
- With various topologies and services

Network context : Hot issues

Problem with L2TP tunnels:

- L2TP is used for INTERNET aggregation. The IP backbone is fully securized. But during link or node failure and scheduled work a lot of L2TP sessions are reset. We need less than 5s as L2TP timer is around 8s.

Problem with MPLS/VPN:

- Traffic from VPN customers is very sensitive in term of convergence because :
  - Video applications over INTRANET are growing fast
  - There are a lot of real time applications : cartography apps, financial apps, etc..
  - SNA interconnections with thousands of users
  - Deployment of "thin clients" (e.g. Citrix)
- Service Convergence should be under the second

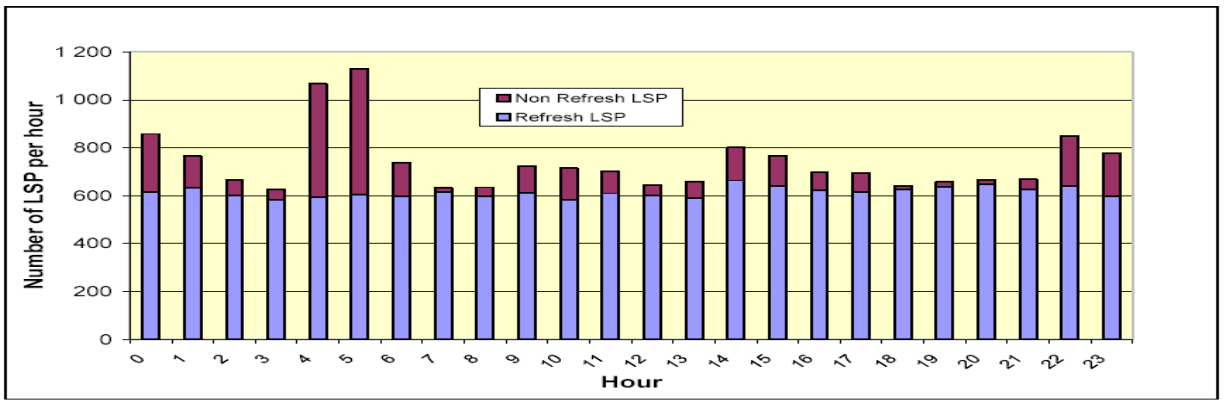
Two improvement areas :

- Improve the **scheduled maintenance** process so that :
  - L2TP session get rerouted before the maintenance operations
  - MPLS-VPN stability is enhanced (no traffic loss, no BGP session reset)
- Improve the IS-IS **convergence** in the IP backbone (PE-P and P-P) so that :
  - Unexpected traffic disruption become as short as possible

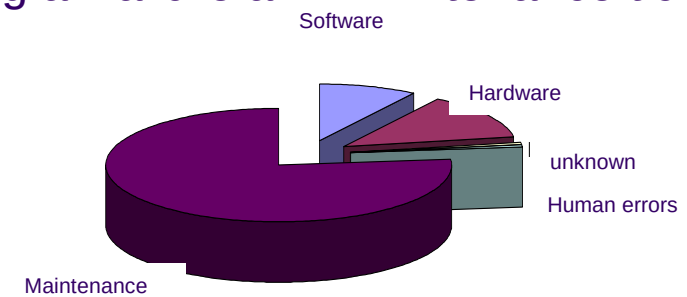
Improve the scheduled work (1/2)

- Taking into account the scheduled work is a priority. Most convergence cases are related to scheduled work.

- Example 1: Study comparing IS-IS activities on maintenance hours compared to daytime. IS-IS changes are much more frequent during maintenance hours (study over 1 month)



- Example 2: Study analyzing all failure and maintenance tickets and the related root causes:



Improve the scheduled Maintenance (2/2)

- We can anticipate the rerouting for a scheduled maintenance. That is of course not possible in case of an unexpected failure event.

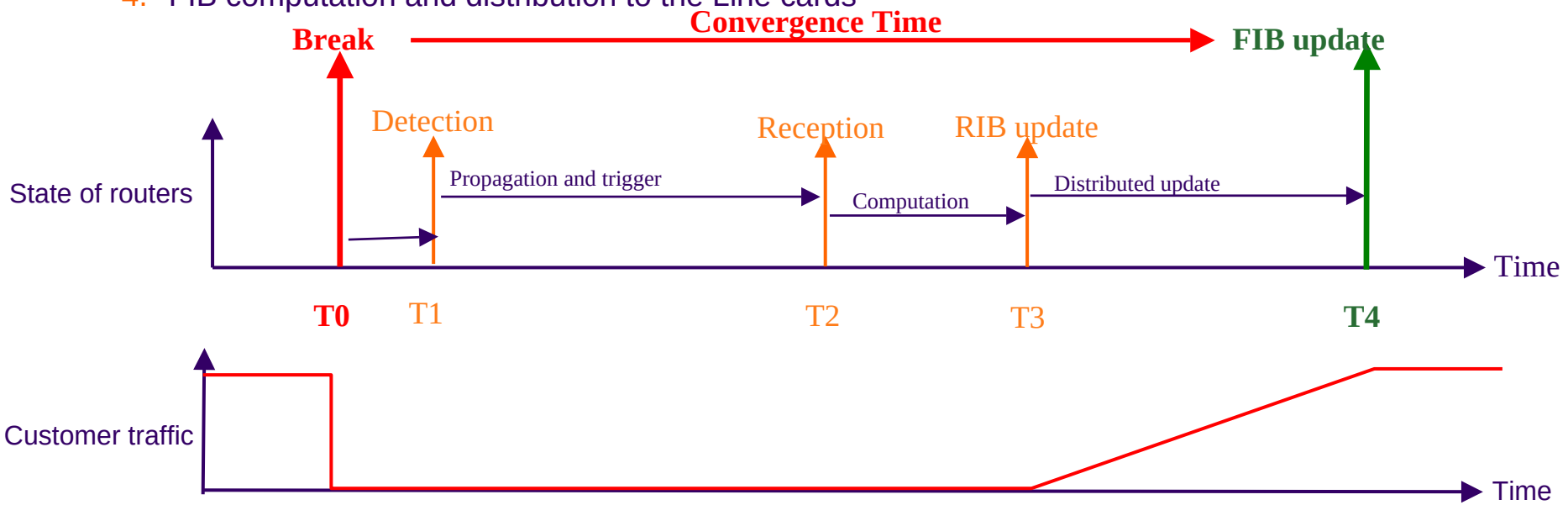
- Exemple of an improved engineering rule in case of a router reload (software upgrade).
  - Before the reload :
    - Step 1 : Set overload bit in the IS-IS configuration
    - Step 2 : Depending on the router, modify some metrics (next-hop or loopback, etc.)
    - Step 3 : Wait and verify that expected rerouting of the traffic as occured.
    - Step 4 : Reload the router
  - After the reload :
    - Step 5 : Don't forget to normalize the router,
    - Step 6 : Verify that nominal routing is again working
  - Not always efficient, can become very complex.

- **We need a safe, simple, efficient solution to allow scheduled work without loss of packets in order to reduce operational costs and disruptions**
  - Maybe dedicated reload feature combined with Graceful restart and Non Stop Forwarding ?
  - Maybe other protocol changes in ISIS and BGP ?

Improve the convergence time

Analysis of the convergence time:

1. Detection Time
2. IS-IS Flooding and SPF triggering
3. SPF Computation and RIB update
4. FIB computation and distribution to the Line cards

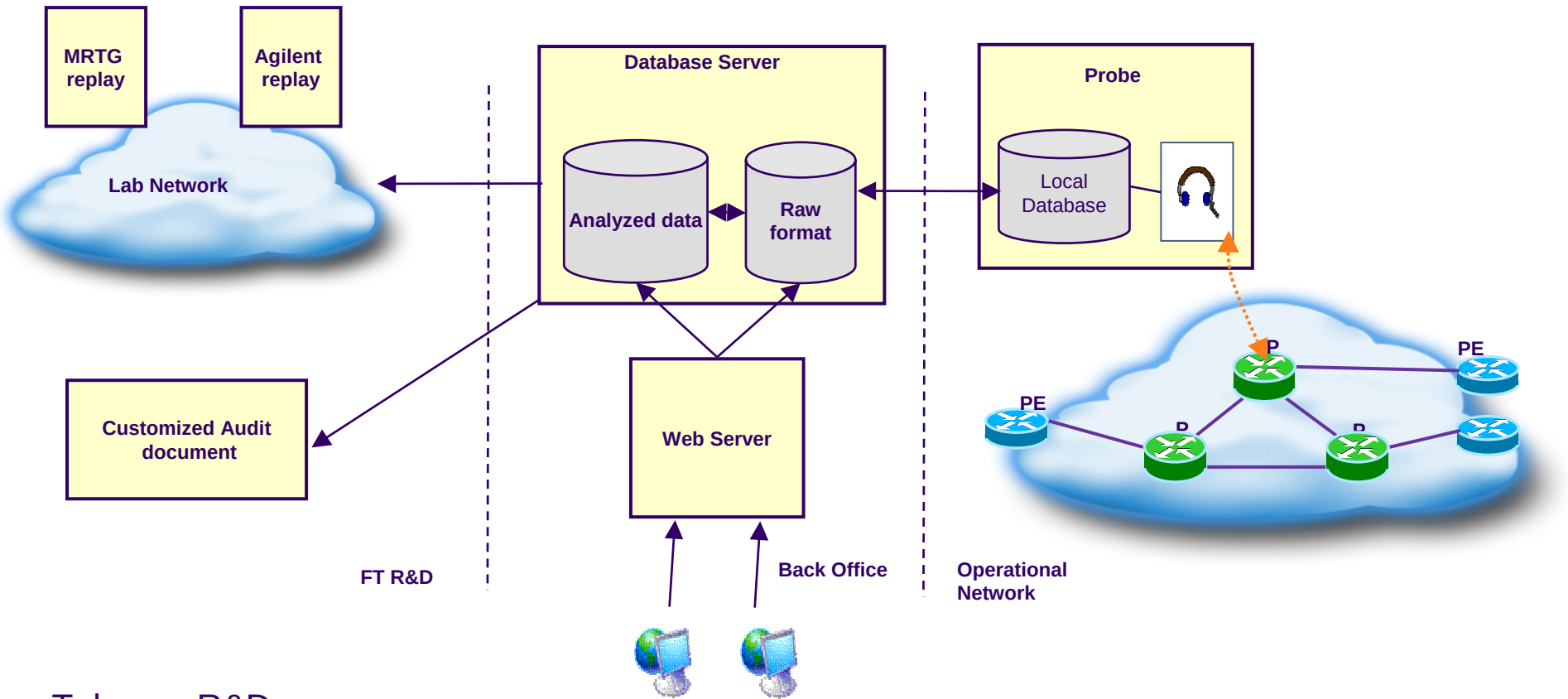


Method to improve the convergence time:

1. Understand the IS-IS activity
2. Enhance IS-IS protocol behavior
3. Improve our test tools
4. Verify and measure on the real network the improvement

Principle for monitoring IS-IS

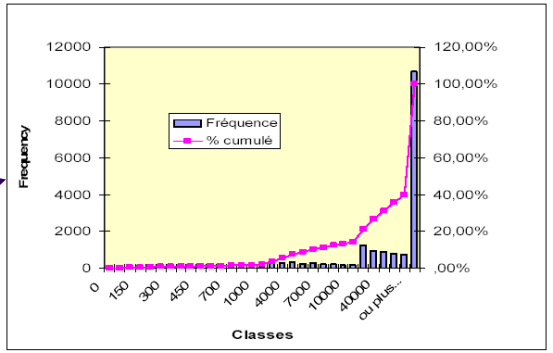
- **Principle: Snoop an IS-IS session between two IS-IS Routers**
  - The probe receives and stores all LSP flooded in the network
  - The database server stores and analyses the LSP
  - The web server displays statistics

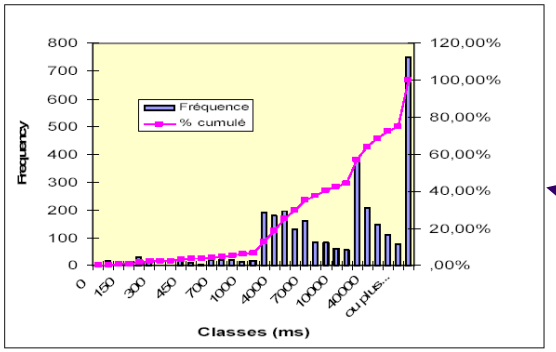


Understand the network behavior

- **Audit network behavior to prove IS-IS optimization feasability with Sceptre:**
  - Inter-LSP arrival

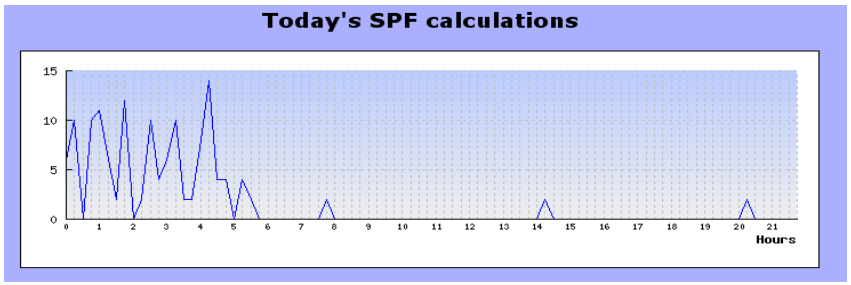
Repartition of the interval between LSPs (ms)

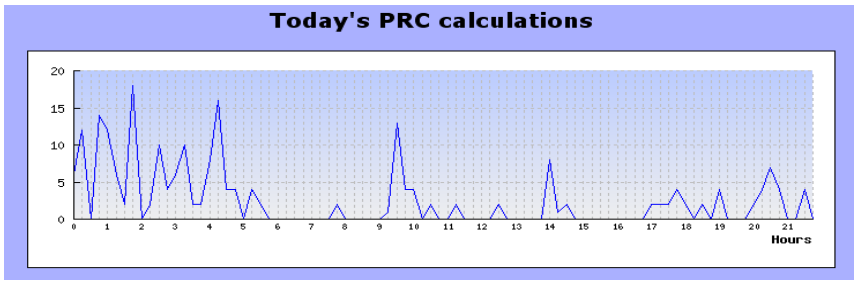




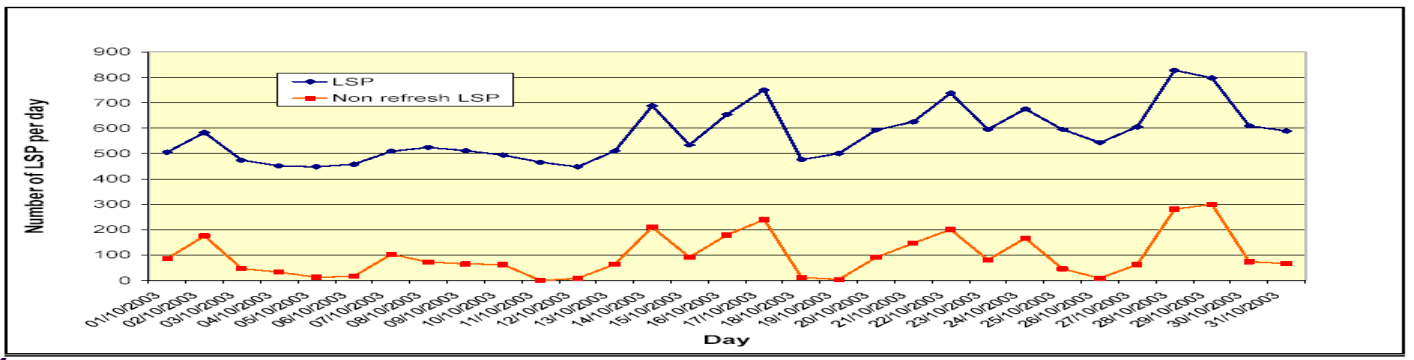
Repartition of the interval between non-refresh LSPs (ms)

  - Number of SPF and PRC computations per day :





  - General evolution of IS-IS activity (here over one month) :



Today possible IS-IS Optimization

Optimizing IS-IS reactions to failures

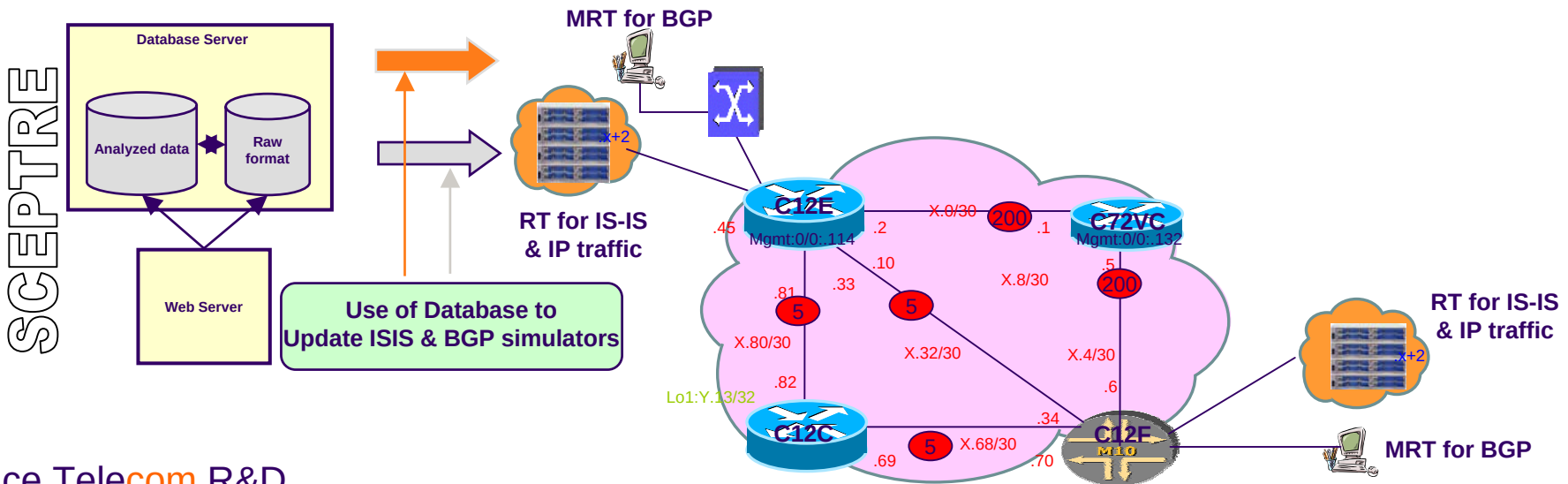
- Detection
  - No possible changes in IS-IS hello : too dangerous
  - POS timer optimization (depending of layer 2 protection timer)
- LSP generation /flooding
  - Small backoff LSP gen interval (avoids spacing too much LSPs)
- SPF and PRC triggering and computation
  - Small back off spf interval

	IS-IS before	IS-IS optimized
Failure detection	2000 ms	<= 5 ms
Delay before flooding LSP	33 ms	<= 5 ms
Delay before SPF computation	5500 ms	1 ms to max time (avg 20ms)
SPF computation time & RIB update	40 to 250 ms	40 to 250 ms

Tools to mesure convergence delay test

- Lab tests with real hardware are also necessary and complementary to control plane auditing and monitoring.
  - In order to evaluate equipment performance
  - In order to check that new equipments can effectively forward and reroute the traffic

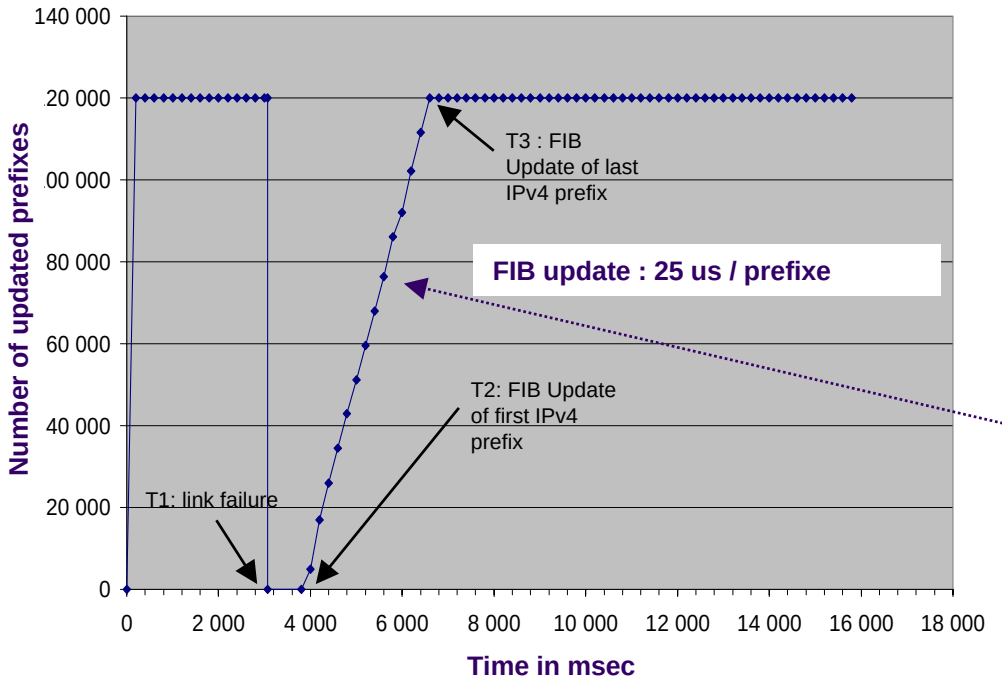
- Example of measurements
  - Line card FIB feeding speed can only be measured in the lab.
  - LSP generation mechanisms evolution depending upon version
  - SPF triggering and computation speed
  - SPF vs. Incremental SPF in realistic network conditions

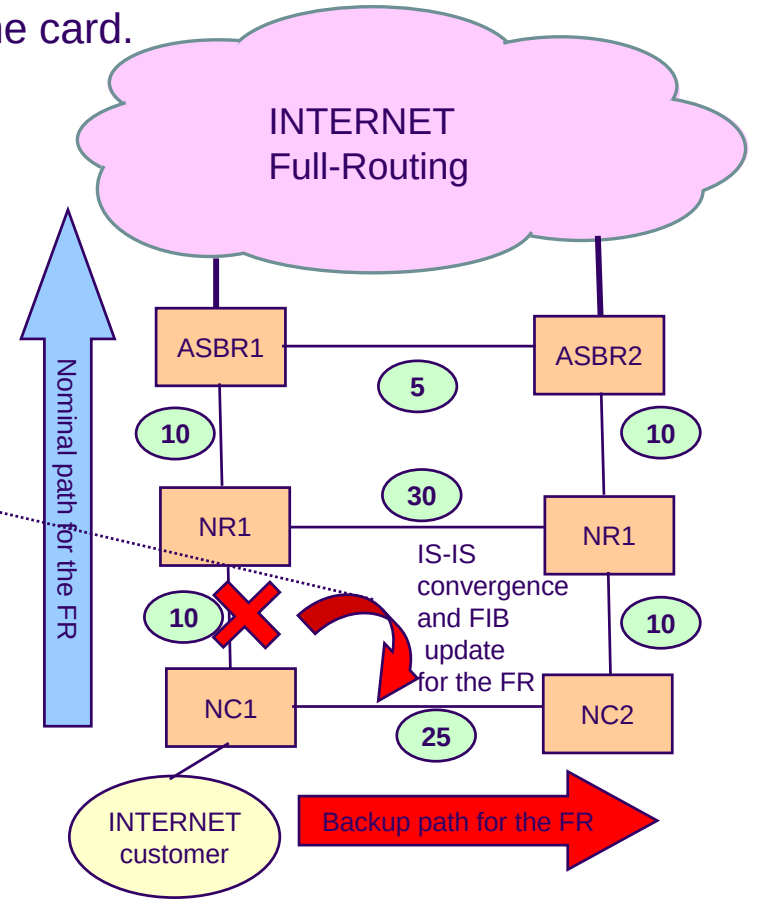


IS-IS convergence : FIB update

- **In some network situations for some equipments, the principal problem is the FIB update :**

  This time depends on the failure (ASBR, link in the backbone) and the number of prefixes that should be modified inside each line card.



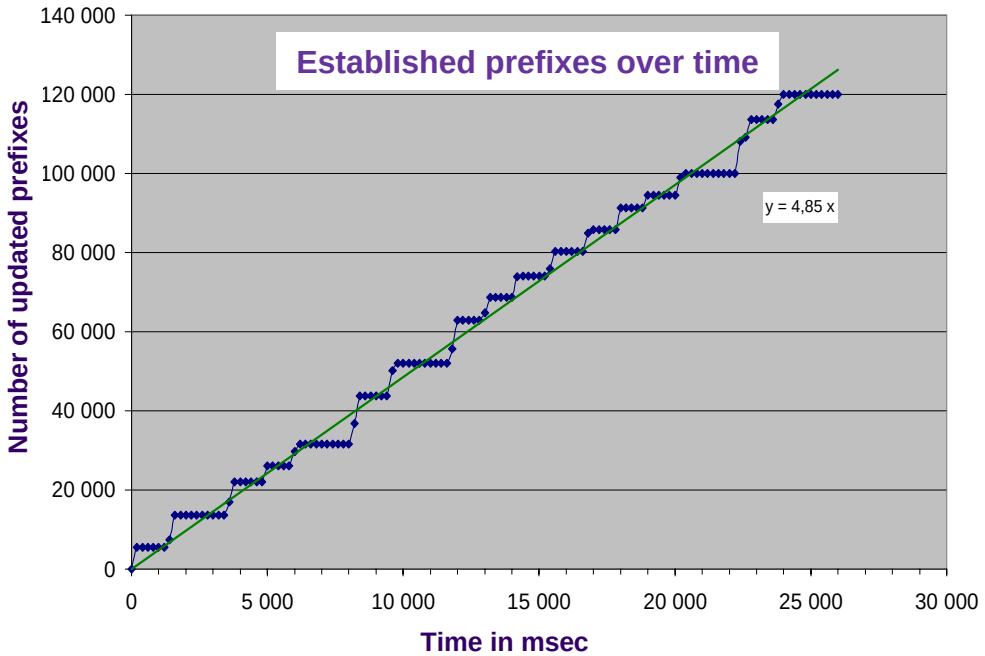


After a change of output interface without change of BGP nexthop in NC1 (decision base on local pref)

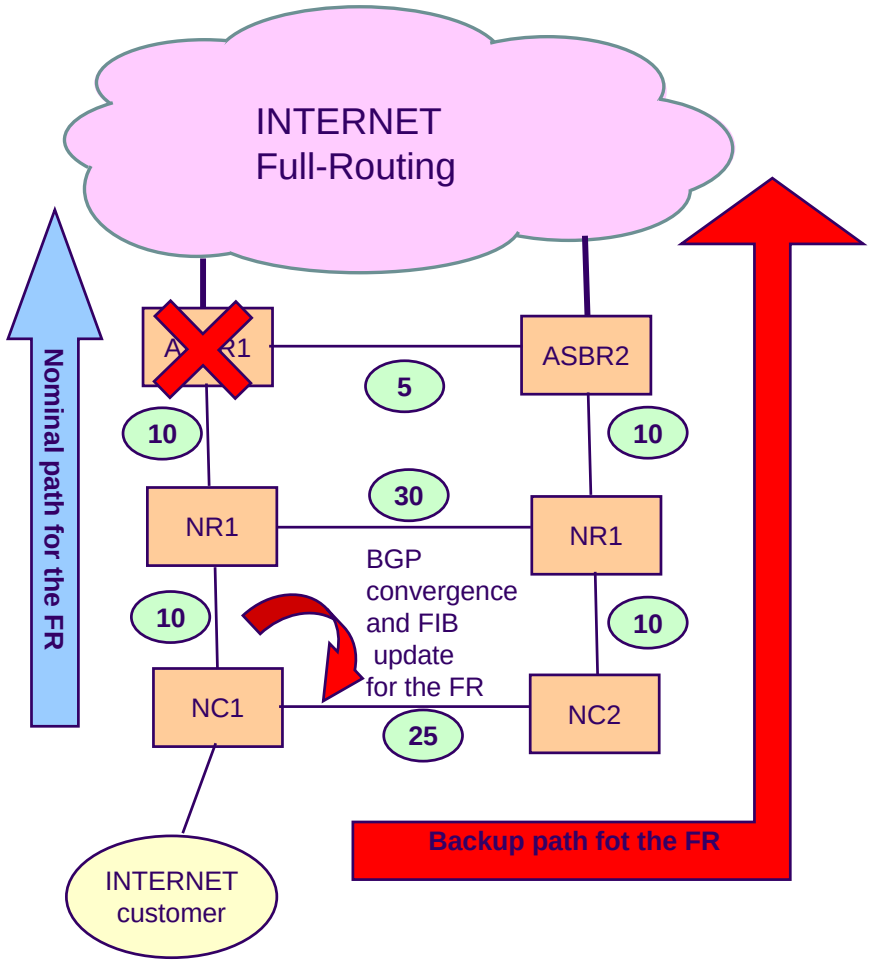
- **The solution is : Two stages lookup , CISCO introduce s this feature in 12.0.27S**

Updating a lot of prefixes

➡ **But in other failure case like :**



Time to update the FIB for all BGP prefixes (with a new next-hop). After an ASBR failure.



➡ **What is the hardware, software solution ? (maybe just a best network design ? ;-)**

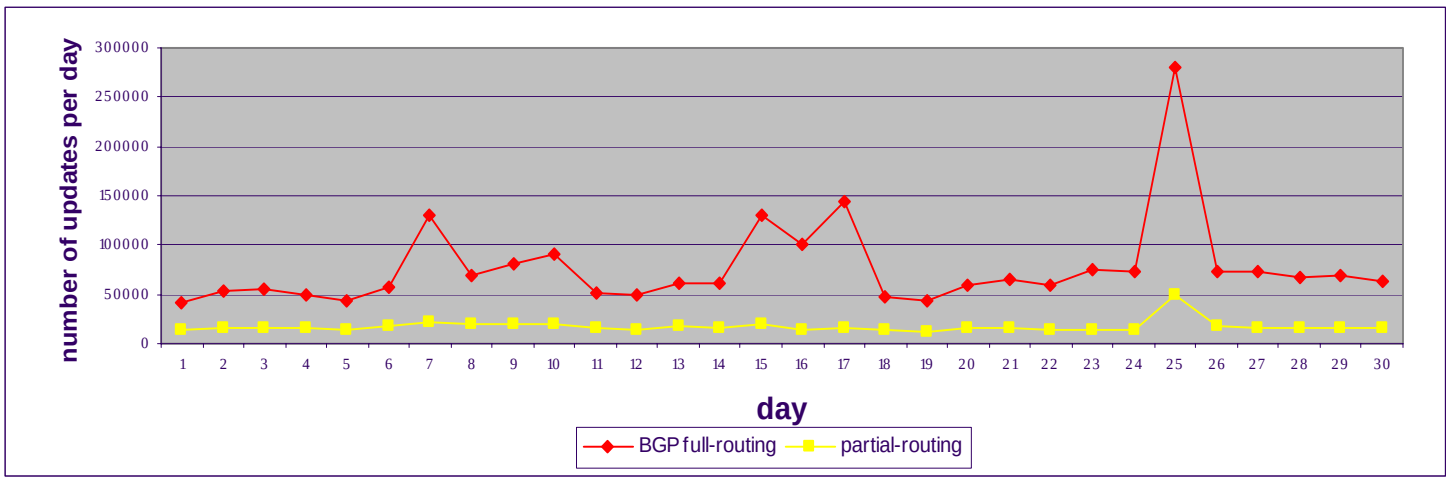
Solution for the FIB Update issue

Some possible solutions

- Hardware improvement: Two stages lookup at the edge or in the core;
- Engineering changes :
  - Reduce the number of IS-IS and BGP prefixes (for example : Partial Routing with default route)
  - Limit the number of next-hop changes.

Partial routing performance (Case of the collect network)

- It is not a perfect solution, but it provides a good improvement to the convergence in the core
- It can stabilize the network by reducing the amount of BGP updates



Summary

For Internet services

- IS-IS optimizations are sufficient to fulfill the service requirements with routers that have two stages look-up. **FT will deploy IS-IS fast convergence**.
- BFD will be a good solution to improve the detection time on GigaEthernet

For MPLS-VPN services

- Some additionnal tests are needed to detail failure cases and discuss the benefits of subsecond IS-IS convergence. There is a strong interaction with LDP.

Outlook

- Interoperability : Test the end-to-end results with different routers (CISCO – JUNIPER)
- Need to consider Fast Reroute to decrease the convergence time for new services ?
- A problem for updating a lot of prefixes in the FIB remains (in some special failure)
- Improve the Inter-AS convergence time