MPLS Protection Switching vs. OSPF Rerouting
A Simulative Comparison

Sandrine Pasqualini[1], Andreas Iselt[1], Andreas Kirstädter[1], and Antoine Frot[2]

[1] Siemens AG, Corporate Technology, Information & Communication, Munich, Germany
{sandrine.pasqualini|andreas.iselt|andreas.kirstaedter}@siemens.com
[2] École Nationale Supérieure des Télécommunications de Bretagne,Brest, France
antoine.frot@aitb.org

Abstract. Resilience is becoming a key design issue for future IP-based networks having a growing commercial importance. In the case of element failures the networks have to reconfigure in the order of a few hundred milliseconds, i.e. much faster than provided by the slow rerouting of current implementations. Several multi-path extensions to IP and timer modifcations have been recently proposed providing interesting alternatives to the usage of of MPLS below IP. In this paper these approaches are first described in a common context and then compared by simulations using very detailed simulation models. As one of the main results it can be shown that an accelerated update of the internal forwarding tables in the nodes together with fast hardware-based failure detection are the most promising measures for reaching the required reconfiguration time orders.

Key words: Resilience, OSPF, MPLS, Simulation.

1 Introduction

The current situation of the Internet is marked by the development and introduction of new real-time connection-oriented services like streaming technologies and mission-critical transaction-oriented services. Therefore, the Internet is gaining more and more importance for the economic success of single companies as well as of whole countries and network resilience is becoming a key issue in the design of IP based networks.

Originally, IP routing had been designed to be robust, i.e. to be able to reestablish connectivity after almost any failure of network elements. However, the applications mentioned only allow service interruptions on the order of a few hundred milliseconds - a time frame that cannot be reached by today's robust routing protocols. Therefore, several extensions and modifications have been proposed recently for speeding up IP protection performance: e.g. a simple reduction of the most important routing timer values or the large-scale introduction of IP multi-path operation with a fast local reaction to network element failures. Increasingly, network operators also deploy a designated MPLS layer below the IP layer having its own rather fast recovery mechanisms and providing failure-proof virtual links to the IP layer.

The most important aspect in the comparison of all these approaches is the resulting recovery speed. In order to thoroughly investigate the time-oriented behaviour of the alternatives we developed very detailed simulation models of the corresponding router/switch nodes. We implemented the single state machines and timing constants as extensions to the basic MPLS and OSPF models of the well-known Internet protocol simulation tool NS-2 [1]. The resulting simulator then was integrated into a very comfortable tool chain that allows the flexible selection of network topologies, traffic demands and protection mechanisms.

The rest of this paper is organized as follows: section 2 first describes MPLS and OSPF starting with MPLS basics and the two most interesting MPLS recovery mechanisms. This is followed by the description of the basic mechanisms of OSPF, the main time constants that were considered in the simulator, and the proposed extensions for faster reaction. In section 3 we describe the simulation framework, the enhancements implemented in the common public domain simulator NS-2 and the resulting tool chain. Section 4 details on the measurements we ran on the selected network topology and discusses the results obtained. Conclusions and recommendations for future hardware and protocol generations are given in section 5.

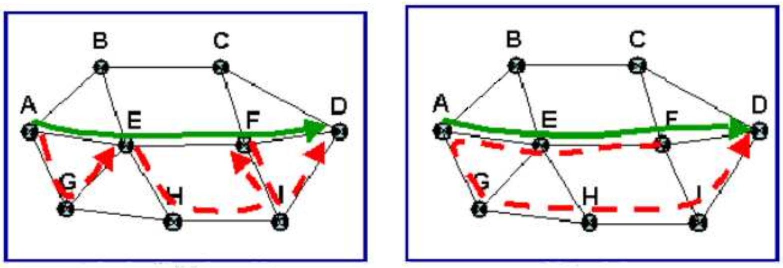
2 Resilience Mechanisms

2.1 Multiprotocol Label Switching (MPLS)

Label Switching. The routing in IP networks is destination-based: routers take their forwarding decisions only according to the destination address of a packet. Therefore, routing tables are huge and the rerouting process takes a correspondig amount of time. With Multiprotocol Label Switching (MPLS) ingress routers add labels to packets. These labels are interpreted by transient routers known as Label Switching Routers (LSR) as connection identifiers and form the basis for their forwarding decision. Each LSR re-labels and switches incoming packets according to its forwarding table. Label Switching speeds up the packet forwarding, and offers new efficient and quick resilience mechanisms. The setup of a MPLS path consists in the establishment of a sequence of labels, called Label Switched Path (LSP) that the packet will follow through the network. This can be simply done using conventional routing algorithms. But the main advantage of Label Switching appears when the forwarding decision takes the Quality of Service or links reservation into consideration. Then more complicated routing algorithms have to be used in order to offer the most efficient usage of the network.

MPLS Recovery. MPLS Recovery methods provide alternative LSPs to which the traffic can be switched in case of a failure. We must distinguish two types of recovery mechanisms: protection Switching and Restoration. The former includes recovery methods where a protection LSP is pre-calculated, just needing a switching of all traffic from the working LSP to the backup LSP after the failure detection. In the latter case, the backup LSP is calculated dynamically after the detection. Another way to classify these recovery mechanisms depends on which router along the LSP takes the rerouting decision: it can be done locally, the node detecting a failure immediately switching the traffic from the working to the backup LSP, or globally when the failure is notified to upstream and downstream LSRs that reroute the traffic. This paper will focus on Protection Switching schemes. Hereby Link Protection, similar to Cisco's Fast Reroute, and the mechanism introduced by Haskin [2] are considered further.

Link Protection provides a shortest backup path for each link of the primary LSP. When a failure occurs on a protected link, the backup path replaces the failed link in the LSP: the upstream router redirects incoming traffic onto the backup path and as soon as traffic arrives on the router downstream of the failed link it will use the primary LSP again. The Haskin scheme uses a global backup path for the LSP from ingress to egress router. When a failure occurs on a protected link the upstream router redirects incoming traffic back to the ingress router, which will be advertised that a failure has occurred. Then these packets are forwarded on the backup path and reach the egress router.



(a) Link Protection

(b) Haskin

Fig. 1. MPLS recovery mechanisms

Routes distribution. There are several possible algorithms to distribute labels through the network such as the Label Distribution Protocol (LDP), extended for Constraint-based Routing (CR-LDP). Another way is to distribute labels by piggybacking them onto other protocols, in particular the Reservation Protocol (RSVP) and its Traffic Engineering extension (RSVP-TE [3]).

2.2 OSPF

Today, one of the most common intra-domain routing protocols in IP networks is OSPF. This section shortly describes the OSPF mechanisms relevant for an understanding of the general behaviour and the various processing times and timers.

Basic OSPF mechanisms. The Hello protocol is used for the detection of topology changes. Each router periodically emits Hello packets on all its outgoing interfaces. If a router has not received Hello packets from an adjacent router within the "Router Dead Interval", the link between the two routers is considered down. When a topology change is detected, the information is broadcasted to neighbours via Link State Advertisements (LSA).

Each router maintains a complete view of the OSPF area, stored as a LSA Database. Each LSA represents one link of the network, and adjacent routers exchange bundles of LSAs to synchronise their databases. When a new LSA is received the database is updated and the information is broadcasted on outgoing interfaces.

Routes calculation: configurable cost values are associated to each link. Each router then calculates a complete shortest path tree[3]. However, only the next hop is used for the forwarding process.

The Forwarding Information Base (FIB) of a router determines which interface has to be used to forward a packet. After each computation of routes, the FIB must be reconfigured.

Main time constants. Considering the previous mechanisms, the convergence behaviour of OSPF in case of a failure can be divided into steps as follows : detection of the failure[4], then flooding of LSAs and - at the same time - scheduling of a SPF calculation, and launching a FIB update. Table 1 lists these times along with their typical values.

Proposed extensions to OSPF. Considering the standardized values, the OSPF protocol needs at least a few seconds to converge. To accelerate the convergence time, it is proposed to investigate the following two options: reduce delays, and associate multipath routing with local failure reaction. In the last years, there were several proposals [8,9] to accelerate OSPF convergence time by reducing the main timers : $T_{spfDelay}$ and $T_{spfHold}$ set to 0, and sub-second T_{Hello} or hardware failure detection. These accelerated variants of OSPF will be refered to in the following sections as $OSPF^{acc.}_{hello}$ when only sub-second hellos are used, and $OSPF^{acc.}_{hard}$ when hardware detection is enabled in addition. A new approach, proposed in [10] is to associate multipath routing with local failure reaction. This would allow to reduce the impact of a link failure by continuing to send traffic on the remaining paths. The OSPF standard [11] already allows to use paths with equal costs[5] simultaneously. In practice it is not straightforward to find link cost assignments yielding equal cost for several paths [12]. [10] presents a new routing scheme which provides each node in the network with two or more outgoing links towards every destination. Two or more possible next hops are then used at each router towards any destination instead of OSPF's single next hop.

[3] Shortest Path First (SPF) calculation

[4] by expiration of the Router Dead Interval or by reception of a new LSA

[5] Equal Cost Multi-Path (ECMP)

Table 1. Main time constants in OSPF

Name	Typically	Short Description
T_{Hello}	10s [4]	Interval between successive Hello packets
T_{Dead}	$4 \times T_{Hello}$ [5,6]	Router Dead Interval
T_{spf}	$\mathcal{O}(n.logn)$ $\mathcal{O}(n^2)^{(a)}$	SPF calculation
$T_{spfDelay}$	5s [5,6]	Minimum time between LSA reception and start of SPF computation
$T_{spfHold}$	10s [5,6]	Minimum time between consecutive SPF computations
T_{lsa}	0.6-1.1ms [7]	Process LSA : check if LSA is new and update LSA database
$T_{lsaFlood}$	33ms [7]	LSA flooding time : process LSA, bundle LSAs and pacing timer
T_{fib}	100-300ms [7]	Update the FIB : from end of LSA processing to end of new routes installation

(a) $2.53 \times 10^{-6} n^2 - 1.25 \times 10^{-5} n + 0.0012$, where n is the number of routers in the area, for details see [7]

In [10] such paths are called *hammocks*, due to their general structure where the multiple outgoing paths at one node may recombine at other nodes. The routing algorithms for calculating the hammocks where designed in order to fulfill the following criteria:

1. The algorithm must propose at least two outgoing links for every node,
2. if the topology is such as it is impossible to fulfill the first requirement, the algorithm should minimize the number of excpetions,
3. the algorithm must provide loop-free routing,
4. and no “single point of failure”[6],
5. it should minimize the maximum path length.

A router detecting a link or port failure can then react locally, immediately rerouting the affected traffic over the remaining next hops. This local mechanism avoids the time-consuming SPF calculation and flooding of LSAs in the entire area in the case of a single link failure. However, if multiple link failures occur and there is no remaining alternative link at a router, the local reaction will trigger a standard OSPF reaction. This multipath variant of OSPF will be refered to in the following sections as $OSPF^{hammock}_{hello}$ and $OSPF^{hammock}_{hard}$, depending on which detection mechanism is used.

[6] Such a node would prevent at least one other node from reaching a destination if it fails

3 Simulation Framework

In order to investigate the recovery performances of OSPF and MPLS, a simulation tool has been implemented. Based on the simulator NS-2 [1], it uses extensions such as the MPLS module MNS [13], the rtProtoLS module [14] and other protocol implementations, e.g. RSVP-TE Hellos. The OSPF implementation derives from rtProtoLS, to which a Hello protocol and timers have been added [15]. And the OSPF extensions were built from this implementation by changing the way the routes are calculated and the reactions to a failure are handled. The simulation scenario is specified in topology and traffic demand files, in NDL format (Network Description Language), an extension of GML [16]. NAM [1] is also used for the visualisation of the network activity. All tools are integrated into a comprehensive simulation framework, easily customizable through a simple GUI. This simulator automates the creation of OSPF or MPLS simulations for NS-2. Figure 2 shows how the different tools are articulated within the simulation framework. Given a topology, the MPLS Paths computation module① builds MPLS working and backup paths, using Dijkstra's algorithm, and exports them in NDL format. Supported recovery schemes are Link Protection, similar to Cisco's Fast Reroute, and the method of Haskin [2]. After giving some parameters, such as triggering link failures, a tool translates all NDL sources into one NS-2 simulation file②. For the OSPF simulations, the NS-2 simulator has been extended to allow local external routing algorithms③. This allows to use existing routing tools and to develop routing independently from NS-2. The results are visualized in NAM ④.

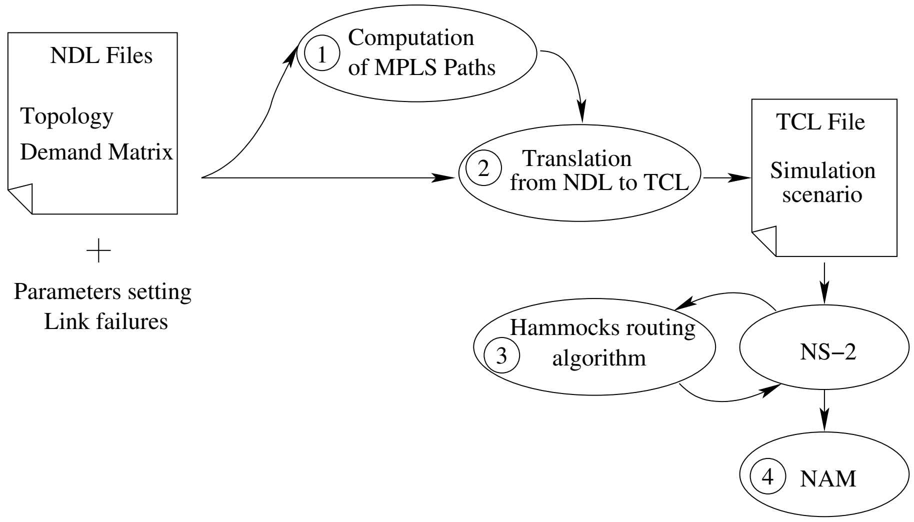


Fig. 2. Tool chain

4 Measurements and results

The focus of the investigations was on the speed of the traffic restoration after a failure. As a main sample network, the Pan-European optical network from the COST 239 project [17] was chosen because of its widespread use for network investigations. This network, shown in Fig.3 contains 11 nodes and 26 links with capacities of 20 Gbit/s. A full-mesh of equal flows between all nodes has been used as demand pattern. To save simulation time, the link bandwidths are scaled down by a factor of 1000. The sources send packet flows with 800kbit/s constant bit rate (CBR) (packets of 500 bytes sent every 5 ms). This allows more than 20 simultaneous flows on one link without any packet loss. The simulation starts with the establishment of the network configuration. For MPLS this includes the set up of paths and backup paths. For OSPF this means the convergence of the OSPF routing protocol. After starting the sources, a link failure is simulated triggering failure detection, dynamic route calculation, if necessary, and switching to alternative routes. To get rid of synchronisation effects of hello timers with failure times, the simulations are repeated with different periods of time between the simulation start and the failure time. The simulation is also repeated for all possible link failures, to average over the effect of different failure locations. To characterise the effect of the failure, the sum of the rates of all traffic received at sinks in the network is considered over the time. Fig. 4 shows the affected traffic and the times for restoration for different MPLS protection switching and IP rerouting approaches both with different timer values for the RSVP refresh messages or for the OSPF hello protocol. Each curve in Fig. 4 shows the sum of all traffic flows in the network. After the occurrence of a failure the sum rate decreases since the traffic that is expected to be carried over the failed link is lost. Just after the link is repaired, shortest routes are used again while packets

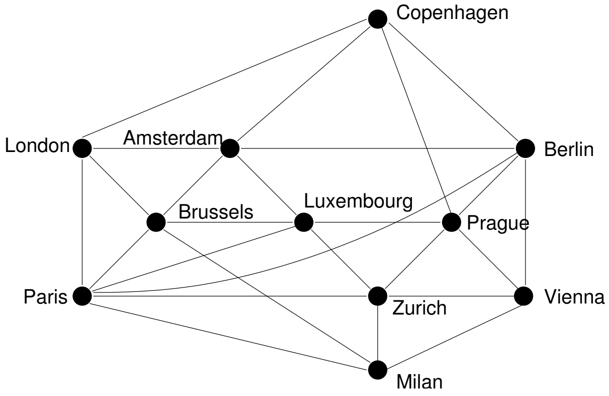


Fig. 3. COST239 network

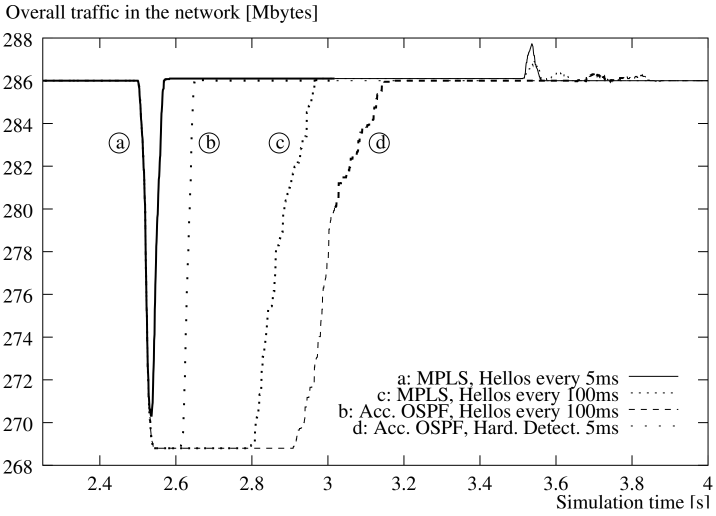


Fig. 4. Comparison between MPLS and accelerated OSPF recovery

are still on the alternate routes, which results in more packets reaching their destination during a few milliseconds. The four curves represent the cases:

- MPLS Link Protection[7] with RSVP-TE standard failure detection intervals of 5ms ⓐ and 100ms ⓒ.
- $OSPF^{acc.}_{hello}$ with modified hello intervals of 100ms ⓓ and $OSPF^{acc.}_{hard}$ with hardware failure detection of 5ms ⓑ.

It can be noticed that standard MPLS protection switching, ⓐ, is much faster than both OSPF mechanisms. Even MPLS ⓒ, with the same T_{Hello} and T_{Dead} timers as $OSPF^{acc.}_{hello}$ is still faster, in the order of 100ms. This results from the computational effort, the signalling delay and mostly from the update of the FIBs, which is more time consuming for the larger tables of OSPF - compared to MPLS. Of course, this is a very implementation dependent parameter and may be addressed in future router developments. The effect of hardware failure detection is shown in Fig. 5. Obviously the hardware failure detection[8] speeds up the OSPF recovery considerably. This figure also shows a difference between shortest path routing ⓕⓖ, and multi-path routing ⓔⓗ, as it is described in [10]. With multi-path routing the traffic is distributed over a fan of paths, including paths longer than the shortest paths. Therefore the probability for such a path to be hit by a single link failure is higher. This results in the increased impact represented by the lower throughput in the case of a failure. Fig. 6 depicts the different times involved in the extended OSPF implementation, with the values used for the simulations. The predominant times here are the detection of

[7] the Haskin cases give similar results regarding reconfiguration time

[8] this timer is set to 5ms, which is realistic regarding current physical possibilities

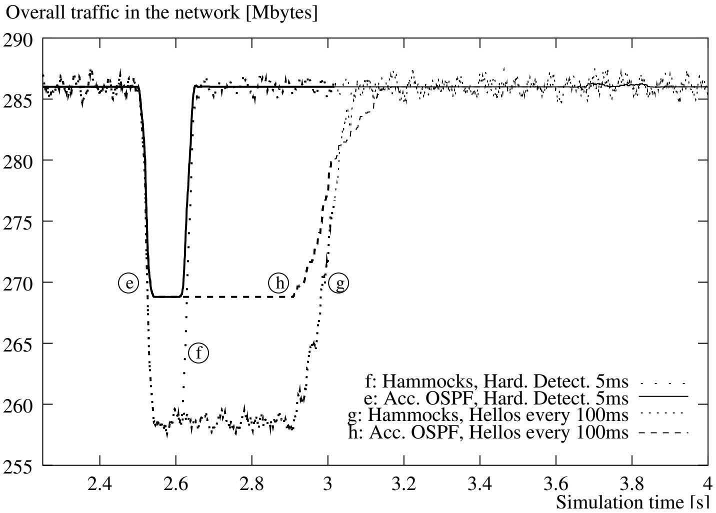


Fig. 5. Comparison between Hammocks and OSPF routing

failures and the updating of the forwarding tables. For larger networks, the LSA processing times also have to be considered. This indicates clearly where future improvements in OSPF and router technology are necessary: failure detection and FIB update. To reduce the failure detection time, hardware failure detection

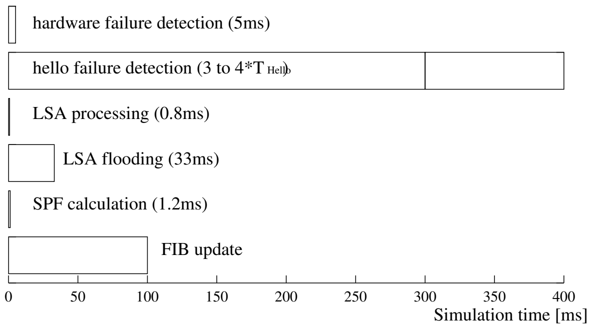


Fig. 6. Relative size of the various times involved in OSPF implementation.

already gives major relief. Moreover, where hardware failure detection does not help, short hello intervals will also allow faster failure detection. In [18] a protocol is proposed allowing the use of short hello intervals independent of the routing protocol. The other major time that has to be improved is the FIB update time. As already mentioned above, this requires changes in the router implementation.

5 Conclusion

The current Internet routing protocol OSPF as it is implemented and used today has major deficiencies with respect to network resilience. The simulative comparison with MPLS-enhanced networks shows the superior time behavior of MPLS resilience. We have outlined that there are several proposed extensions to improve the resilience of routed networks. These proposals include optimization of timers and the use of multi-path routing with local failure reaction. At investigating the extensions by simulation it turned out that they are the first steps in the right direction. From the investigations it can be concluded that there are two major points to be addressed in order to improve the restoration speed of OSPF re-routing: speed-up of failure detection and acceleration of forwarding information base (FIB) update. For the former some very promising approaches, like hardware failure detection and fast hello protocols (e.g. BFD [18]) are already evolving. For the acceleration of the FIB updates the internal router architectures have to be improved. With these extensions OPSF routed networks will be able to reach sub-second restoration speeds in the future.

References

1. UCB/LBNL/VINT. [Online]. Available: http://www.isi.edu/nsnam/
2. D. Haskin and R. Krishnan, "A method for setting an alternative label switched paths to handle fast reroute," IETF," Internet Draft.
3. D. Awduche et al., "RSVP-TE: Extensions to RSVP for LSP tunnels," NWG, RFC 3209.
4. Cisco Corp. [Online]. Available: http://www.cisco.com
5. C. Huitema, *Routing in the Internet*, 2nd ed. Prentice Hall PTR, 2000.
6. J. T. Moy, *OSPF: Anatomy of an Internet Routing Protocol*, nov 2000.
7. A. Shaikh and A. Greenberg, "Experience in black-box ospf measurement," in *ACM SIGCOMM Internet Measurement Workshop (IMW)*, nov 2001.
8. C. Alaettinoglu et al., "Toward millisecond IGP convergence," IETF," Internet Draft.
9. A. Basu and J. Riecke, "Stability issues in OSPF routing." ACM SIGCOMM, aug 2001.
10. G. Schollmeier et al., "Improving the resilience in IP networks," in *HPSR 2003*, jun 2003.
11. J. Moy, "OSPF version 2," NWG, RFC 2328.
12. A. Sridharan et al., "Achieving near-optimal traffic engineering solutions for current OSPF/IS-IS networks." INFOCOM 2003.
13. G. Ahn. Mns. [Online]. Available: http://flower.ce.cnu.ac.kr/~fog1/mns/
14. M. Sun. [Online]. Available: http://networks.ecse.rpi.edu/~sunmin/rtProtoLS/
15. C. Harrer, "Verhalten von IP-routingprotokollen bei ausfall von netzelementen," Master's thesis, Technische Universität München, LKN, apr 2001.
16. M. Himsolt. Gml: A portable graph file format. Universität Passau. [Online]. Available: http://www.infosun.fmi.uni-passau.de/Graphlet/GML/
17. P. Batchelor et al., "Ultra high capacity optical transmission networks," COST 239," Final report of Action, jan 1999.
18. D. Katz and D. Ward, "Bfd for ipv4 and ipv6 (single hop)," IETF," Internet Draft.