Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Prof. Laurent Vanbever
Networked Systems Group

Data-plane driven fast network convergence

Master thesis proposal

With the always increasing level performance required by current applications running on the Internet (*e.g.*, realtime voice or video services), retrieving connectivity upon a failure is of paramount importance for IP networks. When a link or a node fails in an IP network, the routers adjacent to the failing resource must react by distributing new routing information to allow each router of the network to update its forwarding table. To do so, control-plane messages (*e.g.*, with OSPF or IS-IS) are exchanged between the routers, and this will in turn trigger a recomputation of the shortest paths and an update of the routing table on each router. Even for carefully configured networks using link state IGPs, this process can take few tens or hundreds of milliseconds [5], a time during which the forwarding state of the network is not valid, resulting in forwarding loops or blackholes. In addition, potentially hundreds or thousands of forwarding entries may need to be updated (one for each affected prefix). Given the time required to update a FIB entry and the always increasing number of prefixes advertised in the Internet, this process can take from few hundreds of milliseconds up to several tens of seconds, a time during which the traffic is lost [6].

Several techniques have been designed to speed-up de convergence in link-state IP networks. IP fast reroute [4, 7], MPLS fast reroute [1] or PIC [3] can locally fast reroute all the affected traffic in 100 to 200ms only by precomputing backup next-hops (or establishing backup MPLS tunnels) and activate them upon the failure. These techniques *locally* fast reroute, meaning that other routers in the network are not aware about this traffic deflection. To avoid transient loops, these techniques either use loop-free backup paths [7] or signal the backup paths prior to the failure with RSVP-TE [1]. However, loop-free backup paths can be suboptimal, and sometimes such paths just do not exist [4]. On the other hand, signaling backup paths with RSVP-TE also implies to enable this protocol in the network even if it is not used to forward packets when the network is stable.

In this thesis proposal, we propose to go one step further by *globally* fast converging the network upon a failure. By leveraging programmable switches in the network, we believe it is possible to converge the network within few milliseconds only. This global fast convergence will allow the routers to use the best paths as backup paths, without requiring additional control-plane signaling. The key intuition behind this global fast convergence is that data-plane tags can be embedded in the packets by the router detecting the failure. When these data-plane tags are processed by the other routers in the network, they activate backup forwarding rules (possibly prepopulated) which forward the affected traffic towards the best backup path. As the packet containing the data-plane tag go through the backup path, the routers in the path will converge, ensuring a fast and loop-free convergence.

Requirements For this thesis, we ask the student to perform the following tasks:

- A thorough understanding of the existing fast reroute techniques
- Design a data-plane driven fast convergence following the approach described above
- Implement this approach in P4 [2]
- Evaluate the proposed solution when fully or only partially deployed on simulated networks with software switches and routers
- Compare the performance of the proposed solution with the existing solutions
- Make sure the proposed P4 program runs on current P4 enabled switches

Contact

- Thomas Holterbach, thomahol@ethz.ch
- Edgar Costa Molero, cedgar@ethz.ch
- Prof. Dr. Laurent Vanbever, lvanbever@ethz.ch

References

[1] A. Atlas, G. Swallow, and P. Pan. Fast Reroute Extensions to RSVP-TE for LSP Tunnels. RFC 4090, May 2005.

[2] P. Bosshart, D. Daly, G. Gibb, M. Izzard, N. McKeown, J. Rexford, C. Schlesinger, D. Talayco, A. Vahdat, G. Varghese, and D. Walker. P4: Programming protocol-independent packet processors. *SIGCOMM Comput. Commun. Rev.*, 44(3):87–95, July 2014.

[3] C. FilsFils, P. Mohapatra, J. Bettink, P. Dharwadkar, P. D. Vriendt, Y. Tsier, V. V. D. Schrieck, O. Bonaventure, and P. Francois. Bgp prefix independent convergence (pic) technical report. Technical report, Cisco, 2011.

[4] P. Francois and O. Bonaventure. An evaluation of ip-based fast reroute techniques. In *Proceedings of the 2005 ACM Conference on Emerging Network Experiment and Technology*, CoNEXT '05, pages 244–245, New York, NY, USA, 2005. ACM.

[5] P. Francois, C. Filsfils, J. Evans, and O. Bonaventure. Achieving sub-second igp convergence in large ip networks. *SIGCOMM Comput. Commun. Rev.*, 35(3):35–44, July 2005.

[6] T. Holterbach, S. Vissicchio, A. Dainotti, and L. Vanbever. SWIFT: Predictive Fast Reroute. In *ACM SIGCOMM*, Los Angeles, USA, August 2017.

[7] M. Shand and S. Bryant. IP Fast Reroute Framework. RFC 5714, Jan. 2010.