Supercharge me: Boost Router Convergence with SDN

Michael Alan Chang
Princeton University/ETH Zürich
machang@princeton.edu

Thomas Holterbach
ETH Zürich
thomahol@ethz.ch

Markus Happe
ETH Zürich
markus.happe@tik.ee.ethz.ch

Laurent Vanbever
ETH Zürich
lvanbever@ethz.ch

ABSTRACT

By enabling logically-centralized and direct control of the forwarding behavior of a network, Software-Defined Networking (SDN) holds great promise in terms of improving network management, performance, and costs. Realizing this vision is challenging though as SDN proposals to date require substantial and expensive changes to the existing network architecture before the benefits can be realized. As a result, the number of SDN deployments has been rather limited in scope. To kickstart a wide-scale SDN deployment, there is a need for low-risk, high return solutions that solve a timely problem. As one possible solution, we show how we can significantly improve the performance of legacy IP routers, *i.e.* "supercharge" them, by combining them with SDN-enabled devices. In this abstract, we supercharge one particular aspect of the router performance: its convergence time after a link or a node failure.

CCS Concepts

•**Networks** → *Routers; Network performance analysis;*

1. TODAY'S (SLOW) CONVERGENCE

The convergence time of traditional IP routers is directly linked to the time it takes for the router to update its hardware-based Forwarding Information Base (FIB) after it detects the failure. Each FIB entry maps an IP destination to the L2 Next-Hop address (*i.e.*, MAC address) of the chosen IP NH as well as the output in-

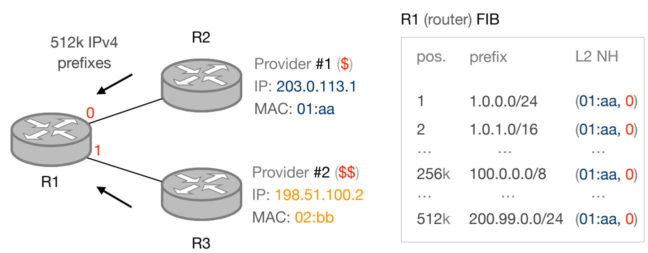

Figure 1: Upon a failure of $R2$, each of the 512k FIB entries have to be updated to restore full connectivity, a time-consuming operation.

terface. In most routers, the FIB is flat, meaning each FIB entry is mapped to a different (but possibly identical in content) L2 NH entry. This is illustrated by Fig. 1, which depicts a simple network where $R1$ is an edge router connected to the router of two providers, $R2$ and $R3$. Each of these provider routers advertise a full Internet routing table composed of more than 512,000 IPv4 prefixes [1]. As $R2$ is cheaper than $R3$, $R1$ is configured to prefer $R2$ for all destinations. In such a case, each of the 512k FIB entries in $R1$ is associated to a distinct L2 NH entry which all contain the physical MAC address of $R2$. Upon the failure of a $R2$, every single entry of $R1$ FIB has to be updated creating a significant downtime. Our measurements on a recent router (see §3) show that it actually takes *several minutes* for $R1$ to fully converge, during which traffic is lost. With the ever rising cost of downtime [2] and as services increasingly rely on high-availability, convergence of the order of minutes is simply not acceptable.

2. SUPERCHARGING CONVERGENCE

Equipping routers with a hierarchical FIB [3] is an obvious solution to the convergence problem mentioned above. In a hierarchical FIB, each IP destination is mapped to a pointer that resolves to the actual L2 NH to be used. Upon failure of a L2 NH, only pointer values have to be updated. Since the number of L2 NH is

SIGCOMM '15 August 17-21, 2015, London, United Kingdom

ACM ISBN 978-1-4503-3542-3/15/08.
DOI: http://dx.doi.org/10.1145/2785956.2790007

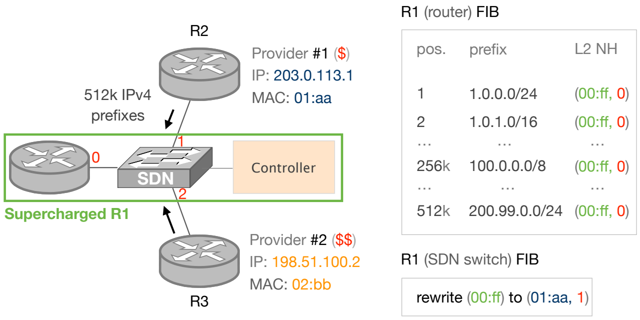

Figure 2: With a supercharged router, upon failure of $R2$, only *one entry*—the pointer value—needs to update to restore full connectivity.

orders of magnitude smaller than the number of FIB entries, convergence is greatly improved. Unfortunately, hierarchical FIB designs also means much more complex hardware, and therefore, more expensive routers.

Fig. 2 illustrates how we can provide *any* router (here $R1$) with a hierarchical FIB, spanning two devices, by combining it with a SDN switch. To provision forwarding entries in this hierarchical FIB, we built a *supercharged controller*. While the controller can rely on (typically) OpenFlow to provision forwarding entries in a SDN switch, dynamically provisioning specific forwarding entries in a router is trickier. Our key insight is that the supercharged controller can use any routing protocol spoken by the router as a provisioning interface. Indeed, FIB entries in a router directs traffic to the L2 NH associated to the L3 NH learned via the routing protocol. To do so, our supercharged controller first interposes itself between the router and its peers (see Fig. 2). Then, it computes primary and backup NH for all IP destinations. Finally, it provisions L2 NH "pointers" by setting the IP NH field to a virtual L3 NH that gets resolved by the router into a L2 NH using ARP. Upon failure of $R2$ in Fig. 2, all the controller has to do to converge is to modify the switch rule to (rewrite(00:ff) to (02:bb,2)) in order to converge *all traffic* to $R3$.

3. EVALUATION

Setup and methodology. Our setup is based on the scenario depicted in Fig. 2. It consists of 3 routers Cisco Nexus 7k C7018 interconnected through a HP E3800 J9575A Openflow-enabled switch. We measured the convergence time of $R1$ with and without supercharging it. To do so, we similarly loaded $R2$ and $R3$ with an increasing number of actual BGP routes collected from the RIPE RIS dataset [4]. In both cases (supercharged and not supercharged), $R1$ was configured to prefer $R2$ for all the destinations. Once all routes were advertised, we started to inject traffic at $R1$ using a FPGA-based generator. We configured $R2$ and $R3$ to send all receiving traffic to another FPGA-based board, acting as sink. We subsequently disconnected $R2$ from the switch, triggering the convergence process at $R1$; subsequently, we measure the time until recovering full connectivity. Results are depicted in Fig. 3.

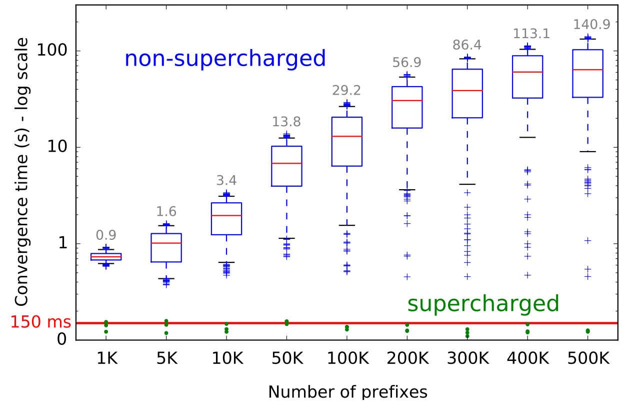

Figure 3: Convergence time with respect to number of advertised prefixes.

The non-supercharged R1 takes ~2.5min to converge in the worst case. The convergence time is linearly proportional to the number of prefixes because FIB entries are updated one-by-one. This worst-case highlights undesirability of the non-supercharged approach: as the FIB grows, so does the convergence time.

The supercharged R1 systematically converges *within 150ms*, for all prefixes. Thanks to its hierarchical FIB design, the supercharged R1's convergence time is constant—irrespective of the number of prefixes. This constitutes a 900× improvement factor over the worst case of the non-supercharged solution.

4. CONCLUSION

We boosted the convergence time of legacy routers by combining them with SDN equipment in a novel way, essentially building a hierarchical forwarding table spanning across devices.

We believe this paper opens up many interesting future directions for integrating legacy routing and SDN devices in a more "symbiotic way". By juxtaposing the agility of the SDN with the tried-and-true routers prevalent in the industry, we take the best of both worlds and take the first steps towards electrifying modern day networks through supercharged networking devices.

5. REFERENCES

[1] "CIDR report," http://www.cidr-report.org/as2.0/.
[2] C. Cérin, C. Coti, P. Delort, F. Diaz, M. Gagnaire, Q. Gaumer, N. Guillaume, J. Lous, S. Lubiarz, J. Raffaelli *et al.*, "Downtime statistics of current cloud solutions," *International Working Group on Cloud Computing Resiliency, Tech. Rep*, 2013.
[3] C. Filsfils, P. Mohapatra, J. Bettink, P. Dharwadkar, P. De Vriendt, Y. Tsier, V. Van Den Schrieck, O. Bonaventure, and P. Francois, "BGP Prefix Independent Convergence (PIC)," Cisco, Tech. Rep., 2011.
[4] "RIPE RIS Raw Data," https://www.ripe.net/data-tools/stats/ris/ris-raw-data.