ETH zürich

Prof. Laurent Vanbever
Networked Systems Group

Anycast Load Distribution at Network Telemetry Data Collection

Master thesis proposal (in collaboration with Swisscom)

Swisscom collects millions of Network Telemetry [1] metrics every second with BMP [2–4], IPFIX [5] and YANG push [6] from thousands of network devices. This poses a challenge in terms of load distribution at the data-collection across servers, Linux network sockets [7] and collector daemons.

To accommodate to the ever-increasing amount of telemetry data, this thesis explores a solution based on Anycast [8] with ECMP [9] to distribute traffic across Layer 3 links and routers to various collection servers. On the server itself, we will extend a Linux network socket option called SO_REUSEPORT which allows to bind multiple sockets to the same IP and port combination (which is normally not possible) thereby distributing the incoming telemetry data between multiple listening collection daemons. Finally, we will research potential extensions to the TCP handshake when combined with SO_REUSEPORT. A normal TCP session would timeout when it is taken over by a new collector daemon (e.g., in case of a network failure or maintenance) and would need to re-establish the session with a full TCP handshake. This process is not only slow, it could also lead to traffic loss which is suboptimal in a network telemetry scenario.

The thesis can be divided into four main parts:

- Familiarize yourself with Swisscom's current Network Telemetry collection process based on BMP, IPFIX and YANG push and how the collected data improves network visibility. You will also understand how Anycast with ECMP works and enables traffic distribution.
- Extend the existing SO_REUSEPORT Linux implementation to better support tasks specific to telemetry data collection. During this work you will be closely supported by Prof. Pierre Francois' team (INSA university in Lyon [10]).
- Test, verify and analyze your improvements in the Swisscom IETF interoperability lab which contains servers and routers from various vendors. If everything works as intended, you will help with the rollout into Swisscom's production network.
- Research and document TCP handshake changes to prevent TCP session timeouts when a session is taken over by a new collection daemon sharing the same Linux network socket with SO_REUSEPORT.

Requirements

- Good understanding of BGP [11] and experience with Linux.
- Motivated to work as part of an interdisciplinary team (people from industry and academia).
- Not afraid to program in C (experts will support you).
- Knowledge of Linux network TCP/IP stack or large scale MPLS VPN [12] networks are a plus but *not* required.

Important To best match Swisscom's schedule the thesis start has to be in March.

Towards the end of your thesis (July 24-30) you can present your results at the IETF 111 GROW working group which is a great opportunity to meet other operators, vendors and people from academia.

Contact If interested, please contact Tobias Bühler (buehlert@ethz.ch). We can then schedule a short call with people from Swisscom to discuss further details.

References

[1] H. Song, F. Qin, P. Martinez-Julia, L. Ciavaglia, and A. Wang. Network Telemetry Framework. https://tools.ietf.org/html/draft-ietf-opsawg-ntf-05.

[2] J. Scudder, R. Fernando, and S. Stuart. BGP Monitoring Protocol (BMP). https://www.rfc-editor.org/info/rfc7854.

[3] T. Evens, S. Bayraktar, M. Bhardwaj, and P. Lucente. Support for Local RIB in BGP Monitoring Protocol (BMP). https://tools.ietf.org/html/draft-ietf-grow-bmp-local-rib-08.

[4] C. Cardona, P. Lucente, P. Francois, Y. Gu, and T. Graf. BMP Extension for Path Status TLV. https://tools.ietf.org/html/draft-cppy-grow-bmp-path-marking-tlv-07.

[5] B. Claise, B. Trammell, and P. Aitken. Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of Flow Information. https://tools.ietf.org/html/rfc7011.

[6] G. Zheng, T. Zhou, T. Graf, P. Francois, and P. Lucente. UDP-based Transport for Configured Subscriptions. https://tools.ietf.org/html/draft-ietf-netconf-udp-notif-01.

[7] Linux socket interface. https://man7.org/linux/man-pages/man7/socket.7.html.

[8] C. Partridge, T. Mendez, and W. Milliken. Host Anycasting Service. https://tools.ietf.org/html/rfc1546.

[9] C. Hopps. Analysis of an Equal-Cost Multi-Path Algorithm. https://tools.ietf.org/html/rfc2992.

[10] National Institute of Applied Sciences of Lyon. https://www.insa-lyon.fr/en/.

[11] Y. Rekhter, T. Li, and S. Hares. A Border Gateway Protocol 4 (BGP-4). https://tools.ietf.org/html/rfc4271.

[12] E. Rosen and Y. Rekhter. BGP/MPLS IP Virtual Private Networks (VPNs). http://www.rfc-editor.org/info/rfc4364.