Contribution ID: **143** Type: **not specified**

From XDP to Socket

Thursday 23 September 2021 08:40 (40 minutes)

In this talk, we describe important challenges in L4 and L7 load balancing for the consistent routing of packets across hosts as well as across sockets within a host, once a packet is received in the XDP based L4LB. We then describe how we leverage recent additions on the BPF programs to address those challenges.

Typically some form of Consistent Hashing is used to pick an end host for incoming packets within an L4 LB [2]. Such mechanisms, however, pose challenges in maintaining routing consistency over a long window of time without sharing routing states among the L4LBs. In Facebook, we devised a novel server-id based routing for completely stateless routing of both TCP and QUIC connection. For routing of TCP packets, we leverage tcp_hdr_opt [1] to encode server_id between the endpoints.

'Zero downtime restart'[3] supported by many L7 Proxies, such as Proxygen in Facebook, require lots of custom userspace solution for routing consistency, especially for UDP payloads. Further, maintaining uniform load across individual sockets and CPU cores in a host is not straightforward without custom solutions. We describe how we leverage SOREUSEPORT_SOCKARRAY to create a framework that allows us to efficiently and effectively address both problems by:
a) Being able to make routing decision in picking up a socket on per packet (UDP) and per connection (TCP) basis
b) Being able to granularly target individual CPU core to handle incoming packets
This has allowed us to run at scale with minimal operation load and further simplify our implementation to execute disruption free restart of L7 proxy [3].

References

1. M. Lau. BPF TCP header option. https://lwn.net/Articles/827672/
2. D. E. Eisenbud, C. Yi, C. Contavalli, C. Smith, R. Kononov, E. Mann-Hielscher, A. Cilingiroglu, B. Cheyney, W. Shang, and J. D. Hosein. Maglev: A fast and reliable software network load balancer. In USENIX Symposium on Networked Systems Design and Implementation (NSDI), Mar. 2016
3. U Naseer, L Niccolini, U Pant, A Frindell, R Dasineni, TA Benson. Zero Downtime Release: Disruption-free Load Balancing of a Multi-Billion User Website SIGCOMM '20: Proceedings of the Annual Conference of the ACM Special
4. Katran - A high performance layer 4 load balancer. https://bit.ly/38ktXD7.

I agree to abide by the anti-harassment policy

I agree

Primary authors: PANT, Udip (Facebook); LAU, Martin (Facebook)

Presenters: PANT, Udip (Facebook); LAU, Martin (Facebook)

Session Classification: BPF & Networking Summit

Track Classification: Networking & BPF Summit (Closed)