Contribution ID: **270** Type: **not specified**

Programmable socket lookup with BPF

Monday, September 9, 2019 12:45 PM (45 minutes)

At Netconf 2019 we have presented a BPF-based alternative to steering packets into sockets with iptables and TPROXY extension. A mechanism which is of interest to us because it allows (1) services to share a port number when their IP address ranges don't overlap, and (2) reverse proxies to listen on all available port numbers.

The solution adds a new BPF program type BPF_INET_LOOKUP, which is invoked during the socket lookup. The BPF program is able to steer SKBs by overwriting the key used for listening socket lookup. The attach point is associated with a network namespace.

Since then, we have been reworking the solution to follow the existing pattern of using maps of socket references for redirecting packets, that is REUSEPORT_SOCKARRAY, SOCKMAP, or XSKMAP. We expect to publish the next version of BPF_INET_LOOKUP RFC patch set, which addresses the feedback from Netconf, in August.

During LPC 2019 BPF Microconference we would like to briefly recap on how BPF-driven socket lookup compares to classic bind()-based dispatch, TPROXY packet steering, and socket dispatch on TC ingress currently in development by Cilium.

Next we would like discuss low-level implementation challenges. How to best ensure that packet delivery to connected UDP sockets remains unaffected? Can a BPF_INET_LOOKUP program co-exist with reuseport groups? Is there a possibility of code sharing with REUSEPORT_SOCKARRAY implementation?

Following the implementation discussion, we will touch on performance aspects, that is what is the observed cost of running BPF during socket lookup both in SYN flood and UDP flood scenarios.

Finally, we want to go into the usability of user-space API. Redirection with a BPF map of sockets raises a question who populates the map, and if existing network applications like NGINX need to be modified in any way to receive traffic steered with this new mechanism.

The desired outcome of the discussion is to identify steps needed to graduate the patch set from an RFC series to a ready-for-review submission.

I agree to abide by the anti-harassment policy

Yes

I confirm that I am already registered for LPC 2019

Primary authors: SITNICKI, Jakub (Cloudflare); BAUER, Lorenz (Cloudflare); MAJKOWSKI, Marek (Cloudflare)

Presenters: SITNICKI, Jakub (Cloudflare); BAUER, Lorenz (Cloudflare); MAJKOWSKI, Marek (Cloudflare)

Session Classification: Networking Summit Track