Contribution ID: **147** Type: **not specified**

Dynamic Encapsulation Using eBPF

Thursday, 23 September 2021 07:00 (40 minutes)

Prior to LWT (Lightweight Tunnels) and modern eBPF, the only way to send encapsulated packets to multiple destinations was achieved by creating multiple tunnel devices which didn't scale well when thousands of different destinations were needed.

In the past Google solved this problem by introducing custom patches on top of the ip gre device to allow sockets to provide the destination address and encapsulation protocol to change the encapsulation headers in flight, but thanks to advancement of eBPF this logic can be completely implemented outside of the kernel in a less intrusive way and with all of the benefits that come with eBPF.

In this presentation I'm going to talk about how eBPF was used to encapsulate packets using the eBPF TC filter and the cgroup hooks, discuss what the differences are between this approach and LWT, explain how this feature was easily extended to support a more interesting feature: "encapsulation headers reflection" which is used to store the encapsulation headers of incoming traffic and reflect them on the responses making it transparent for the application. During the talk I'm also going to discuss the pain points found during the implementation which lead us to non obvious solutions.

The goal is to have an open discussion about how the problem was solved and the obstacles faced and highlight possible eBPF/kernel features that would have been nice to have i.e BPF_MAP_TYPE_NS_STORAGE (namespace storage).

I agree to abide by the anti-harassment policy

I agree

Primary authors: VAZQUEZ, Brian (Google); LI, Coco (Google); FOMICHEV, Stanislav (Google); DE BRUIJN, Willem (Google)

Presenters: VAZQUEZ, Brian (Google); LI, Coco (Google); FOMICHEV, Stanislav (Google); DE BRUIJN, Willem (Google)

Session Classification: BPF & Networking Summit

Track Classification: Networking & BPF Summit (Closed)