Contribution ID: **136** Type: **not specified**

Socket migration for SO_REUSEPORT

Monday, September 20, 2021 8:40 AM (40 minutes)

This talk presents our recent work available in the v5.14 kernel, which improves the SO_REUSEPORT functionality.

The SO_REUSEPORT option was introduced in v3.9. In the former version, only one socket is allowed to listen() on any given TCP port. The traditional technique for a high-performance server is to have a single process that accept()s and distributes connections to other processes or to have multiple processes that accept() connections from the same single socket. However, the accept() syscalls to a single listen()ing socket can be a bottleneck. The SO_REUSEPORT option allows multiple sockets to listen() on the same port and addresses the bottleneck.

If the option is enabled, the kernel distributes connections evenly to each listen()ing socket when SYN packets arrive. Once the kernel has committed a connection to a listen()ing socket, it does not change later. Thus, when a listen()ing socket is close()d, the not yet accept()ed connections are aborted even if other sockets still listen() on the same port.

This talk shows how the SO_REUSEPORT mechanism works with SYN processing, when it causes connection failures, how we can work around it with BPF, and how we address it with the new socket migration feature and the extension of BPF.

I agree to abide by the anti-harassment policy

I agree

Primary author: IWASHIMA, Kuniyuki (Amazon)

Presenter: IWASHIMA, Kuniyuki (Amazon)

Session Classification: BPF & Networking Summit

Track Classification: Networking & BPF Summit (Closed)