Contribution ID: **460** Type: **not specified**

The case for zero-copy in containers: Accelerating KubeVirt and enabling memory providers

Saturday 13 December 2025 15:40 (40 minutes)

For building high-performance datapaths, zero-copy mechanisms are inevitable. Up until today, their usability from network namespaces is either non-existent or very slow. In this talk, we present a solution to natively "lease"a physical NIC's hardware queue to a virtual device (such as netkit) in order to enable applications in containers/Pods to fully utilize io_uring or devmem zero-copy for TCP as well as AF_XDP. While the former two are needed to significantly accelerate applications inside Pods, the latter use-case is targeted at KubeVirt Pods to launch QEMU/KVM instances backed by AF_XDP. While both mechanisms are tackling different use-cases, the NIC queue leasing approach presents itself as a common underlying solution. We detail the design, API extensions, and implementation details for the networking core as well as virtual drivers to support both technologies.

For accelerating KubeVirt with AF_XDP, the work does not yet stop there given we also need a new AF_XDP TX hook for policy enforcement. We therefore propose a revamp of the XDP API to better address future needs. The work includes a conversion to bpf_mprog to enable multi-attach capabilities (first introduced via tcx) along with support for a queue-range attachment of BPF programs, the TX-side attachment itself and other extensions needed. We conclude with a discussion of how these changes impact the current / legacy API.

Primary authors: PROTOPOPOV, Anton (Isovalent); BORKMANN, Daniel (Isovalent); WEI, David (Meta)

Presenters: PROTOPOPOV, Anton (Isovalent); BORKMANN, Daniel (Isovalent); WEI, David (Meta)

Session Classification: Networking Track

Track Classification: Networking Track