Contribution ID: **155** Type: **not specified**

Optimizing UDP for Content Delivery with GSO, Pacing and Zerocopy

Wednesday, November 14, 2018 11:00 AM (35 minutes)

UDP is a popular foundation for new protocols. It is available across operating systems without superuser privileges and widely supported by middleboxes. Shipping protocols in userspace on top of a robust UDP stack allows for rapid deployment, experimentation and innovation of network protocols.

But implementing protocols in userspace has limitations. The environment lacks access to features like high resolution timers and hardware offload. Transport cost can be high. Cycle count of transferring large payloads with UDP can be up to 3x that of TCP.

In this talk we present recent and ongoing work, both by the authors and others, at improving UDP for content delivery.

UDP Segmentation offload amortizes transmit stack traversal by sending as many as 64 segments as one large fused large packet. The kernel passes this through the stack as one datagram, then splits it into multiple packets and replicates their network and transport headers just before handing to the network device.

Some devices can offload segmentation for exact multiples of segment size. We discuss how partial GSO support combines the best of software and hardware offload and evaluate the benefits of segmentation offload over standard UDP.

With these large buffers, MSG_ZEROCOPY becomes effective at removing the cost of copying in sendmsg, often the largest single line item in these workloads. We extend this to UDP and evaluate it on top of GSO.

Bursting too many segments at once can cause drops and retransmits. SO_TXTIME adds a release time interface which allows offloading of pacing to the kernel, where it is both more accurate and cheaper. We will look at this interface and how it is supported by queuing disciplines and hardware devices.

Finally, we look at how these transmit savings can be extended to the forwarding and receive paths through the complement of GSO, GRO, and local delivery of fused packets.

Presenter: DE BRUIJN, Willem (Google)

Session Classification: Networking Track