Packet-Level Telemetry in Large Datacenter Networks

SIGCOMM 2015 Public Review

Geoffrey M. Voelker
University of California, San Diego

Data center networks are both complex and subject to high-performance application demands. Inevitably problems arise where such complexity impacts performance and correctness and leads to problems that are difficult to track down. To troubleshoot these kinds of problems, Everflow provides "packet-level network telemetry": the ability to passively and actively measure data center nework behavior at the granularity of individual packets. Passive measurement enables continuous monitoring and detection of background issues (e.g., performance or forwarding anomalies) and a historical record when a problem is encountered. Active measurement enables focused exploration of issues or confirmation of causes. As a platform, Everflow efficiently implements its passive and active measurement techniques. On this platform Everflow supports applications that both use captured data to troubleshoot problems, as well as control which packets to passively capture and actively inject.

Everflow builds upon a long line of work developing troubleshooting techniques to explain and identify root causes of mysterious network behavior. For instance, recent previous work, such as NetSight, argued for the need to monitor at the packet granularity. The authors' own work with Duet showed how to take advantage of existing switch hardware to efficiently reshuffle network traffic at line rates. And systems like OFRewind demonstrated the value of replaying traffic to reproduce problems.

What makes Everflow interesting to the SIGCOMM community is how it extends these directions. How do you monitor every packet in the network? Everflow's answer is that you don't. The system does monitor all control packets (e.g., BGP, RDMA), but only key packets for all data flows (TCP SYN, FIN, RST) by default. When operators encounter unusual behavior, at any time they can have Everflow install filtering rules that then capture all packets of flows of interest. In addition, Everflow employs Duet to reshuffle monitored packets to both balance the load of capturing packets as well as to ensure capture consistency (e.g., packets from the same flow always go to the same capture server). Finally, Everflow supports packet replay, but crucially extends this capability to enable operators to inject any kind of packet anywhere in the network.

Why read this paper? In addition to the techniques and methodology used to scale packet-level tracing to the data center, two other aspects also make the paper a very interesting read. The first is the guided probing technique, which effectively turns the entire data center network into a precise and fine-grained active measurement and debugging tool. At any location in the network, operators can craft arbitrary packets, inject them, and monitor how they behave as they flow through the network.

The second is Section 7, an in-depth case study and experiences section. The authors use Everflow to explore and solve a variety of different challenging problems, from uncovering loops that nominally should not exist to attributing the root cause of errors to cutting-edge NICs. These experiences convincingly convey the daunting complexity of troubleshooting data center networks in practice, and the value of measurement platforms like Everflow in investigating mysterious behavior and ultimately proving that the proverbial network butler indeed did it.

Of course, this paper will not be the last on data center network measurement. Indeed, the authors themselves are investigating how to extend the system to troubleshoot problems across data centers. Advances in hardware in commodity switches, such as timestamping functionality that the authors already anticipate, could potentially improve accuracy and reduce system overhead further. And an interesting open question is how systems like Everflow can be applied and extended to troubleshoot problems that span the path from data center networks, across Internet ISPs, to clients.