Simplifying Datacenter Network Debugging with PathDump

Praveen Tammana†, Rachit Agarwal‡, Myungjin Lee†
†School of Informatics - University of Edinburgh, ‡Computer Science - Cornell University

Open source available at: https://github.com/PathDump/

Datacenter networks are complex
- Increasingly larger scale
- Over 100,000 servers, over 10,000 switches
- Each server with 10 to 40 Gbps NIC
- High utilization: > 100 Tbps aggregate traffic

Complexity due to the need for
- High availability
- High performance

Latency matters. Amazon found every 100ms of latency cost them 1% in sales. Google found an extra .5 seconds in search page generation time dropped traffic by 20%. A broker could lose $4 million in revenues per millisecond if their electronic trading platform is 5 milliseconds behind the competition.

Network debuggers are even more complex
- Increasingly, network debugging functionality is pushed into the network due to improved switch programmability
- Existing tools employ complex in-network techniques such as data plane snapshot, per-switch per-packet logs, packet mirroring/sampling, and dynamic rule updates

PathDump: A minimalistic network debugger
- Partition debugging functionality between switches and end-hosts
  - In-network tools focus on a smaller set of problems
  - Keeping networks and debugging tools as simple as possible
  - Before forwarding a packet, check a condition
  - If met, embeds its ID into packet header

Example: Load imbalance diagnosis
- Equal-Cost Multi-Path (ECMP) forwarding
  - Popular network load-balancing scheme in DCN

Conclusion and future work
- DCNs are complex; and their debuggers are even more complex
- Carefully partitions debugging functionality between network switches and edge devices
- Requires no complex operations from network switches
- Debugs a large class of network problems
- Future work: Real-time network debugging