Vertigo: A Priority-aware Burst-tolerant Datacenter Fabric

JOHNS HOPKINS UNIVERSITY

Sepehr Abdous, Erfan Sharafzadeh, Sougol Gheissi, Soudeh Ghorbani

1. Datacenters: A critical infrastructure in contemporary networks

- **Datacenters** host many applications we run today.
- **Cisco's study:** datacenter traffic represents 95% of global traffic.
- **Goal:** Low latency + High throughput.

2. Datacenter traffic is bursty

- **Bursts can be as short as 10s of micro-seconds (micro-bursts):** Despite datacenters low average utilization, μbursts increase queueing time, impose packet drops, hurt latency, and violate SLA.
- **Managing μbursts is challenging** due to characteristics like micro-second granularity and vast range of root causes.

3. Packet deflection to the rescue

What if the packets are deflected to a randomly chosen neghboring switch instead of being dropped during congestion?

- **Intuition:** By preventing packet drops, the consequences of μbursts can be mitigated using the spare capacity of the network.
- **Challenges:**
  - Excessive packet reordering.
  - Higher cost per packet drop.
  - Increased round-trip time (RTT).
  - Performance degradation under load.
- **Common solutions to some of the challenges:**
  - Disabling fast retransmission to avoid the consequences of reordering [3, 2].
  - Using powerful congestion control techniques, like DCTCP [1], to avoid packet drops.

Contact information → Email: **sabdous1@jh.edu**

4. Disabling fast retransmission increases the cost of packet drop

Avoiding the loss of throughput by disabling fast retransmission **increases the average delay imposed by packet drops** as every drop triggers a retransmission timeout (Figure 1).

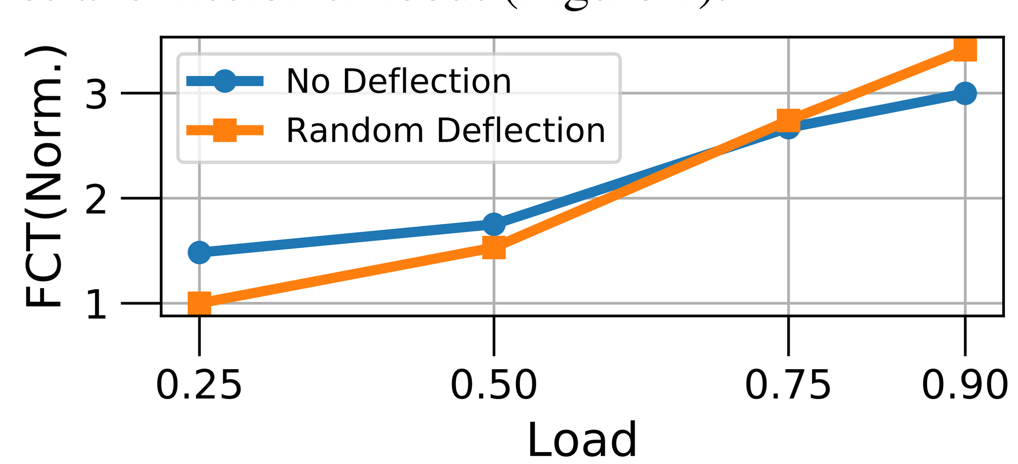

Figure 1: Under 90% load, using random deflection increases the Flow Completion Time (FCT) by 6% due to excessive packet drop.

5. Random deflection hurts the latency of the mice flows ($\leq 100KB$)

While deflection helps large flows by preventing packet drops, it hurts the FCT of the mice flows by increasing the queue occupancy (Figure 2).

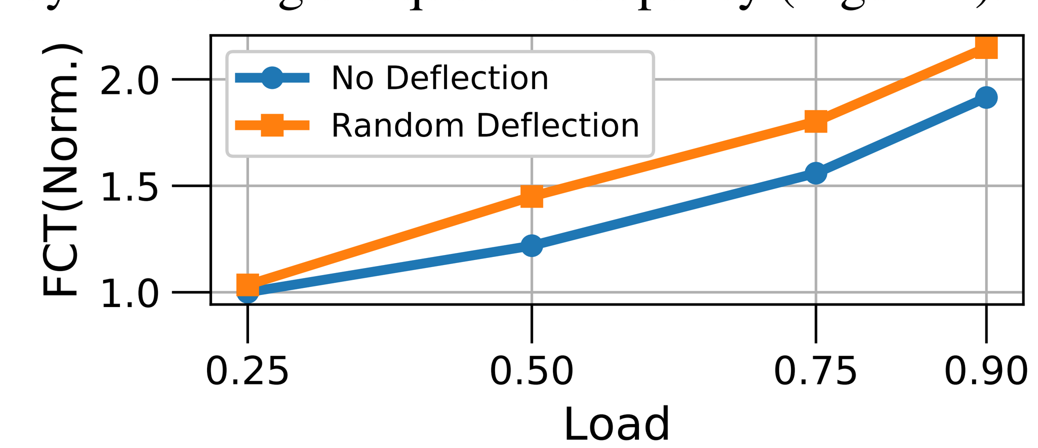

Figure 2: Under 90% load, the FCT of the mice flows increases by 12% due to deflection.

6. Vertigo: Priority-aware deflection

- **Main objective:** Deflect and, in cases of extreme congestion, drop the packets that impose the least latency.
- **Edge of the network** tags packets based on their position inside the flows which enables the core to differentiate between packets of small and large flows and react to them differently.
- **Core of the network** assigns priorities to the packets based on their tags and forwards, deflects and drops packets with regards to their priorities. Vertigo assigns lower priorities to the packets of large flows that are likely contributing to the heavy load and long-lasting congestion. Accordingly, the delays of queueing, deflection, and packet drop are imposed to low-priority packets.

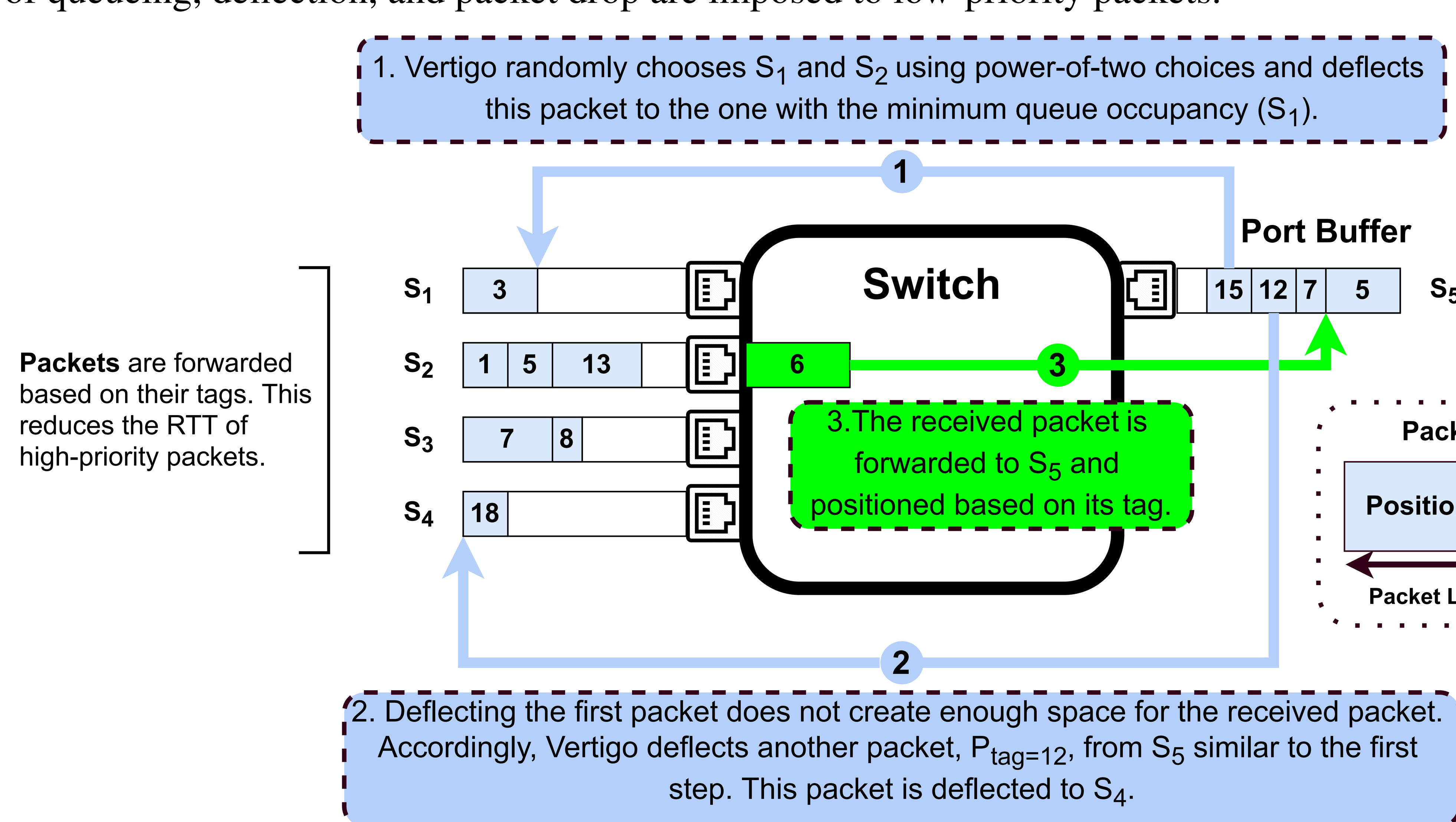

Figure 3: An example of how Vertigo performs when a packet with $tag = 6$ arrives and should normally be forwarded to S_5.

7. Initial Results

Vertigo performs better than random deflection while facing packet drops.

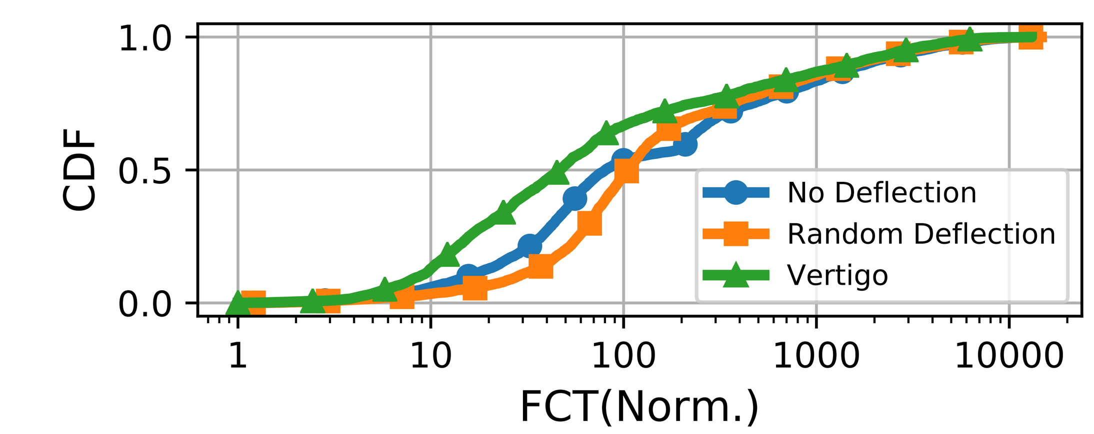

Figure 4: Under 90% load, Vertigo achieves 19% less mean FCT than random packet deflection (DIBS).

Vertigo keeps the mean FCT of mice flows low.

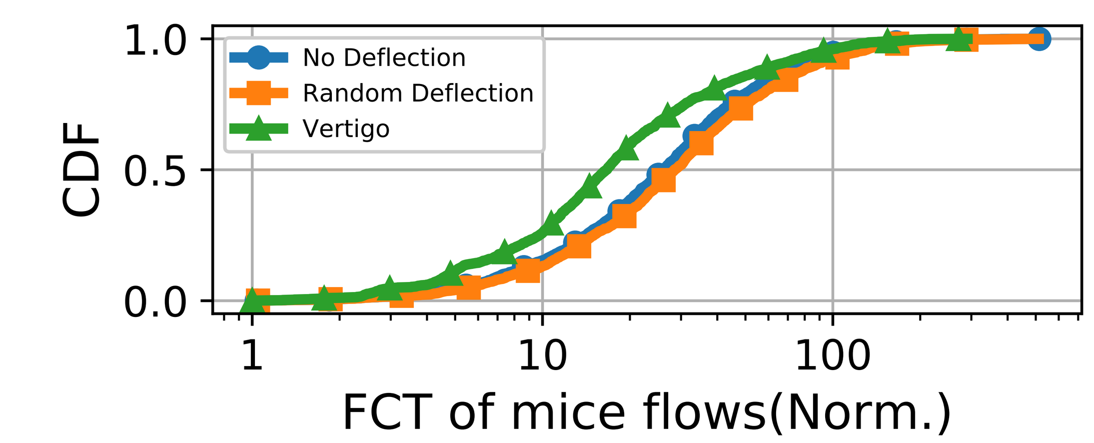

Figure 5: Under 90% load, Vertigo achieves 34% lower mean FCT for mice flows than random deflection.

8. Ongoing and Future work

We are trying to mitigate the reordering while keeping fast retransmission benefits. **Check out our other poster: "Valinor".**

We are evaluating Vertigo in large-scale datacenter topology against various common workloads using network simulations. Additionally, we plan to implement Vertigo on commodity programmable fabric.

References

[1] Mohammad Alizadeh, Albert Greenberg, David A. Maltz, Jitendra Padhye, Parveen Patel, Balaji Prabhakar, Sudipta Sengupta, and Murari Sridharan. Data center tcp (dctcp). SIGCOMM '10, 2010.

[2] X. Shi, L. Wang, F. Zhang, K. Zheng, and Z. Liu. Pabo: Congestion mitigation via packet bounce. IEEE ICC '17, 2017.

[3] Kyriakos Zarifis, Rui Miao, Matt Calder, Ethan Katz-Bassett, Minlan Yu, and Jitendra Padhye. Dibs: Just-in-time congestion mitigation for data centers. EuroSys '14, 2014.