Motivation

Multipath TCP and Optical Switching in Data Centers

Multipath TCP and optical switching are respectively explored recently. Both are emerging technologies and represent the future trend.

- Multipath TCP provides tradeoff between reliability and utilization. But it requires additional modification of architecture and topology.
- Optical Switching allows flexible strategies on topology control and demand response, which can get higher bandwidth and efficiency. But its reconfiguration delays will lead to high latency.

Benefits when multipath TCP and optical switching are combined:

- Optical switching enables multiplexing technology which provides preconditions of architecture and topology for multipath TCP.
- Multipath TCP has lower flow completion time which reduces the adverse effect of high latency in optical switching.

Challenges

In spite of all the benefits, both multipath TCP and optical switching will make it more complex and challenging for data center networking.

Topology: relay nodes, extra overhead, efficient throughput

- BCube and DCell need relay servers to support multipath, but it makes capacity scheduling more complicated.
- In Fattree and VL2 that do not need relay nodes, multipath leads to extra overhead such as SYN and coding headers.
- The extra overhead leads to the decrease of efficient throughput, though the flow completion time is reduced.

Failures Handling: performance, robustness, reliability

- When failures occur, the performance and robustness are reduced considering SYN&coding headers and flow drops&retransmission.
- Failures in nodes (including servers and switches) and links result in poor reliability compared to single path forwarding.

Challenges

Congestion Control

- Apart from node and link failures, network congestion occurs more frequently in multipath, especially when the traffic loads are heavy.

Resource Allocation

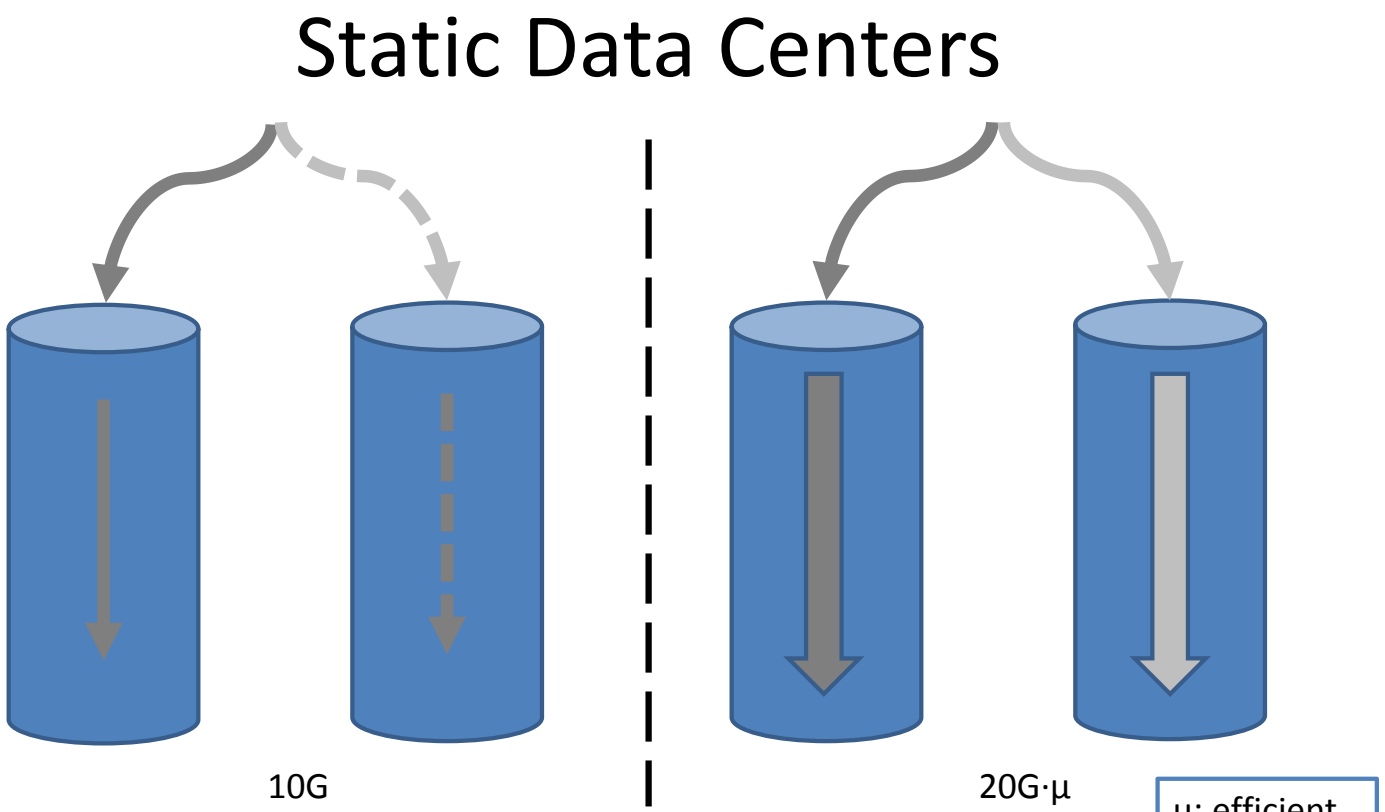
- Both multipath and flexible switching make it more complicated for data center networking, like addressing, routing, scheduling, etc.

Therefore, link failures (including network congestion) of multipath TCP have significant influence on the reliability and efficiency of optical data centers. The followings can be introduced to address these problems.

- Flow Backup: encoding, synchronization, retransmission
- Block ACK: congestion control, failure detection, state feedback

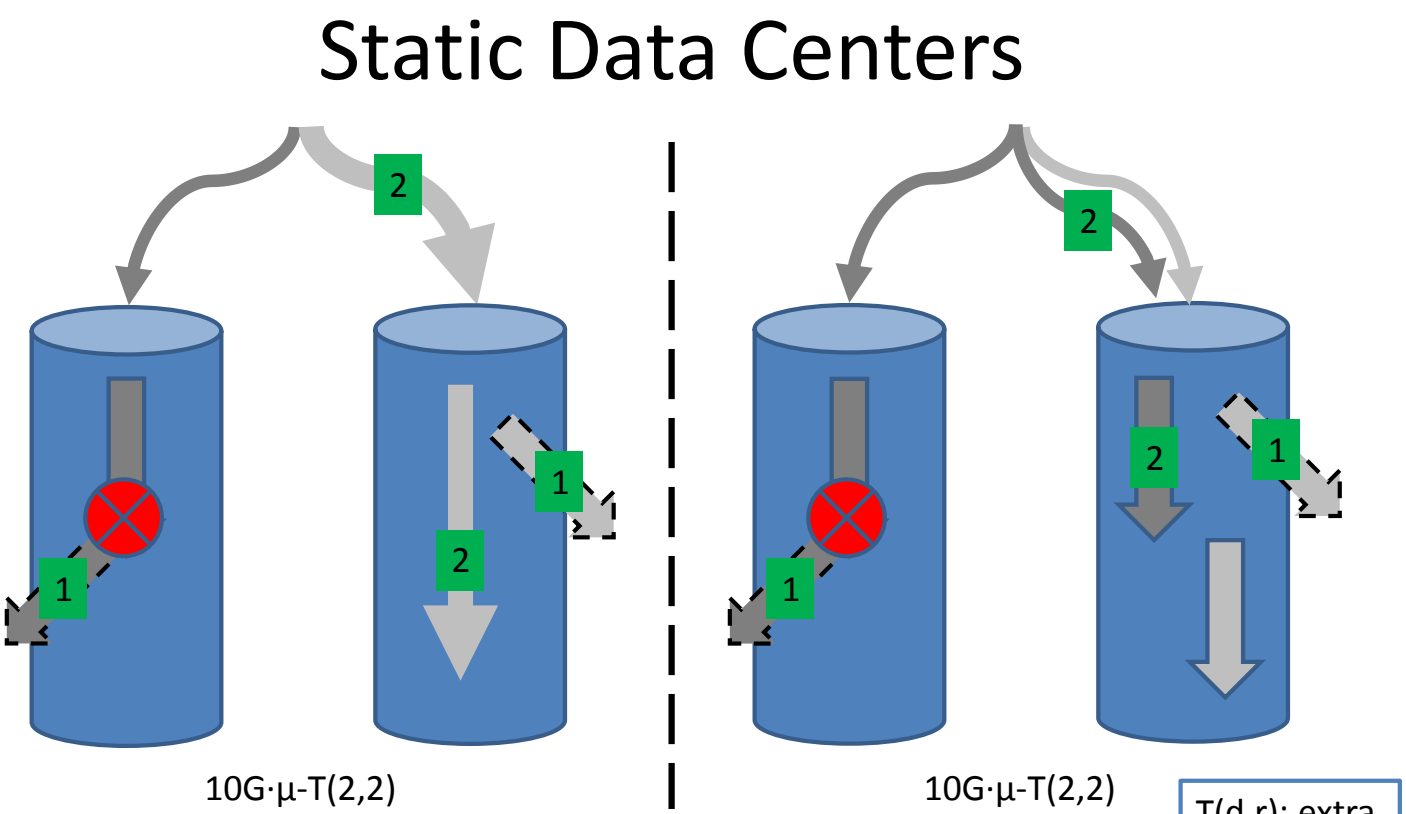
Multipath TCP in Data Centers

Trade-off between reliability and utilization



Multipath TCP in Data Centers

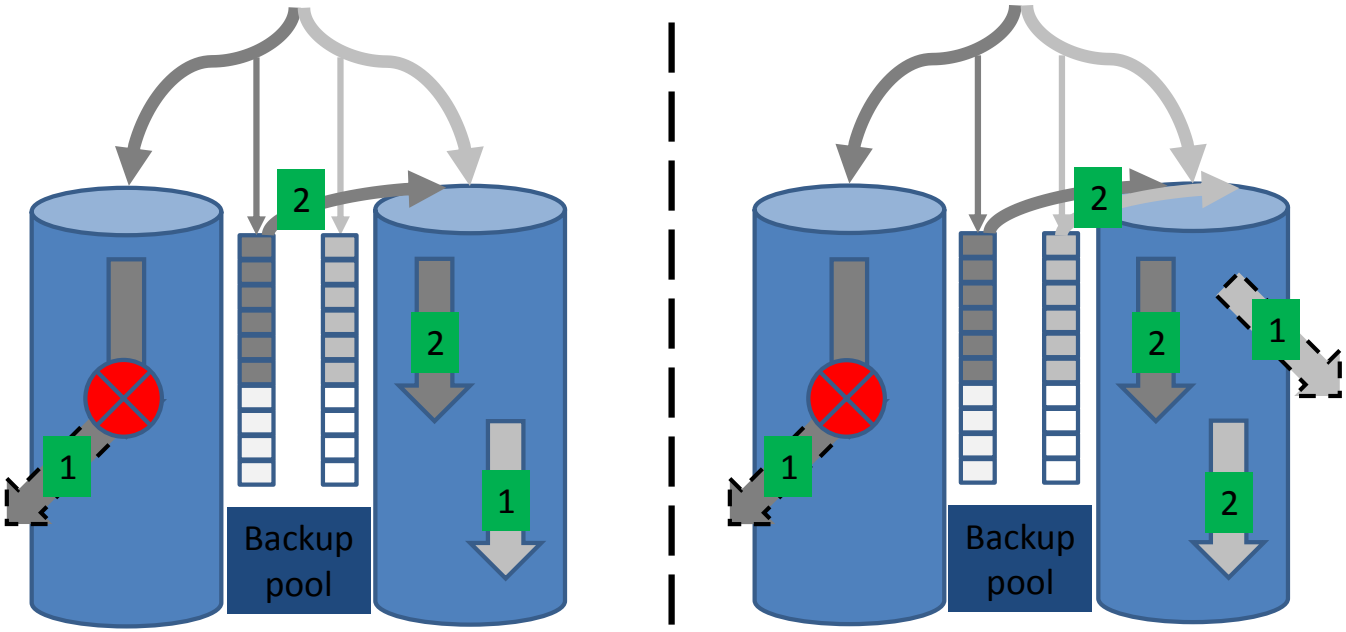
Low reliability due to drops and retransmission



Failures Handling in Data Centers

High reliability through flows backup

Static Data Centers



10G·μ-T(1,1)

Multi-path failures by backup:

1. Drop one subflow;
2. Retransmit one subflow;
3. Decoding after transmit.

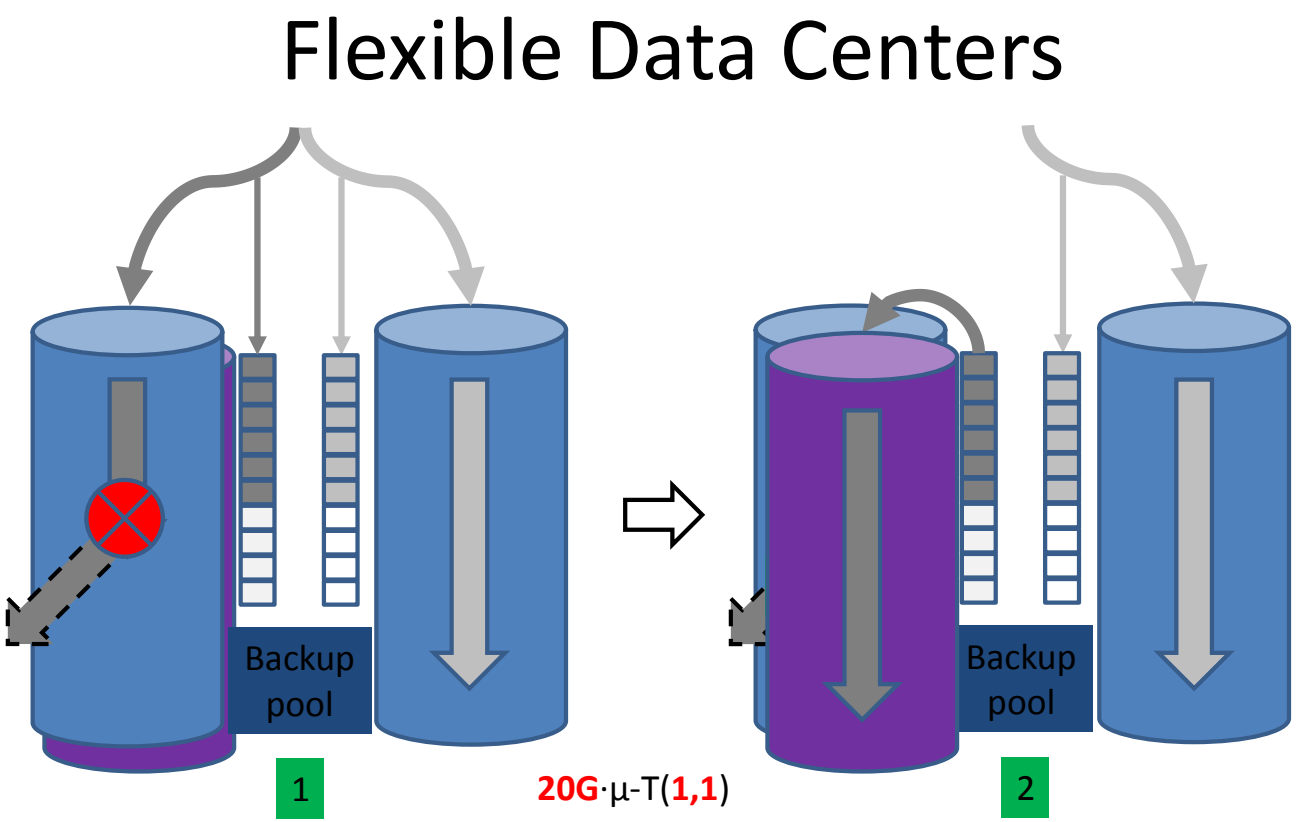
10G·μ-T(2,2)

Multi-path failures by backup:

1. Drop two subflows;
2. Retransmit two subflows;
3. Decoding **as transmitting**.

Failures Handling in Data Centers

High efficiency through flexible switching



Multi-path failures by backup:

1. Drop **one subflow**;
2. Retransmit **one subflow**;
3. Decoding **as transmitting**.

Multipath TCP and Failures Handling

Basic Procedures

Topology & Congestion Control, Packet Scheduling & Decoding

1. Split flow(s) into subflows and push into backup pool
2. Congestion control by length of **flow backup** or **block ACK**
3. Topology control and scheduling by **flexible switching**
4. Add premix to subflows, make backup and transmit subflows
5. Decoding according to premix or flow backup

- Flow backup: copy the transmitting subflows and premix
- Block ACK: several backup subflows use one combined ACK
- Flexible switching: on-demand response by block ACK and backup

Multipath TCP and Failures Handling

Adopted Technologies

Failures handling technologies

- **Flow backup** reduces the overhead of drops and retransmission: T(2,2) $\rightarrow$ T(1,1).
- **Flexible switching** improves the capacity and reduces the delay: 10G $\rightarrow$ 20G.

The above failures handling methods improve the decoding efficiency: after transmit $\rightarrow$ as transmitting.

Multipath TCP technologies

- **Block ACK** reduces extra overhead: ACK/subflow $\rightarrow$ ACK/block, and gives state feedback for failure detection and path selection.

The multipath routing and scheduling of optical data centers have not been explored, especially on flexible switching (topology control) and failures handling (including congestion control).