Packet reordering in multi-path transport scenarios

Nathalie Romo Moreno

LIFE IS FOR SHARING.

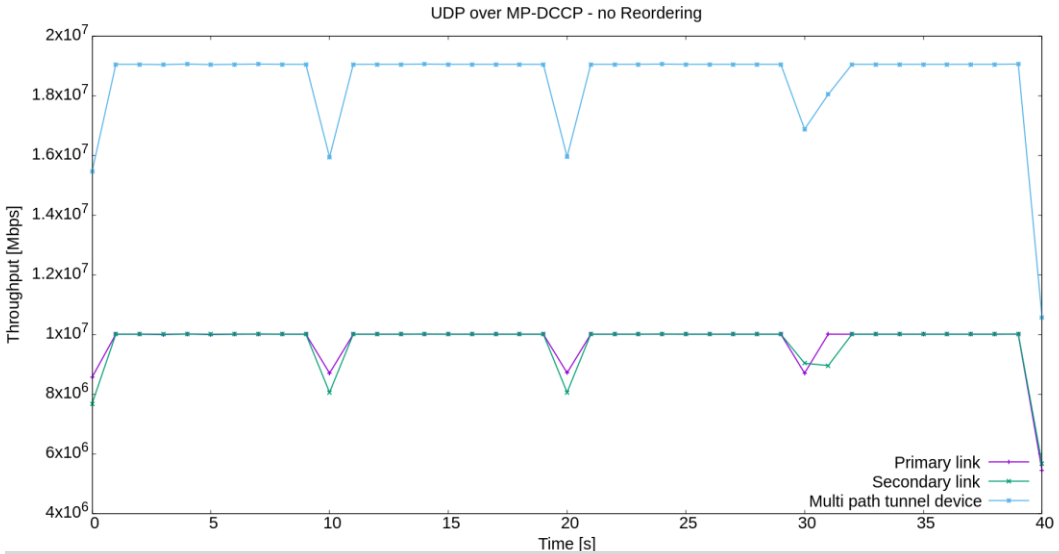
Introduction

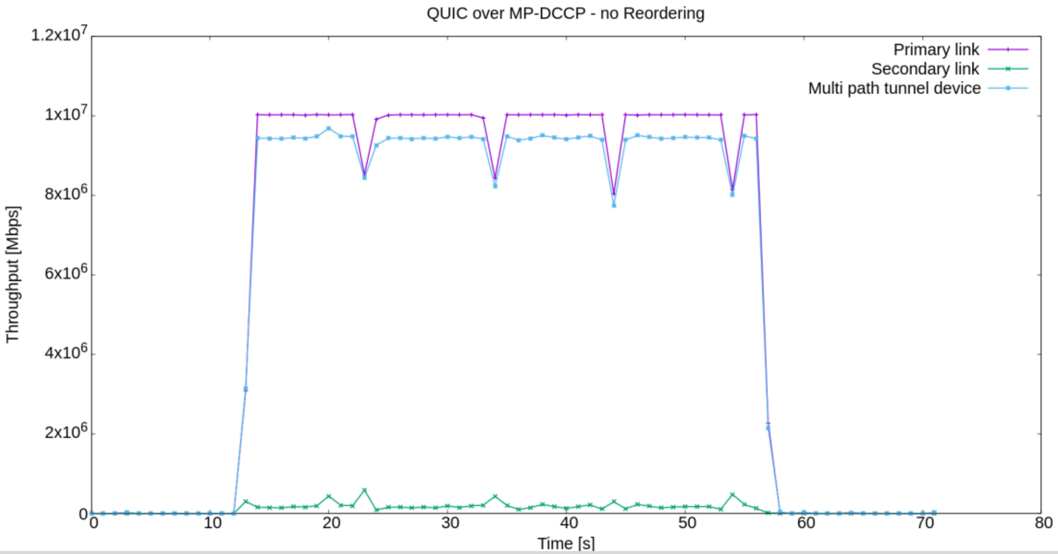
- **Multipath (MP) transport compared to single-path transport experiences heterogenous transport characteristics, with specifically extraordinary latency/jitter and significant out-of-order delivery**
- **MP protocols exposes this characteristics to higher layer application or in case of intermediary MP transport (MNO bundled 5G + Wi-Fi aka 3GPP ATSSS) to carried e2e traffic.**
  - MPTCP/MP-QUIC: Jitter dominated by path latency difference
  - MP-DCCP/MP-QUIC+DATAGRAM/CMT-SCTP+PR: Jitter and out-of-order delivery dominated by path latency difference
- Services running over the Internet are designed to cope with the characteristics of single path transport, thus, the changes introduced , e.g., within the ATSSS, **might result in QoE and QoS degradation**

Conducted work to investigate demand for MP reordering and latency difference compensation in an intermediary MP transport scenario

- Demonstrate the impact of no reordering with different traffic types carried over a multipath transport
- Evaluate different solutions to correct out of order delivery and jitter -> For further detail on the algorithms tested refer to https://datatracker.ietf.org/doc/html/draft-amend-iccrg-multipath-reordering-03
- Tests were executed using the MP-DCCP framework (https://github.com/telekom/mp-dccp/), but results are applicable to other multipath solutions, e.g., MP-QUIC
- TCP and UDP traffic generated with iperf
- QUIC traffic generated using the quic-go implementation available at https://github.com/lucas-clemente/quic-go

No Reordering – Carried traffic: UDP vs QUIC





Settings: Path latency difference: 15ms | MP-DCCP subflows CC : CCID5 (BBR) | QUIC traffic CC: New Reno | Priority based steering mode | Per path Bw: 10Mbps

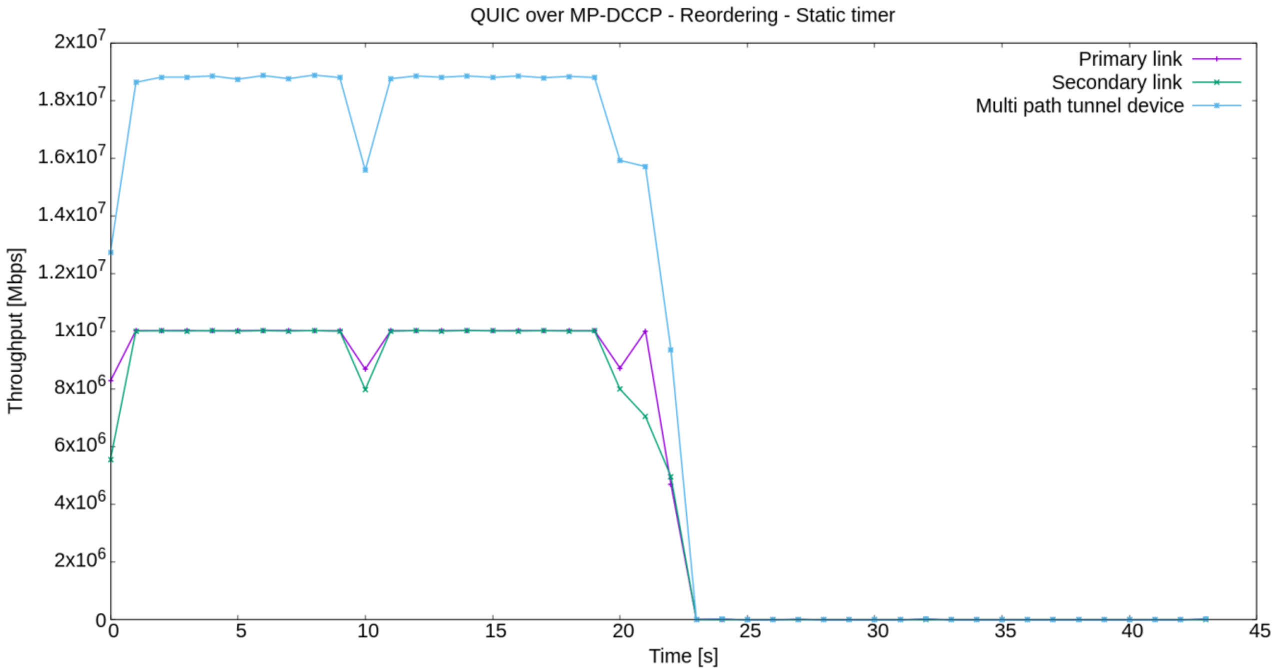
Carried plain UDP traffic -> Full aggregated Bandwidth utilized

Carried QUIC traffic -> Single path Bandwidth utilized

- Out of order packets cause DUPACK -> packets marked as loss
- Congestion control and reliability mechanisms present in the QUIC traffic react decreasing cwnd and triggering retransmission

- **Unlike plane UDP, QUIC traffic has some demand on in order deliver, and therefore fails to use the aggregated bandwidth due to the impact of the packet scrambling**

Reordering with Static timer – Carried traffic: QUIC



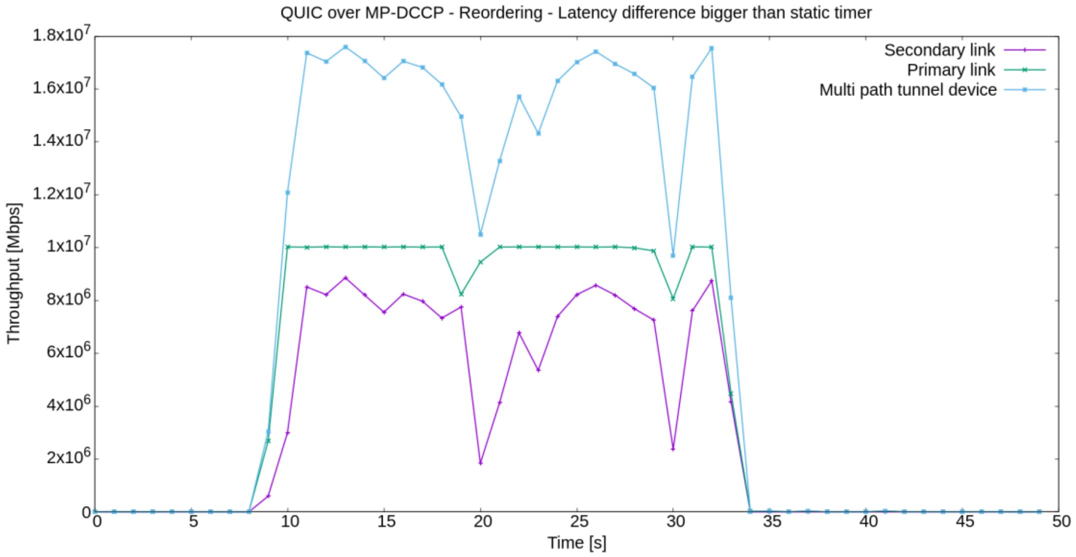
Reordering algorithm:

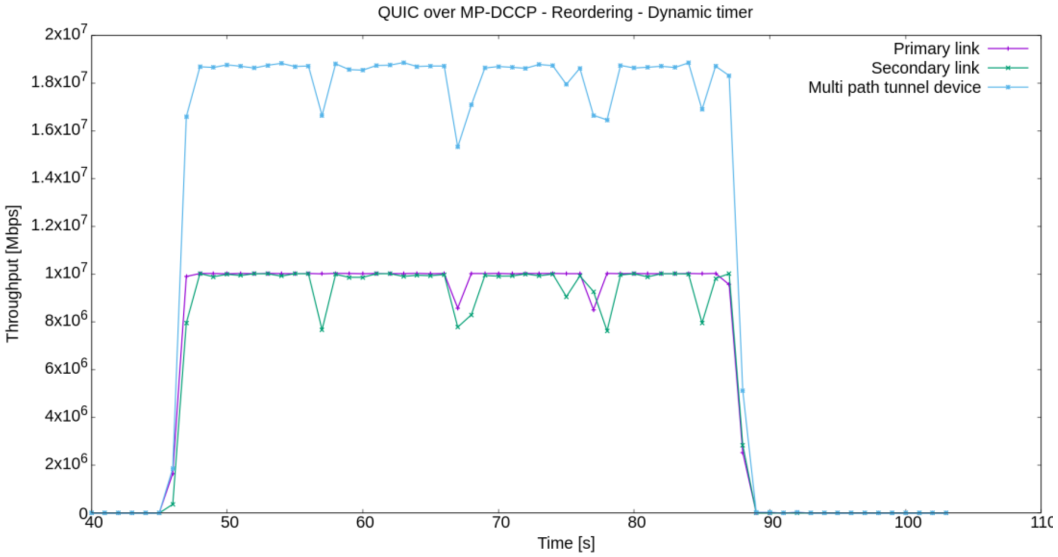
- Connection sequence numbers are used to verify in order arrival. When a gap is detected, a buffer is used to store received packets until the missing one(s) arrive, or a fixed timer expires.

Settings: Path latency difference: 15ms | MP-DCCP subflows CC : CCID5 (BBR) | QUIC traffic CC: New Reno | Priority based steering mode | Per path Bw: 10Mbps | | Timer: 50ms

- **Scrambling correction eliminates DUPACKs and in consequence the reaction of the carried QUIC Congestion Control .**
- **As a result, the application manages to utilize the full bandwidth available in both paths**

Reordering - Static vs Dynamic timer



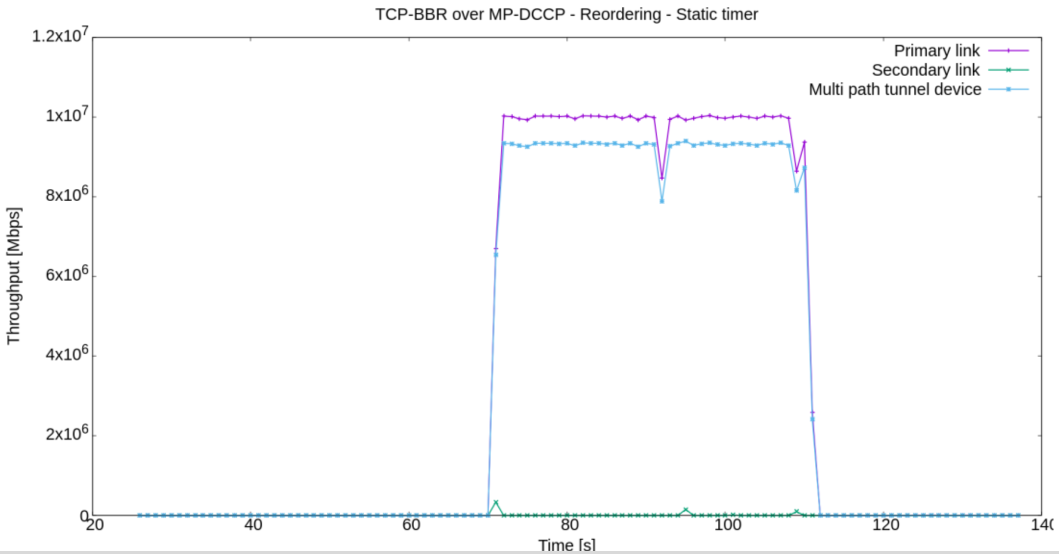


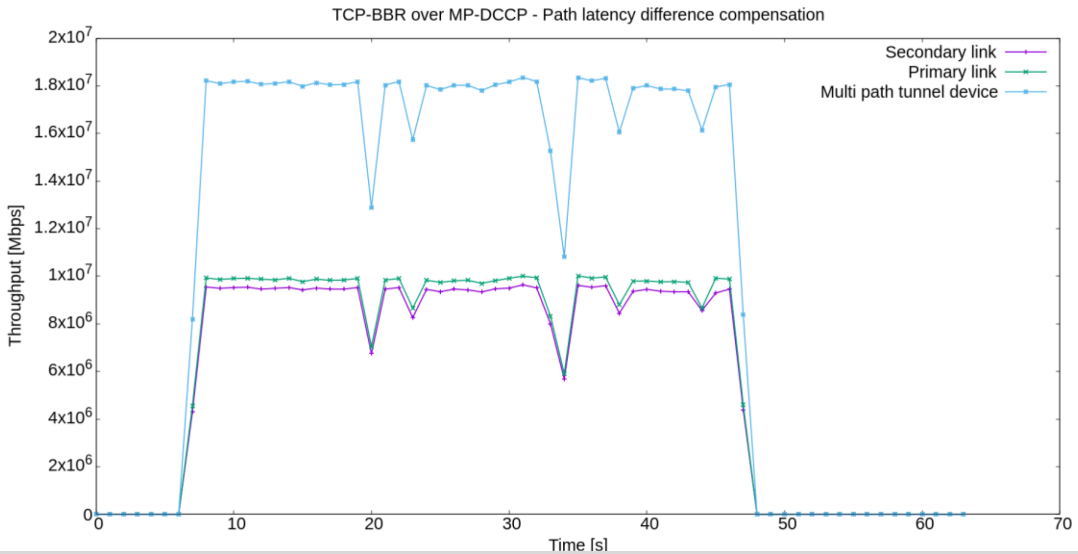
Settings: Path latency difference: 20ms | MP-DCCP subflows CC : CCID5 (BBR) | QUIC traffic CC: New Reno | Priority based steering mode | Per path Bw: 10Mbps | Timer*: 15ms

- Fixed expiration timer works well under static known conditions, however, changes in the network might cause the latency difference to go above or below the configured timer, leading to performance degradation or increment of the end-to-end latency

- The dynamic expiration timer used to buffer out of order packet is updated based on the latency difference of the paths estimated from the RTT measurements provided by the CC in place
- **As the timer adapts to the network conditions, the scrambling correction works good enough to achieve full aggregation**

Reordering Static vs Delay equalization





Settings: Path latency difference: 15ms | MP-DCCP subflows CC : CCID5 (BBR) | QUIC traffic CC: New Reno | Priority based steering mode | Per path Bw: 10Mbps | Timer*: 50ms

- Sequence number based re-ordering does not compensate for latency difference between paths, i.e, the receiving end observes latency changes whenever the steering mode alternates between both paths (jitter).
- **Latency sensitive CC algorithms like BBR react to the latency jumps by throttling throughput under the assumption of buffer bloat.**

- The delay equalization mechanism works by introducing delay to the incoming packets in the fastest path, until its latency equals the one corresponding to the slowest path
- It does not guarantee a total in order delivery, but reduces the reordering effort and **provides a smooth end-to end latency, improving aggregation performance in the case of traffic controlled by BBR**

Conclusion

- The impact on the performance varies depending on the characteristics of the carried traffic/service, and different solutions are suitable for different traffic types
- Within the ATSSS splitting scenario, the characteristics of the traffic carried over the multi path network might be unknow, therefore, certain in-network reordering mechanisms are required to guarantee an optimal performance
- 3GPP SA2 WG concluded in-network reordering support is required after contributing results in S2-2203965
- S2-2203965 document recommends to use a combination of sequence number based re-ordering with dynamic expiration timer , delay equalization and fast packet loss detection mechanisms to guarantee optimal performance when the characteristics of the carried traffic are unknow.

Reordering under IETF scope?

Reordering and latency difference compensation for MP protocols is not standardized, similar to scheduling

→ Both is so far out of scope of serious standardization in IETF* and therefore implementation specific

→ Without specification consistent behaviour between client and server in down and uplink is NOT ensured.

→ Scheduling and reordering are mainly agnostic to IETF protocols: MPTCP, MP-DCCP, MP-QUIC, CMT-SCTP.

Where is the habor in IETF to address multipath overarching topics such as reordering and scheduling?

* With [1], [2] individual informational drafts are available in ICCRG