On the hardware cost of end-to-end latency variation control for time-slotted optical networks

Mijail Szczerban [1][2], Abed-Elhak Kasbari[2], Achour Ouslimani[2], Sébastien Bigo[1], Nihel Benzaoui[1]

[1] Nokia Bell Labs, 7 Route de Villejust, Nozay 91120, France, mijail.szczerban_gonzalez@nokia.com

[2] Quartz Lab, ENSEA, 6 Avenue du Ponceau, 95000 Cergy, France

Abstract *We experimentally study the impact of latency variation (jitter) compensation mechanism (JCM) on the network hardware in time-slotted optical networks. We implement a JCM that enforces latency variation below 30 ns, and we experimentally found trade-offs between guaranteed latency value, reservation granularity and memory requirements.*

Introduction

Time-critical applications such as 5G fronthaul [1] and high-frequency trading [2], require end-to-end performance guarantees to operate correctly. These guarantees typically translate into bounded absolute latency, that can go as low as few tens of microseconds; and low latency variation (i.e. jitter) typically one order of magnitude below [3]-[5]. These applications call for a new form of networking which we refer to as deterministic networking, where low jitter is achieved by per-flow hardware network slicing and by using a jitter compensation mechanism [5]. However, time performance guarantees come at the expense of additional hardware and system complexity. In this work we assess the incurred hardware cost. We experimentally study the relation between minimum achievable latency, flow transmission period, resource reservation granularity and buffer memory utilization to guarantee quasi-constant latency in reservation-based time-slotted optical networks.

Hardware network slicing

To guarantee performance, we implement end-to-end, per flow and hardware-based network slicing, as shown in Fig. 1. This implies that each time-critical flow running on the network is provided with dedicated resources from source to destination. This includes network interfaces at source and destination nodes, per-flow transmission and reception queues; and reserved per-flow transmission time slots in the optical line. Thus, physical resources -electrical and optical- are guaranteed end-to-end, preventing undesired interaction between flows and making the performance of each flow dependent only on its own properties, disregarding network utilization. To ensure hard slicing, optical transparent data plane [6] or cut-through transit traffic forwarding [7] are required to prevent contention and variable delay at intermediate nodes. We rely on Time-division multiplexing (TDM) to allocate transmission resources periodically and implement hard network slicing in the optical line[5],[8]. Nonetheless, TDM also comes at the expense of latency variation. Incoming frames need to wait until an optical time slot becomes available for transmission, which results in flow latency variations.

Jitter compensation mechanism

To cope with latency variations intrinsic to TDM systems, we implement a jitter compensation mechanism (JCM) whereby each frame is time-stamped with nanosecond precision at the insertion node [5]. It allows for a precise per-frame estimation of the experienced latency at the destination node. The experienced latency is compared with the per-flow target latency (defined a priori) and the difference between the target and experienced latency is compensated by buffering the frame accordingly at a reception

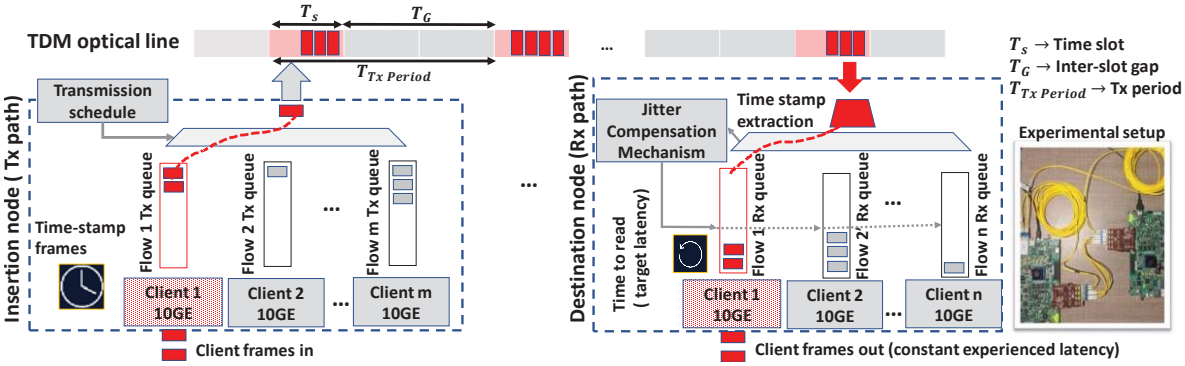

Fig. 1: Per-flow dedicated networking resources enabling per-flow network slicing. Experimental setup picture.

First In, First Out (FIFO) queue. To ensure constant latency at the reception client interface, the target latency is to be larger than the maximum delay that any frame belonging to the flow can experience within the network. The end-to-end hard network slicing that we implemented guarantees a fixed worst case delay, since information transmission is ensured per-flow, and flows are buffered independently. Nonetheless, this latency determinism comes at the expense of additional buffer memory needed for jitter compensation. It also requires modifications of the control plane for per-flow resource allocation. We proved in previous work that edge-networks implementing per-flow hard slicing can reconfigure in tens of microseconds when real-time control plane strategies are applied [9].

Experimental evaluation

To experimentally identify the trade-offs between memory requirements, target latency and TDM parameters (i.e. optical transmission period and time slot duration), we built an experimental testbed emulating the flow journey throughout CBOSS all-optical intra data centre network [6]. We used FPGA Xilinx Kintex UltraScale KCU105 boards to implement data plane features, including 10G Ethernet client interfaces (source and destination client), with integrated constant bit rate flow generator and real-time performance monitoring, TDM functionalities to encapsulate and schedule the transmission of client information into the optical line, as well as the real-time JCM.

In order to evaluate the transmitter (Tx) and receiver (Rx) queue memory utilization, we varied two parameters. First, the transmitter period ($T_{Tx\ Period}$), between consecutive time slots (of duration T_s) reserved for the flow under study, as shown in Fig. 1. To maintain the reserved transmission throughput, each time $T_{Tx\ Period}$ was modified, T_s and the inter-time slot gap (T_G) were also modified proportionally. Frames arriving to the Tx buffer just after a transmission time slot need to wait T_G before being sent, in consequence, it should be expected that the worst-case latency and Tx buffer utilization grow with T_G. The second parameter that was varied is the flow target latency (T_{Target}). We guarantee frame transmission by strictly allocating larger line capacity than the flow average data rate (D). To guarantee constant latency, we ensure $T_{Target} \geq T_G$, for the correct operation of the JCM. Note that the propagation delay is excluded for the sake of simplicity, and hard network slicing ensures constant propagation delay as in [6, 7]. The above-mentioned conditions were assumed to induce the following analytical equations.

Results

First, we evaluate the memory required for jitter compensation at the destination node. Fig. 2 shows the Rx FIFO utilization in (U_{Rx}) in kilobytes as we vary T_{Target} for different $T_{Tx\ Period}$ (and T_G) assigned to the flow under study. In theory, the maximum Rx FIFO utilization ($U_{Rx\ MAX}$) occurs when the RX FIFO stores all the client frames sent during T_{Target}, which is the maximum time any frame can be stored to compensate the jitter, thus there is direct relation between both, as predicted by Eq. 1 in Fig. 2. $U_{Rx\ MAX}$ increases with T_{Target} and is independent of $T_{Tx\ Period}$.

Regarding the average utilization ($U_{Rx\ AVG}$), it is inversely proportional to T_G. Indeed, the larger the delay the frames experience at Tx, the shorter the time in average they are buffered at Rx to meet the target latency (T_{Target}). Eq. 2 defines $U_{Rx\ AVG}$ assuming reserved line capacity just over D. Eq. 2a describes the evolution of $U_{Rx\ AVG}$ in the region of correct operation for the JCM, when $T_{Target} \geq T_G$, otherwise, Eq. 2b is valid, but in this region the JCM cannot compensate for the jitter of frames that have experienced latency larger than T_{Target}. From Eq. 2 we can deduce that average memory utilization at Rx can be adjusted by adapting T_G ($thus,\ T_{Tx\ Period}$).

As observed in Fig. 2, the analytical values obtained from Eq. 1, match the experimental results for $U_{Rx\ MAX}$. In the case of $U_{Rx\ AVG}$, the threshold of operation of the JCM (when $T_{Target} = T_G$), defines two regions, a linear region where the JCM can effectively compensate latency variations, and, a second region where only the frames that experience a shorter latency than T_{Target} are buffered at the reception. The experimental results obtained follow the analytical model given by Eq. 2a and Eq. 2b with an offset of typically one packet size.

Second, we experimentally investigated the relation between Tx and Rx FIFO utilization when

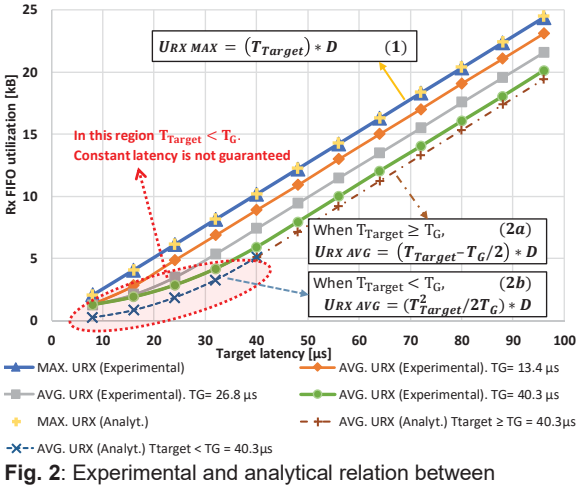

Fig. 2: Experimental and analytical relation between U_{Rx} and T_{Target} for different transmission period.

transmission parameters are varied. Fig. 3 shows Tx and Rx FIFO utilization in kilobytes for different $T_{Tx\ Period}$. In this case, the target latency was kept constant at 64 µs, thus, to guarantee constant latency $T_G < 64\mu s$. As described in Eq. 1, $U_{Rx\ MAX}$ does not depend on the transmission parameters, while $U_{Rx\ AVG}$ has an inverse relation with T_G (and $T_{Tx\ Period}$) as deduced from Eq. 2. Regarding the Tx FIFO side, the utilization (U_{Tx}) is directly related to T_G, since the Tx FIFO stores all frames until the next time slot is available.

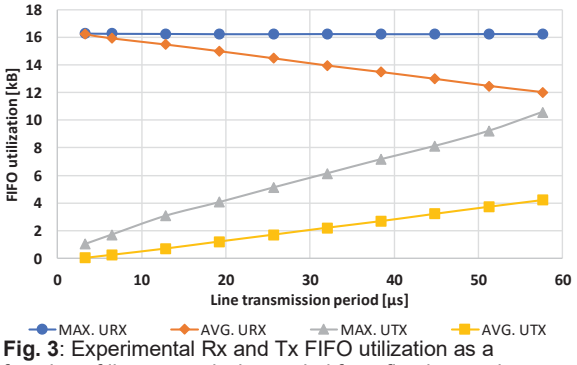

Fig. 3: Experimental Rx and Tx FIFO utilization as a function of line transmission period for a fixed target latency.

$$U_{Tx\ MAX} = (T_G) * D = (T_{Tx\ Period} - T_S) * D \quad (3)$$

It is important to note that $U_{Tx\ AVG}$ follows an inverse relation with respect to $U_{Rx\ AVG}$ when the transmission period is varied as shown in Fig. 3. Thus, in conditions where one endpoint is more stressed in terms of memory utilization than the other, we can redistribute the memory utilization by adjusting $T_{Tx\ Period}$. For example, having a fixed target latency, if we want to decrease $U_{Rx\ AVG}$, we could increase the line transmission period. In this example, $U_{Tx\ AVG}$ would increase while $U_{Rx\ AVG}$ decreases.

Third, we explored the relation between transmission reservation granularity, the achievable guaranteed latency and optical time slot utilization. Reducing the size of elementary reservable resources is advantageous for per-flow network slicing because it allows to better distribute the communication capacity among more flows. Increasing the number of time slots in a periodic reservation window (W), increases the resource reservation granularity; e.g. if the optical line runs at 10 Gb/s, a window of 10 time slots would enable an elementary reservation of 1Gb/s if only one periodic slot is reserved for the flow. Following the same logic, a window of 100 slots would enable an elementary reservation of 100 Mb/s. The window size increase allows to make reservations adapted for low throughput flows -with reduced capacity waste- by applying larger $T_{Tx\ Period}$, nonetheless, it also increases the worst-case latency for these flows. For this experiment, the throughput of the flow under study was 100 Mb/s and the time slot of 1.6 µs. We allocate one time slot per reservation window for the flow (elementary reservation). The target latency was adapted for each window size to account for the corresponding worst-case latency. Fig. 4 shows the experimental results: bars with maximum, average and minimum latency when no JCM is applied (Rx buffers not used), the evolution of the minimal guaranteed latency after compensating for the jitter (in yellow) and the U_{Rx} (light blue). As we increase W, $T_{Tx\ Period}$ also increases, thus, average and maximum latency grow. The JCM can equalize the latencies, but as we deduced from Eq. 1, the increase of the target latency (slightly larger than W in this experiment) leads to larger memory use. The guaranteed latency is measured when JCM is applied (in yellow), latency variations after JCM are below 30 ns. We observe that larger granularity, increases the slot utilization (efficiency) since the reserved capacity matches better the flow needs but the latency variation increases as well. If the goal is to provide low guaranteed latency, then the reservation W should be short (efficiency would be negatively affected). If the goal is to maximize utilization and flow count, then, the window size should be large (guaranteed latency would be larger, and memory needed at Rx FIFO would increase).

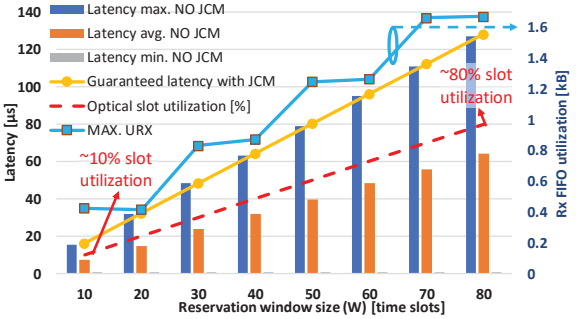

Fig. 4: Experimental evolution of latency and $U_{Rx\ max}$ against the transmission window size and its relationship with optical slot utilization efficiency.

Conclusions

We reported deterministic latency (jitter <30 ns) through per-flow hardware network slicing and by using JCM. Nonetheless, determinism comes at a cost. We found a directly proportional relation between the guaranteed target latency of the JCM and the RX FIFO memory requirements. Additionally, we found the inverse relation between the Tx and Rx FIFO average memory utilization when the transmission period is adjusted, thus, providing a way to redistribute memory usage among source and destination node. An increase of reservation granularity (number of fix-length time slots in a reservation window) increases the time slot utilization but also increases the achievable target latency.

References

[1] Y. Pointurier, N. Benzaoui, W. Lautenschlaeger and L. Dembeck, “End-to-End Time-Sensitive Optical Networking: Challenges and Solutions,” *Journal of Lightwave Technology,* vol. 37, no. 7, pp. 1732-1741, 2019.

[2] C. Moallemi and M. Sağlam, “The Cost of Latency in High-Frequency Trading,” *Operations Research,* vol. 61, no. 5, pp. 1069-1257, 25 April 2013.

[3] IETF RFC 8578, *Deterministic Networking Use Cases,* E. Grossman, Ed., 2019.

[4] 5G Americas White Paper, *New Services & Applications with 5G Ultra-Reliable Low Latency Communications,* 2018.

[5] N. Benzaoui, M. Szczerban, J. M. Estarán, H. Mardoyan, W. Lautenschlaeger, U. Gebhard, L. Dembeck, S. Bigo and Y. Pointurier, “Deterministic Dynamic Networks (DDN),” *Journal of Lightwave Technology,* vol. 37, no. 14, pp. 3465-3474, 15 July 2019.

[6] N. Benzaoui, J. M. Estarán, E. Dutisseuil, H. Mardoyan, G. D. Valicourt, A. Dupas, Q. P. Van, D. Verchere, B. Uščumlić, M. Szczerban, P. Dong, Y. Chen, S. Bigo and Y. Pointurier, “CBOSS: bringing traffic engineering inside data center networks,” *IEEE/OSA Journal of Optical Communications and Networking,* vol. 10, no. 7, pp. 117-125, 13 July 2018.

[7] W. Lautenschlaeger, L. Dembeck and U. Gebhard, “Prototyping Optical Ethernet—A Network for Distributed Data Centers in the Edge Cloud,” *IEEE/OSA Journal of Optical Communications and Networking,* vol. 10, no. 12, 27 December 2018.

[8] H. Uzawa, K. Honda, H. Nakamura, Y. Hirano, K.-i. Nakura, S. Kozaki and J. Terada, “Dynamic bandwidth allocation scheme for network-slicing-based TDM-PON toward the beyond-5G era,” *IEEE/OSA Journal of Optical Communications and Networking,* vol. 12, no. 2, pp. A135-A143, February 2020.

[9] M. Szczerban, N. Benzaoui, J. Estarán, H. Mardoyan, A. Ouslimani, A.-E. Kasbari, S. Bigo and Y. Pointurier, “Real-time Control and Management Plane for Edge-Cloud Deterministic and Dynamic Networks,” *IEEE/OSA Journal of Optical Communication and Networking,* vol. 12, no. 11, November 2020.

[10] W. A. Khan, L. Wisniewski, D. Lang and J. Jasperneite, “Analysis of the requirements for offering industrie 4.0 applications as a cloud service,” in *2017 IEEE 26th International Symposium on Industrial Electronics (ISIE)*, Edinburgh, 2017.