Poster: Per-Hop Bridge-Local Latency Bounds with Strict Priority Transmission Selection

Alexej Grigorjew, Florian Metzger, Tobias Hoßfeld
University of Würzburg, Germany
{alexej.grigorjew | florian.metzger | tobias.hossfeld}@uni-wuerzburg.de

Johannes Specht
University of Duisburg-Essen, Germany
johannes.specht@uni-due.de

Franz-Josef Götz, Feng Chen, Jürgen Schmitt
Siemens AG, Germany
{franz-josef.goetz | chen.feng | juergen.jues.schmitt}@siemens.com

***Abstract*—This work presents a proven per-hop latency bound for real-time networks with admission control that does not rely on shaping or timed gates. It can be applied in a distributed control plane using only bridge-local information, making it an ideal candidate for brownfield real-time network installations. A brief comparison with Asynchronous Traffic Shaping (ATS) shows that the achieved network utilization is comparable to that of ATS in some scenarios, while it only lags behind in scenarios with accumulating bursts where re-shaping could actually reduce the interference of the observed streams.**

I. INTRODUCTION TO BOUNDED LATENCY CONCEPTS

Deterministic networking has recently raised a lot of interest due to the need for bounded latency and reliable transmission in industrial and automotive use cases. The IEEE Time Sensitive Networking (TSN) group is tasked with the standardization of mechanisms to enable deterministic guarantees in standard Ethernet networks, such as priority transmission selection [1] and time synchronization [2]. Similarly, the IETF Deterministic Networking (detnet) group is focused on the deterministic operation of a layer 3 control and data plane.

Data plane mechanisms. Tied to the transmission selection are shaping and scheduling features, including the Credit-Based Shaper (CBS, 802.1Qav [3]), timed gates (802.1Qbv [4]), and Asynchronous Traffic Shaping (ATS, 802.1Qcr [5]). These concepts can be divided into coarse-grained per-priority mechanisms (CBS, timed gates), with a lower hardware implementation complexity, and fine-grained per-stream reshaping (ATS), that require individual per-stream state. All mechanisms are intended to limit the amount of concurrent traffic and provide a basis for latency computations.

Control plane architectures. Deterministic latency requires the reservation of network resources and the dissemination of stream requirements in the network. This raises the question of responsibility and decision making. IEEE 802.1Qcc [6] suggests three types of control plane architectures, from fully distributed to fully centralized decision making. Centralized approaches can better optimize resource allocation with global knowledge, but they are a single point of failure and may have scalability issues. Fully distributed models can be difficult to provide guaranteed latencies with only bridge-local stream information available. Hybrid models typically distribute information collection to switches, but then communicate this towards a central decision making instance, thus not avoiding the scalability issues.

Some data plane mechanisms are tightly linked to certain control planes. For CBS, the Stream Reservation Protocol (SRP, 802.1Qav [3]) was developed with a fully distributed model in mind. The latency bound proposed in its application profile (802.1BA [7]) was later proven to be inaccurate [8]. For timed gates, literature commonly assumes centralized approaches with full network knowledge. Finally, asynchronous models (e.g. ATS) are intended for use with distributed reservation protocols, but they are generally also compatible with centralized models.

Control plane granularity. The information exchange and decision making of the control plane can also be fine-grained on a per-stream basis, or coarse per-priority. This is not necessarily tied to the type of data plane mechanism. For example, timed gates can only provide per-priority control for transmission scheduling, but the control plane can still collect information from all streams in the network and schedule them in a way that isolates individual streams and provides per-stream decision making. In contrast, CBS does use per-stream information, but decides based on aggregated information.

Bridge-local latency bounds without re-shaping. The data plane shaping mechanisms in TSN were developed to limit the interference of streams in real-time networks, but they require new hardware features and are not applicable in existing brownfield installations. A wide range of literature provides latency bounds for networks without per-hop re-shaping, including holistic approaches [9], a trajectory approach [10], and deterministic network calculus [11][12][13]. In general, bridge-local computation of per-hop latency bounds without re-shaping is often regarded with caution. It is commonly deemed challenging to provide deterministic networking with just strict priority transmission selection and no central authority, as it is difficult to estimate the amount of possible interference with limited information.

Annex to ISBN 978-3-903176-28-7© 2020 IFIP

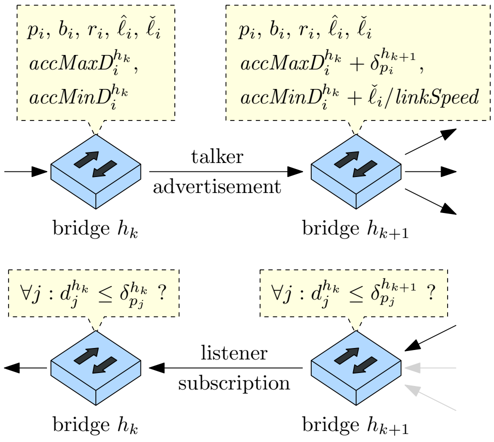

Figure 1. Resource reservation of stream i: talker advertisement traffic specification, and admission control during listener subscription.

Contribution. This paper shows that bridge-local computation of per-hop latency bounds is possible at low complexity, and that it provides feasible performance with respect to real-time bandwidth utilization when compared to ATS. The presented mechanism is explained in detail in [14]. It only uses coarse-grained priority queuing, and does not require central control or network-wide information as reservations are distributed on a bridge-to-bridge basis. It enhances each stream's information during dissemination to compute a proven latency bound for each priority level. The applied traffic model is based on the burst size b_i and traffic rate r_i. It is equivalent to commonly applied token-bucket or rate-latency traffic models, such as being used for ATS and network calculus.

II. DISTRIBUTED ADMISSION CONTROL

The latency formula for d_i presented in [14] and Eq.g 1 does not provide an independent upper bound for latency, but it is intended to check whether a given latency guarantee δ_{p_i} is met under the current conditions. Therefore, it relies on the pre-configured latency guarantee δ_{p_i} for its computation of the accumulated maximum latency $accMaxD_x$. As such, it is designed to be used in a network with active *admission control* that prevents the reservation of new streams if any existing stream would exceed its guarantee in the process.

Fig. 1 shows a simplified reservation process including admission control. In the talker advertisement, the stream's traffic specification (priority p_i, burst size b_i, traffic rate r_i, max/min frame sizes $\hat{\ell}_i$, $\check{\ell}_i$) is broadcasted to the entire network. In the process, each bridge h adjusts the accumulated maximum latency $accMaxD_i^h$ and accumulated minimum latency $accMinD_i^h$ by adding the pre-configured per-hop delay guarantee $\delta_{p_i}^h$ and the minimum required transmission time $\check{\ell}_i/linkSpeeds$ respectively. This information is stored by the bridge until the listener subscription returns on the same path. It is now used to compute the latency bound of all reserved streams for the current situation. If the computed bound d_x^h is below the guaranteed delay $\delta_{p_x}^h$ for all streams x, the subscription of i is forwarded towards the next hop. The real-time transmission only begins if the talker actually receives the subscription.

III. COMPARED LATENCY BOUNDS

Both latency bounds compared in this paper only consider transmission and queuing delay. Other delay factors, such as processing and propagation delay, are assumed to be independent of the amount of traffic and therefore omitted here.

Priority transmission selection. Based on the traffic specifications from the talker advertisements, the worst case queuing and transmission delay of all frames from a stream i at hop h_k is bounded by:

$$d_i^{h_k} \le \sum_{x\in\mathcal{H}_i} \frac{y_{i,x} b_x}{linkSpeed} + \sum_{x\in\mathcal{E}_i} \frac{z_x b_x}{linkSpeed} + \max_{x\in\mathcal{L}_i} \frac{\hat{\ell}_x}{linkSpeed} \tag{1}$$

Here, the sets $\mathcal{H}_i$, $\mathcal{E}_i$, and $\mathcal{L}_i$ represent all streams with higher, equal, and lower priority, respectively (including i itself). The terms $y_{i,x}$ and z_x represent the number of bursts from higher and equal priority streams x. Eqs. 2 and 3 can be used such that the upper bound in Eq. 1 is valid.

$$y_{i,x} = \left\lceil (accMaxD_x^{h_k} - accMinD_x^{h_{k-1}} + \delta_{p_i}) \cdot r_x/b_x \right\rceil \tag{2}$$

$$z_x = \left\lceil (accMaxD_x^{h_k} - accMinD_x^{h_{k-1}}) \cdot r_x/b_x \right\rceil \tag{3}$$

For a formal proof and further details, refer to [14].

Asynchronous traffic shaping. As the above concept is compared to ATS, Eq. 4 presents its latency bound that relies on per-hop re-shaping of the traffic [15].

$$d_i^{h_k} \le \frac{\sum_{x\in\mathcal{H}_i\cup\mathcal{E}_i} b_x + \max_{x\in\mathcal{L}_i} \hat{\ell}_x - \check{\ell}}{linkSpeed - \sum_{x\in\mathcal{H}_i} r_x} + \frac{\check{\ell}}{linkSpeed} \tag{4}$$

Therefore, $\check{\ell}$ is set to the smallest worst-case Ethernet frame size of 64 B. For more details, refer to [15]. It features a fluid traffic model instead of looking at counts of bursts in the network, thus $y_{i,x}$ is vaguely represented by $\sum_{x\in\mathcal{H}_i} r_x$ in the denominator. Due to per-hop re-shaping on the data plane, only one burst b_x from each equal-priority stream $x \in \mathcal{E}_i$ can interfere at any time, i.e. the factor z_x is equal to 1 for ATS.

IV. COMPARISON OF NETWORK CAPACITY

A small, linear topology consisting of three bridges is used for the capacity comparison. Both corner bridges are connected to three hosts each, representing the talkers and listeners of this evaluation. A variable number of streams is deployed at random between these hosts, the source and destination of each stream is chosen at random. This means that most streams share the bottleneck links of the center bridge at 1 Gbit/s. The individual traffic specification of each stream is chosen from a set of predefined values, featuring two priority levels (2 and 3), five burst sizes (128 B-1522 B), and two traffic rates (mostly 512 kB/s, with 318 kB/s in the 1522 B burst case).

The experiment is conducted with four different per-hop delay guarantee configurations, ranging from 100 µs to 2000 µs for the high priority, and from 250 µs to 8000 µs for the low priority. For each configuration, 20 repetitions with different

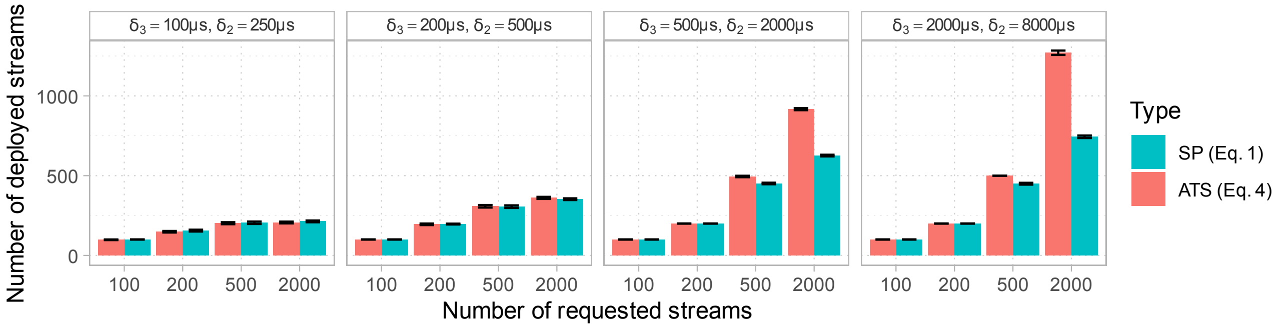

Figure 2. Comparison of network capacity with priority transmission selection and ATS. The number of streams and the delay guarantees δ_{p_i} are varied. Bars show mean values from 20 repetitions, whiskers show 99.5% confidence intervals.

stream traffic specifications are created and the streams consecutively attempt to reserve resources as described in Section II. The mean number of accepted reservations for both ATS and strict priority (SP) transmission is reported in Figure 2 with whiskers representing 99.5% confidence intervals.

The capacities under these different configurations allow three major observations. (1) With ATS, the admission control can always accept at least as many streams as the basic priority transmission selection. This is to be expected, as it is designed to reduce interference, especially between equal-priority streams. (2) For some configurations, ATS and SP accept a very similar amount of streams, especially for those configurations with very tight delay guarantees ($\delta_3 = 100\,\mu\text{s}$, $\delta_2 = 250\,\mu\text{s}$ and $\delta_3 = 250\,\mu\text{s}$, $\delta_2 = 500\,\mu\text{s}$). In these scenarios, the end-to-end delays are so small that bursts cannot accumulate, the shaping from ATS shows no benefit as it is not necessary yet. (3) For looser guarantees ($\delta_3 = 500\,\mu\text{s}$, $\delta_2 = 2000\,\mu\text{s}$ and $\delta_3 = 2000\,\mu\text{s}$, $\delta_2 = 8000\,\mu\text{s}$), ATS can accept significantly more streams than SP with up to 70% difference in our scenarios. In these configurations, the accumulated maximum latency becomes large enough that Eq. 1 considers multiple bursts of each stream as interference. Note that the re-shaping done by ATS requires hardware support, and SP can still be a feasible alternative in these scenarios, depending on the amount of required real-time utilization.

V. CONCLUSION

This work presents a brief categorization of deterministic networking concepts and shows that simple latency bounds can be derived without sophisticated mechanisms. It suggests a distributed methodology for resource reservation that does not require shaping or a central controller, but works with priority queuing alone. Finally, it provides an overview of its efficiency by conducting example stream reservations and comparing the achieved utilization with a sophisticated shaping mechanism. The results show that priority queuing is indeed a feasible trade-off with lower hardware requirements, and in some cases it can even match the efficiency of Asynchronous Traffic Shaping.

REFERENCES

[1] "IEEE Standard for Local and Metropolitan Area Network – Bridges and Bridged Networks", *IEEE Std 802.1Q-2018 (Revision of IEEE Std 802.1Q-2014)*, 2018.

[2] "IEEE Standard for Local and Metropolitan Area Networks – Timing and Synchronization for Time-Sensitive Applications", *IEEE P802.1AS-Rev*, 2019.

[3] "IEEE Standard for Local and Metropolitan Area Networks – Virtual Bridged Local Area Networks – Amendment: Forwarding and Queuing Enhancements for Time-Sensitive Streams", *IEEE Std 802.1Qav-2010*, 2010.

[4] "IEEE Standard for Local and Metropolitan Area Network – Bridges and Bridged Networks – Amendment 25: Enhancements for Scheduled Traffic", *IEEE Std 802.1Qbv-2015*, 2015.

[5] "Draft Standard for Local and Metropolitan Area Networks — Bridges and Bridged Networks — Amendment: Asynchronous Traffic Shaping", *IEEE P802.1Qcr/D2.0*, 2019.

[6] "IEEE Standard for Local and metropolitan area networks – Bridges and Bridged Networks – Amendment 31: Stream Reservation Protocol (SRP) Enhancements and Performance Improvements", *IEEE Std 802.1Qcc*, 2018.

[7] "IEEE Standard for Local and metropolitan area networks – Audio Video Bridging (AVB) Systems", *IEEE Std 802.1BA-2011*, pp. 1–45, 2011.

[8] C. Boiger, "Class A bridge latency calculations", in *IEEE 802 November Plenary Meeting*, 2010.

[9] J. J. Gutiérrez, J. C. Palencia, and M. G. Harbour, "Holistic schedulability analysis for multipacket messages in afdx networks", *Real-Time Systems*, vol. 50, no. 2, pp. 230–269, 2014.

[10] S. Martin, P. Minet, and L. George, "The trajectory approach for the end-to-end response times with non-preemptive fp/edf", in *International Conference on Software Engineering Research and Applications*, Springer, 2004, pp. 229–247.

[11] J.-Y. Le Boudec and P. Thiran, *Network calculus: a theory of deterministic queuing systems for the internet*. Springer Science & Business Media, 2001, vol. 2050.

[12] S. Bondorf, P. Nikolaus, and J. B. Schmitt, "Quality and cost of deterministic network calculus: Design and evaluation of an accurate and fast analysis", *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, vol. 1, no. 1, p. 16, 2017.

[13] K. Jang, J. Sherry, H. Ballani, and T. Moncaster, "Silo: Predictable message latency in the cloud", *ACM SIGCOMM Computer Communication Review*, vol. 45, no. 4, pp. 435–448, 2015.

[14] A. Grigorjew, F. Metzger, T. Hoßfeld, J. Specht, F.-J. Götz, J. Schmitt, and F. Chen, "Technical report on bridge-local guaranteed latency with strict priority scheduling", Institut für Informatik, Tech. Rep., 2020.

[15] J. Specht and S. Samii, "Urgency-based scheduler for time-sensitive switched ethernet networks", in *2016 28th Euromicro Conference on Real-Time Systems (ECRTS)*, Jul. 2016, pp. 75–85.