Aligning rTWT with 802.1Qbv: a Network Calculus Approach

Carlos Barroso-Fernández
cbarroso@pa.uc3m.es
Universidad Carlos III de Madrid
Leganés, Madrid, Spain

Jorge Martín-Pérez
jorge.martin.perez@upm.es
Universidad Politécnica de Madrid
Madrid, Madrid, Spain

Constantine Ayimba
ayconsta@it.uc3m.es
Universidad Carlos III de Madrid
Leganés, Madrid, Spain

Antonio de la Oliva
aoliva@it.uc3m.es
Universidad Carlos III de Madrid
Leganés, Madrid, Spain

ABSTRACT

Industry 4.0 applications impose the challenging demand of delivering packets with bounded latencies via a wireless network. This is further complicated if the network is not dedicated to the time critical application. In this paper we use network calculus analysis to derive closed form expressions of latency bounds for time critical traffic when 802.11 Target Wake Time (TWT) and 802.1Qbv work together in a shared 802.11 network.

CCS CONCEPTS

• **Networks** → **Network performance analysis**; *Network performance modeling*.

KEYWORDS

Network Calculus, Scheduling, 802.11Qbv, restricted Target Wake Time (rTWT)

ACM Reference Format:
Carlos Barroso-Fernández, Jorge Martín-Pérez, Constantine Ayimba, and Antonio de la Oliva. 2023. Aligning rTWT with 802.1Qbv: a Network Calculus Approach. In *The Twenty-fourth International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing (MobiHoc '23), October 23–26, 2023, Washington, DC, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3565287.3617606

1 INTRODUCTION

Unscheduled channel access in 802.11 networks negatively impacts their reliability particularly in dense deployments. This is a particularly challenging phenomenon given the imminent use of wireless networks for industry 4.0 applications [5]. TWT reduces contention by allowing the AP to instruct Station (STA)s to turn their transceivers on or off at agreed intervals. The discipline thus transforms a dense deployment to one with a manageable number of STAs during a TWT session therefore reducing contention and improving the radio channel access delay. Moreover, 802.11be will extend it to restricted Target Wake Time (rTWT) to completely eliminate contention[1]. This method has been proposed as a scheduling mechanism in [1] that achieves improvement in overall network throughput but does not consider the priority of packets. This approach is therefore inadequate for Time-Sensitive Networking (TSN) flows requiring bounded latency. We propose an enhancement, incorporating TSN, by aligning channel access via rTWT with 802.1Qbv scheduling. The authors of [3] provide a proof of concept implementation of TWT for time aware scheduling that isolates priority traffic from best effort traffic achieving promising results on bounded latency and jitter. We leverage network calculus analysis [2] to prove that bounded latencies, required by TSN, can be achieved using this approach when both high and low priority packets are present in a STA with some violation probability that depends on channel quality. In Section 2 we model the 802.1Qbv opening gates using Network Calculus [2, 6], and in Section 3 the wireless errors using stochastic scaling [4]. We discuss results in Section 5 and conclusions in Section 6.

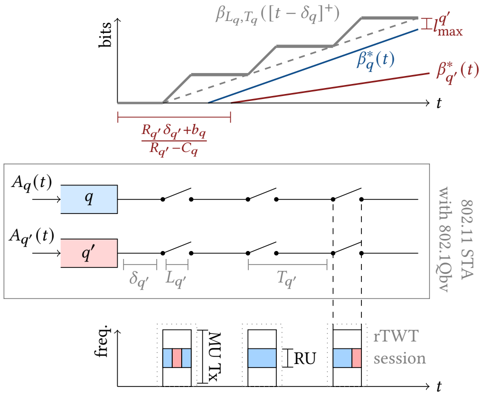

Figure 1: An STA (middle) with two 802.1Qbv queues of high q and low priority q'. The opening times of 802.1Qbv gates are aligned with the rTWT sessions (bottom). The STA has a single RU in the MU Tx scheduled within each TWT session. On top, the resulting network calculus service curves $\beta^*_q, \beta^*_{q'}$.

2 MODELLING 802.1QBV AS TDMA CURVES

The Time-Aware Shaper (TAS) in 802.1Qbv decides when to open and close each queue using a Gate Control List (GCL). The latter defines both the periodicity T_q of each queue and how long, L_q,

MobiHoc '23, October 23–26, 2023, Washington, DC, USA

ACM ISBN 978-1-4503-9926-5/23/10.
https://doi.org/10.1145/3565287.3617606

---

[1] In 802.11ax contention can be ameliorated with techniques such as trigger-based Orthogonal Frequency Division Multiple Access (OFDMA).

it is open for transmission. Using Time Division Multiple Access (TDMA), a user queue q has a transmit service curve (cummulative bits processed over time), $\beta_{L,T}(t)$, given by (1)

$$\beta_{L,T}(t) = C \cdot \max\left\{\left\lfloor \frac{t}{T} \right\rfloor L, t - \left\lceil \frac{t}{T} \right\rceil (T - L)\right\} \tag{1}$$

with C being the transmission rate and T, L the periodicity and length of the TDMA transmission opportunities. The service curve in 802.1Qbv for a queue q is given by $\beta_{L_q,T_q}([t-\delta_q]^+)$ where $\delta_q > 0$ is the time at which queue q gate is first opened.

Note that in 802.11Qbv, two or more queues may have their gates opened concurrently[2]. We use an affine lower envelope of the TDMA service curve (1) to handle gate concurrency as shown in the (top) dashed line of Figure 1. The TDMA affine lower envelope for queue q in 802.1Qbv is the rate-latency function

$$\beta_q(t) = \frac{C \cdot L_q}{T_q + L_q}[t - \delta_q]^+. \tag{2}$$

3 CONCURRENT 802.1QBV OPENED GATES

When high priority queue q and low priority queue q' are opened concurrently, packets of q are transmitted ahead of those from q'.

Corollary 0.1 (802.1Qbv High priority service curve). *Given two queues q, q' whose gates are opened concurrently; the high priority queue q has a strict service curve*

$$\beta_q^*(t) = R_q \left[t - \left(\delta_q + \frac{l_{\max}^{q'}}{R_q}\right)\right]^+ \tag{3}$$

with $R_q = \frac{CL_q}{T_q + L_q}$ and $l_{\max}^{q'}$ the maximum packet size of the low priority queue q'.

Proof. We mimic the proof of [2][Proposition 1.3.4]. Take the affine lower envelope of queue q defined in (1). Consider s as the start of the backlog period of queue q, hence, the arrival and departing curve at queue q is the same at that time $D_q(s) = A_q(s)$.

For 802.1Qbv is non-preemptive, if a low priority packet from q' arrives before a high priority packet from q in the interval $(s, t]$; the high priority packet will wait, i.e.

$$D_q(t) - D_q(s) \geq \beta_q(t - s) - l_{\max}^{q'} \tag{4}$$

given $D_q(s) = A_q(s)$ and $A, D \in \mathcal{F} = \{f : \mathbb{R}^+ \to \mathbb{R}, \forall t \geq s : f(t) \geq f(s), f(0) = 0\}$; we know

$$D_q(t) \geq A_q(s) + \left[R_q \left[(t - s) - \delta_q\right]^+ - l_{\max}^{q'}\right]^+ \tag{5}$$

If $(t - s) > \delta_q$ (5) becomes

$$D_q(t) \geq A_q(s) + R_q \left[(t - s) - \left(\delta_q + \frac{l_{\max}^{q'}}{R_q}\right)\right]^+ \tag{6}$$

If $(t - s) \leq \delta_q$ (5) becomes $D_q(t) \geq A_q(s)$. That is, the departing flow of the high priority queue q satisfies

$$D_q(t) \geq A_q \otimes \beta_q^*(t) \tag{7}$$

with $\beta_q^*(t)$ being the rate-latency service curve with rate R_q and latency $\delta_q + \frac{l_{\max}^{q'}}{R_q}$. □

Similarly, it is also possible to obtain the service curve for the traffic in the low priority queue q' when such queue is opened concurrently with the higher priority queue q.

Corollary 0.2 (802.1Qbv Low priority service curve). *Given two queues q, q' whose gates are opened concurrently; if the flow at q has a strict affine arrival curve γ_{C_q,b_q}, then the low priority queue q' has a strict service curve*

$$\beta_{q'}^*(t) = \left(R_{q'} - C_q\right)\left[t - \frac{R_{q'}\delta_{q'} + b_q}{R_{q'} - C_q}\right]^+ \tag{8}$$

with $R_{q'} = \frac{CL_{q'}}{T_{q'} + L_{q'}}$.

Proof. We again mimic the proof presented in [2] [Proposition 1.3.4]. Take the affine lower envelope of queue q' as defined in (1). Consider s' the start of the busy period, that is, $s' < s$ with s the start of the backlog period. In the interval $(s', t]$ the low priority queue departing traffic satisfies

$$D_{q'}(t) - D_{q'}(s') = \beta_{q'}(t - s') - \left[D_q(t) - D_q(s')\right] \tag{9}$$

Given that at s' there is no backlog, and the strict arrival curve for the high priority queue q, we know

$$\begin{aligned} D_q(t) - D_q(s') &= D_q(t) - A_q(s') \\ &\leq A_q(t) - A_q(s') \leq \gamma_{C_q,b_q}(t - s') \end{aligned} \tag{10}$$

Then, we substitute (10) in (9) to obtain

$$\begin{aligned} D_{q'}(t) - D_{q'}(s') &= D_{q'}(t) - A_{q'}(s') \\ &\geq \left[R_{q'}\left[(t - s) - \delta_{q'}\right]^+ - \left(C_q(t - s') + b_q\right)\right]^+ \end{aligned} \tag{11}$$

Again, depending on $(t - s)$ the above expression changes. For $(t - s) > \delta_{q'}$, (11) becomes

$$D_{q'}(t) \geq A_{q'}(s') + \left[\left(R_{q'} - C_q\right)(t - s') - \left(R_{q'}\delta_{q'} + b_q\right)\right]^+ \tag{12}$$

and with $(t - s') \leq \delta_{q'}$, inequality (11) becomes $D_{q'}(t) \geq A_{q'}(s')$. Overall, we can tell that the departing flow at the low priority queue q' satisfies

$$D_{q'}(t) \geq A_{q'} \otimes \beta_{q'}^*(t) \tag{13}$$

with $\beta_{q'}^*(t)$ being the rate-latency service curve with rate $R_{q'} - C_q$ and latency $\frac{R_{q'}\delta_{q'} + b_q}{R_{q'} - C_q}$. □

4 RETRANSMISSIONS IN 802.11

So far, we have obtained the service curves $\beta_q^*(t), \beta_{q'}^*(t)$ of high/low priority traffic upon the opened/closed 802.1Qbv gates. By aligning gate opening with the rTWT sessions as depicted in Figure 1, channel contention is avoided since the STA has a dedicated Resource Unit (RU) in 802.11ax Multi User transmission (MU) Tx, and the rTWT session access is restricted.

However, packet errors may still occur due to channel impairments resulting in retransmissions. To deal with the latter, we use the stochastic scaling approach proposed in [4]. The idea is to send the departing traffic of 802.1Qbv $D_q(t), D_{q'}(t)$ to a wireless channel modeled with a scaling process S as shown in Figure 2. If the Tx does not succeed, the STA waits time W (as maximum) before retransmission of the i$^{\text{th}}$ packet. This process is modelled with the scaling curve δ_W.

---

[2] In this paper, we consider 2 queues for brevity. Our analysis can be easily extended to N concurrent queues.

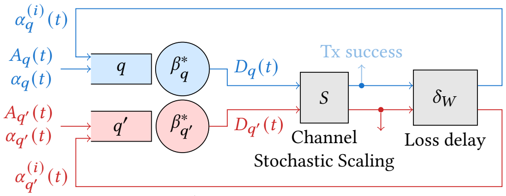

Figure 2: When the q, q' are opened, packets $D(t)$ traverse the stochastic wireless channel S. On Tx error, δ_W represents the time it takes to realize the packet loss. Then, packets are fed-back for retransmission leading to re-entrant arrivals $\alpha^{(i)}(t)$.

Following [4], we consider a binary symmetric channel modelled with stochastic scaling process $S(b) = \sum_i^b X_i$, $X_i \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$, with p the loss probability. We now define a stochastic scaling curve $S^\varepsilon(b) = pb + 1 - \varepsilon$ that, according to [4, Theorem 1], satisfies

$$\mathbb{P}\left(\sup_{0\le a\le b}\left\{S(b) - S(a) - S^\varepsilon(b-a)\right\} \le 0\right) \ge 1-\varepsilon, \quad \forall b \ge 0 \quad (14)$$

that is, the queue experiencing Tx errors $S(D(t))$ is upper bounded by the outgoing traffic (which is bounded by $S^\varepsilon(D(t))$) with probability $1-\varepsilon$.

The scaled flow $S(D(t))$ will result in retransmissions, namely, we have $\alpha^{(i)}(t)$, $i = 0, 1, \ldots, N$ being the arrival curves for the i$^{\text{th}}$ retransmissions, where $\alpha^{(0)} = A(t) = \gamma_{r,b}$ represents the original input flow. The higher the Rtx index, the higher the priority, thus each retransmission flow has arrival and service curves

$$\beta_q^{(i)} = \left[\beta - \sum_{k=i+1}^{N} \alpha^k\right]^+, \qquad \alpha_q^{(i)} = S^\varepsilon\left(\alpha^{(i-1)} \oslash \beta^{(i-1)}\right) \oslash \delta_W. \quad (15)$$

Once the system is solved, we define the aggregated arrival curve for the queue q as $\alpha_q^{(Tot)} = \sum_{j=0}^N \alpha_q^{(j)}$. At this point, given the aggregated arrival curve and the service curve, Network Calculus provides bounds to obtain the maximum delay experienced by queue q using the following expression:

$$h(\alpha, \beta) := \sup_{s\ge 0}\left\{\inf\{u \ge 0 : \alpha(s) \le \beta(s+u)\}\right\}.$$

Note that previous curves are derived from the stochastic scaling process that model the channel, hence, they are not deterministic. In fact, as [4] highlights, the reliability of the delay bound (i.e., the probability that it holds) will be

$$\mathbb{P}\left(d(t) \le h(\alpha^{(Tot)}, \beta)\right) \ge (1-\varepsilon)^N \qquad \forall t \in \mathbb{R}^+, \quad (16)$$

with $d(t)$ representing the real delay experienced by packets at time t.

5 RESULTS

To highlight the importance of these results, we analyze two scenarios: with (i) $T = 3$, $L = 1$ and (ii) $T = 3$, $L = 2$. In both, we calculate the aggregated arrival curves for one non-overlapping queue and two overlapping queues (one with higher priority than the other). The rest of the parameters are $C = 10$, $\alpha^{(0)} = \gamma_{0.1,0.001}$ and $N = 3$. Additionally, the violation probability of the stochastic scaling curve ε is $3.3344 \cdot 10^{-4}$.

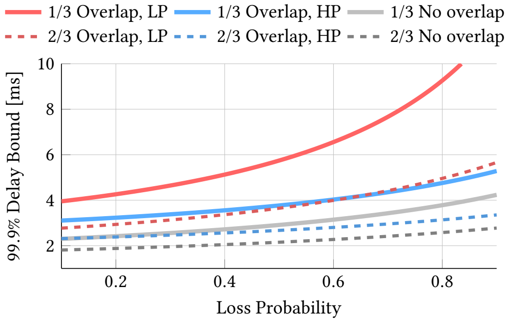

Figure 3: Delay bound vs. wireless loss probability for different $\frac{L}{T}$ ratios. The Delay Bound holds with a 99.9% probability. We test non overlapping setups with single 802.1Qbv queues (gray), and overlapping setups with concurrent high (HP) and low priority (LP) queues (blue and red).

Figure 3 shows the maximum delay experienced by the queues as a function of the loss probability p. Note that these bounds hold with a 0.999 probability. We observe that the delay is reduced with greater overlap. The high priority queue is less impacted compared to the low priority queue. Moreover, comparing the $L/T = 1/3$ non-overlapping case with $2/3$ overlapping case, the high priority queue in the latter case experiences less delay at the expense of the low priority queue.

6 CONCLUSIONS AND FUTURE WORK

We have presented stochastic bounds for TSN packets in a 802.11 deployment, with queue concurrency and retransmissions due to channel impairments. Our expressions assume the alignment of 802.1Qbv queues with 802.11 TWT sessions.

This paper provides a preliminary, simplified assessment of dense Industry 4.0 TSN wireless deployments. We intend to extend our analysis to more complex scenarios and develop a proof of concept experiments to validate our results.

ACKNOWLEDGMENT

This work has been partially funded by the European Commission Horizon Europe SNS JU PREDICT-6G (GA 101095890) Project and the Spanish Ministry of Economic Affairs and Digital Transformation and the European Union-NextGenerationEU through the UNICO 5G I+D 6G-EDGEDT and 6G-DATADRIVEN.

REFERENCES

[1] Chen, Q., and Zhu, Y.-H. Scheduling channel access based on target wake time mechanism in 802.11ax wlans. *IEEE Transactions on Wireless Communications 20*, 3 (2021), 1529–1543.

[2] Le Boudec, J.-Y., and Thiran, P. *Network calculus: a theory of deterministic queuing systems for the internet.* Springer, 2001.

[3] Schneider, B., Sofia, R. C., and Kovatsch, M. A proposal for time-aware scheduling in wireless industrial iot environments. In *NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium* (2022), pp. 1–6.

[4] Wang, H., Schmitt, J., and Ciucu, F. Performance modelling and analysis of unreliable links with retransmissions using network calculus. In *Proceedings of the 2013 25th International Teletraffic Congress (ITC)* (2013), pp. 1–9.

[5] Zhang, K., Shi, Y., Karnouskos, S., Sauter, T., Fang, H., and Colombo, A. W. Advancements in industrial cyber-physical systems: An overview and perspectives. *IEEE Transactions on Industrial Informatics 19*, 1 (2023), 716–729.

[6] Zhao, L., Pop, P., and Craciunas, S. S. Worst-Case Latency Analysis for IEEE 802.1Qbv Time Sensitive Networks Using Network Calculus. *IEEE Access 6* (2018), 41803–41815.