Beyond QoS: Integrating Brain-Aware Constraints into Delay Bounds for uVR and Machine-Type Communication

Benedetta Picano
Dpt. of Information Engineering
Università di Firenze
Firenze, Italy
https://orcid.org/0000-0003-4970-1361

Tommaso Pecorella
Dpt. of Information Engineering
Università di Firenze
Firenze, Italy
https://orcid.org/0000-0002-0009-8154

***Abstract*—Next-generation computing platforms are increasingly expected to accommodate a wide range of immersive applications, integrating advanced technologies such as extended reality and ultra-realistic virtual reality (uVR). The design and optimization of these distributed systems present significant challenges, particularly when managing the coexistence of traffic flows generated by both human and machine sources. This work introduces an end-to-end delay analysis framework focused on human-driven traffic in environments characterized by mixed-source flows. The human-generated streams—central to the analysis—coexist with machine-type communications that differ in terms of latency requirements. The delay analysis is carried out by deriving per-flow stochastic bounds, expressed in terms of the probability of receiving timely service. This is achieved through the use of stochastic network calculus and martingale-based methods, with human perceptual constraints explicitly incorporated into the theoretical model. Simulation-based validation confirms that the proposed bound accurately reflects actual end-to-end behavior. Moreover, the results demonstrate that integrating human perception into the analysis leads to improved performance compared to conventional approaches that do not account for cognitive aspects.**

***Index Terms*—next-generation computing systems, end-to-end delay, human-centric applications**

I. Introduction

In line with the vision driving next-generation applications, distributed computing infrastructures are expected to support a new class of highly demanding and disruptive services, often characterized by stringent latency requirements and a strong focus on user experience. Among the most emblematic examples are interactive and immersive gaming platforms, advanced virtual reality (VR) environments, and the emerging ultimate VR (uVR), which aims to stimulate all five human senses [1]–[7].

While uVR holds great promise due to its increasing market relevance and potential impact across multiple domains—including social interaction, education, healthcare, and entertainment [8]—it also introduces a paradigm shift in system design. Unlike traditional service generations where Quality of Service (QoS) metrics were sufficient to assess system performance, immersive applications demand that users themselves perceive the service as satisfactory, thus emphasizing Quality of Experience (QoE) as the primary evaluation criterion [9]–[11]. This necessitates a human-centered approach, placing the user at the core of system design and optimization.

Quality of Experience (QoE) aims to reflect the degree of satisfaction experienced by users [11], and is widely recognized as one of the most critical and challenging aspects of next-generation services. As the interaction between humans and digital applications becomes increasingly central, it is essential to account for user perception during the design, management, and optimization of future distributed computing platforms.

At the same time, ensuring strict Quality of Service (QoS) guarantees—such as low latency and high reliability—remains fundamental for proper service delivery. However, it has been observed that beyond certain thresholds, further QoS improvements may not translate into perceptible gains for users, due to the intrinsic limits of human perception [1]. The mapping between objective QoS levels and perceived QoE is in fact nontrivial and cannot be directly inferred.

For this reason, the integration of human-centric considerations into resource allocation strategies has gained momentum. Recent interdisciplinary research has shown that users are often unable to fully perceive or benefit from QoS enhancements, particularly in scenarios like video streaming, where both data rate and latency are involved [1], [12], [13]. These findings suggest that future networks could greatly benefit from adaptive mechanisms that consider behavioral and cognitive constraints in human perception.

In this context, end-to-end (e2e) delay emerges as a key metric to characterize the behavior of distributed systems supporting uVR services [7]. Furthermore, studying such systems under mixed traffic conditions—including human-generated and machine-type flows with heterogeneous delay sensitivities—offers valuable insights for improving the overall efficiency of resource usage.

Stochastic Network Calculus (SNC) has recently emerged as a powerful analytical framework for end-to-end delay char-

Identify applicable funding agency here. If none, delete this.

acterization, particularly in scenarios involving non-Markovian traffic patterns [14], [15].

Building upon the methodology introduced in [16], this work presents a novel formulation of a stochastic bound on the end-to-end delay, derived within the SNC framework. A distinguishing aspect of the proposed analysis is the explicit integration of human Quality of Experience (QoE) considerations, in a setting where uVR traffic shares network resources with machine-type (MT) communication flows.

The main contributions of this paper, can be outlined as follows.

- Derivation of the complementary cumulative distribution function (CCDF) of the end-to-end delay bound for human-centric uVR traffic flows coexisting with concurrent MT services. The analysis accounts for two distinct scheduling disciplines. The stochastic bounds are computed by applying SNC principles combined with martingale-based envelope techniques.
- We incorporate human perceptual constraints into the derived stochastic bound, capturing the cognitive limitations in delay sensitivity. The results demonstrate enhanced resource utilization for both uVR and MT traffic types, while still meeting the end-to-end delay requirements imposed by each flow class.

The rest of the manuscript is organized as follow. Section II illustrates the system model. In Section III the e2e delay bound is formulated, and the performance comparison is investigated in Section IV. Then, conclusions are drawn in Section V.

II. System Model

The reference scenario consists of a small base station (SBS) co-located with an edge computing node (EC), tasked with supporting both ultimate VR (uVR) services and machine-type (MT) applications. We consider two uVR users, associated with data flows $\mathcal{H}$ and $\mathcal{G}$ respectively, and a single MT flow denoted by $\mathcal{C}$. All users and devices are assumed to be located at a distance d_0 from the EC node.

As depicted in Figure **??**, each uVR user accesses the EC node through a dedicated uplink channel, operating synchronously and without interference from other uVR access channels. In contrast, MT devices share a common synchronous uplink channel, which follows a slotted ALOHA access scheme—described in detail later. The results of computation, for both uVR and MT tasks, are transmitted back to the respective users via a shared downlink channel.

This setup gives rise to a heterogeneous compound batch arrival process, in which incoming service requests are managed according to a First-In-First-Out (FIFO) scheduling policy. Importantly, all requests within the same batch are served before the system processes the next batch of arrivals [17], [18].

Informed by the findings of [1], we model user-specific perceptual thresholds to reflect the limitations of human delay sensitivity. For each uVR user i, the parameter β_i represents the minimum end-to-end delay variation that can be cognitively perceived. When the experienced delay is below this threshold, user i is unlikely to notice any improvement in service quality. As detailed in [1], this threshold emerges from complex cognitive mechanisms involving attention, effort, and contextual awareness.

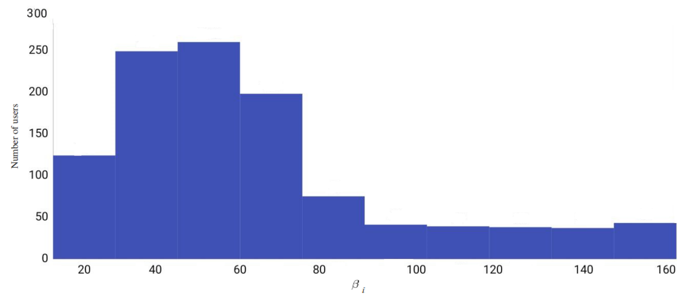

Fig. 1. Distribution of β_i considering 1000 uVR users

Human perceptual response to network delay is commonly modeled through a multi-modal statistical distribution [19], as supported by the findings in [1]. In this context, the probability that a user i perceives a delay lower than a specific threshold $D_i(\epsilon')$ is expressed as:

$$\mathbb{P}\{\beta_i < D_i(\epsilon')\} < \epsilon', \tag{1}$$

where D_i represents the effective end-to-end delay experienced by user i, and ϵ' denotes the probability that user i is able to perceive improvements in delay below $D_i(\epsilon')$.

Based on the analytical framework introduced in [1] and extended in [16], the expression for $D_i(\epsilon')$ is derived as:

$$D_i(\epsilon') = \mu_k(d+1) - \sqrt{Q_{d+1}(1-2\epsilon')\, \boldsymbol{e}_{d+1}^T C_k \boldsymbol{e}_{d+1}}, \tag{2}$$

where d denotes the dimensionality of the human feature vector $\boldsymbol{x}_i \in \mathbb{R}^d$, which characterizes user i (e.g., age, location, or other contextual factors). The term $Q_{d+1}(\cdot)$ refers to the quantile function of the chi-square distribution with $d+1$ degrees of freedom. The vector $\boldsymbol{e}_{d+1} \in \mathbb{R}^{d+1}$ is defined such that its $(d+1)$-th component equals 1, while all other elements are zero. The parameters $\mu_k(d+1)$ and C_k represent the $(d+1)$-th component of the mean vector and the covariance matrix, respectively, associated with the k-th mode of the brain perception distribution [1].

By following the numerical estimation procedure described in [1], the perceptual delay thresholds can be quantitatively evaluated. The resulting distribution of human perception is illustrated in Figure 1.

III. Human-Centric E2E Bound

A. Delay Model

Let p_t denote the probability that a uVR user has a data packet to transmit in a given time slot. In contrast, MT devices associated with flow $\mathcal{C}$ employ a contention-based slotted ALOHA protocol to access a shared uplink channel. A saturated traffic condition is assumed for the MT side, meaning that each device always has a packet ready for transmission to the EC node. However, a new packet is generated only if

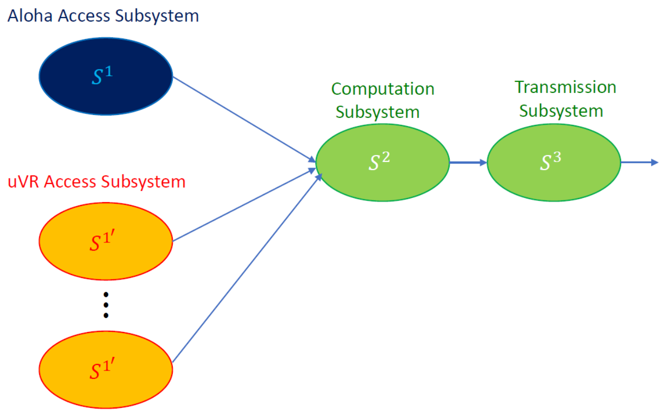

Fig. 2. Tandem system.

the previous one has already been transmitted; otherwise, no additional packet is queued.

Let p_m represent the per-slot transmission probability for any data packet originating from an MT device. This value applies uniformly to both initial transmissions and retransmissions, with no distinction made between the two cases.

The probability that a packet successfully reaches the EC node from a specific MT device, given a total of M devices in the set $\mathcal{C}$, is expressed as:

$$p_S = \binom{M}{1} p_m (1 - p_m)^{M-1}, \tag{3}$$

which corresponds to the probability that exactly one device transmits in a slot while all others remain silent.

Once packets from both MT devices and uVR users are processed by the computation subsystem, the corresponding outputs are transmitted back to their respective sources via a shared downlink channel. This channel constitutes the transmission subsystem and is common to all users.

As a result, the system can be modeled as a tandem queueing structure composed of three sequential subsystems: (i) the access subsystem, responsible for uplink transmissions; (ii) the computation subsystem, which executes task processing; and (iii) the transmission subsystem, which delivers the results back to the users. The presence of the computation subsystem introduces an additional delay component, due to the time a packet may spend waiting to be processed at the EC node. Likewise, the transmission subsystem contributes with a fixed delay, as each result must be sent back to the corresponding uVR user or MT device. This contribution is deterministic and equal to one slot for both traffic types.

In contrast, the access subsystem impacts the e2e delay differently depending on the nature of the traffic. For uVR flows, the access delay is constant and equal to one slot, whereas for MT flows it is random, as governed by the success probability in (3). As a result, the e2e delay analysis must distinguish between uVR and MT traffic.

Focusing on the e2e delay experienced by a specific (tagged) uVR flow i, the arrival of its packets at the EC subsystem is influenced by possible simultaneous transmissions from the remaining $v - 1$ uVR users and from MT devices. This leads to the following statistical distribution:

$$A_c^1 = \begin{cases} 0, & (1-p_t)^{v-1}(1-p_S)^M \\ 1, & \binom{v-1}{1} p_t (1-p_t)^{v-2}(1-p_S) + (1-p_t)^{v-1} p_S \\ \dots & \\ v-1, & p_t^{v-1}(1-p_s)^M + \binom{v-1}{v-2} p_t^{v-2}(1-p_t) p_S \\ v, & p_t^{v-1} p_S. \end{cases} \tag{4}$$

Similarly, when the tagged flow corresponds to traffic generated by an MT device, the arrival process of cross traffic at the computation subsystem can be characterized by the following statistical distribution:

$$A_c^2 = \begin{cases} 0, & (1-p_t)^{v} \\ 1, & \binom{v}{1} p_t (1-p_t)^{v-1} \\ \dots & \\ v-1, & \binom{v}{v-1} p_t^{v-1}(1-p_t) \\ v, & p_t^{v}. \end{cases} \tag{5}$$

Let S^2 and S^3 be the curve of the service time at the computation site, and of the service time at the transmission subsystem, respectively. S^2 is assumed to be modeled as a hypoexponential distribution with two stages for the two traffic type. Differently, S^3 is assumed to be deterministic.

B. Analysis for FIFO Policy

Under the FIFO scheduling assumption, the analysis focuses on the delay experienced by a generic tagged flow i, which can belong to either the uVR traffic set ($\mathcal{H}$, $\mathcal{G}$) or the MT traffic set ($\mathcal{C}$). The cross traffic is described by an aggregated arrival envelope A_{c_i}, whose distribution depends on the type of the tagged flow, as defined in (4) and (5).

Following SNC theory [20], [21], and assuming the existence of martingale envelopes for both arrivals and service, the available service to flow i is bounded below by the convolution of its arrival envelope and the service curve S^2. A probabilistic bound on the queueing delay is then derived as a supremum of the backlog process, which depends on the difference between the aggregate arrivals and the available service.

Martingale bounds for arrival and service processes are formulated using exponential supermartingale approximations. Their combination under independence assumptions and the use of the stopping theorem [22] yields a bound on the delay violation probability for the tagged flow. In particular, the bound for an MT flow is expressed as:

$$\mathbb{P}(W_{MT}(n) \geq k) \leq e^{-\theta^* k K_s} \Gamma,$$

where θ^* is a tunable parameter related to the rate function, and Γ aggregates the expectations of the martingale processes and normalization terms.

An analogous expression is obtained for the uVR flows, with the only difference being the absence of the access subsystem randomness, resulting in a different normalization factor Γ_1:

$$\mathbb{P}(W_{uVR}(n) \geq k) \leq e^{-\theta^* k K_s} \Gamma_1.$$

To conclude the e2e delay characterization for the tandem system, the analysis incorporates the deterministic delay introduced by the transmission subsystem. For MT flows, this results in a reliability expression:

$$\mathcal{R} = 1 - \mathbb{P}(W_{MT}(n) \geq \hat{k}_0), \quad \text{with } \hat{k}_0 = k - \tau,$$

where τ denotes the transmission delay.

For uVR flows, both access and transmission subsystems contribute a constant delay τ, leading to:

$$\mathcal{R} = 1 - \mathbb{P}(W_{uVR}(n) \geq \hat{k}_1), \quad \text{with } \hat{k}_1 = \beta - 2\tau,$$

where β is the perception threshold of the user.

C. Analysis for EDF Policy

Under the Earliest Deadline First (EDF) scheduling discipline, the computation subsystem prioritizes requests based on the smallest remaining deadline. Following the formulation in [20], [21], the service available to a tagged flow i is expressed as:

$$S^i = \left[S^{Tot}(m,n) - \sum_{j \neq i} A^j(m, \hat{m} + [d_j - d_i]) \right]_+ \cdot \mathbf{1}_{\hat{n} > x},$$

where d_i and d_j represent the deadlines associated with the uVR and MT flows, and the indicator function accounts for time window constraints.

As in the FIFO case, stochastic bounds on the delay are derived using martingale envelope processes. Specifically, supermartingale approximations are applied to the arrival processes M_{A^i}, $M_{A^{-i}}$, and the service process M_{S^2}, with exponential expressions involving rate parameters K^{A^i} and K_s. The additional offset y accounts for deadline differences between the tagged flow and the remaining flows.

Depending on the relative timing between n and k, two distinct cases are considered:

1) For $n \geq k$ and $n < k - y$, the delay bound for the tagged MT flow is:

$$\mathbb{P}(W^i_{MT}(n) \geq k) \leq e^{-\theta^* k K_s} \Gamma_2,$$

while for a uVR flow it becomes:

$$\mathbb{P}(W^i_{uVR}(n) \geq k) \leq e^{-\theta^* k K_s} \Gamma_3,$$

where Γ_2 and Γ_3 include expectations over the martingale envelopes.

2) For $n \geq n - y$, the delay bounds incorporate additional exponential decay terms due to interference from flows with closer deadlines:

$$\mathbb{P}(W^i_{MT}(n) \geq k) \leq e^{-\theta^* k K_s - \sum_{q \neq i} A^q(n-k) K^{A^q} y K^{A^q}} \Gamma,$$

$$\mathbb{P}(W^i_{uVR}(n) \geq k) \leq e^{-\theta^* k K_s - \sum_{q \neq i} A^q(n-k) K^{A^q} y K^{A^q}} \Gamma_1.$$

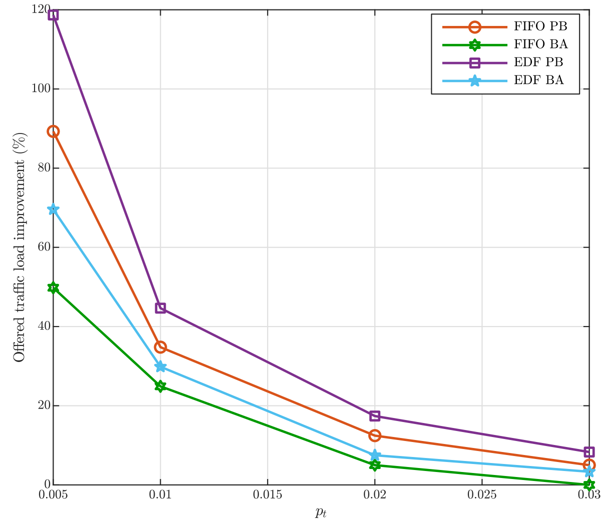

Fig. 3. Offered traffic load improvement as a function of p_t and $p_m = 0.001$.

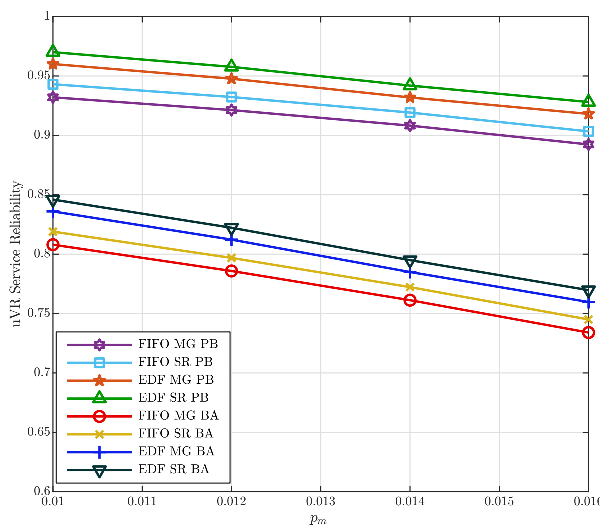

Fig. 4. uVR Service probability as a function of p_m.

As in the FIFO case, the total e2e delay must account for the additional deterministic delays introduced by the access and transmission subsystems. Therefore, the variable k is replaced by the adjusted thresholds $\hat{k}_0$ and $\hat{k}_1$, as defined in the orevious case.

This completes the derivation of e2e delay bounds under EDF scheduling, integrating deadline-awareness into the stochastic analysis framework.

IV. Numerical Results

A performance comparison is carried out between the proposed analytical delay bounds and the outcomes observed

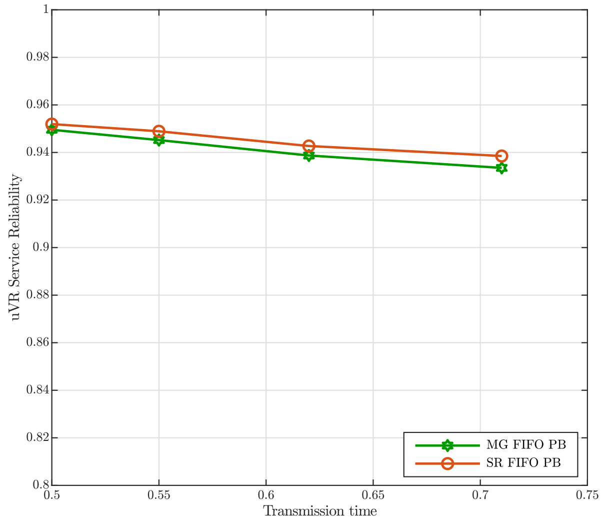

Fig. 5. uVR service reliability as a function of the bandwidth for p_t equal to 0.01 and p_m equal to 0.005 (4 devices).

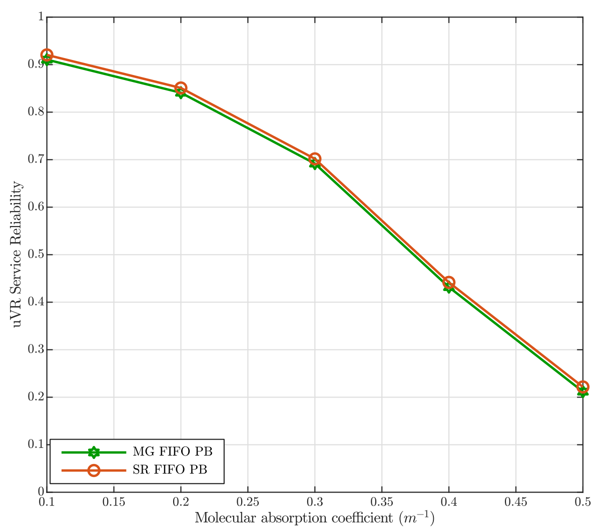

Fig. 6. uVR service reliability as a function of the molecular absorption

through simulation. To emphasize the benefits of incorporating human perception into the e2e delay analysis, a brain-agnostic (BA) baseline is also evaluated. This comparison highlights the accuracy and added value of the human-aware formulation. To contextualize the proposed approach within a realistic setting, we adopt the experimental configuration described in [7], which models a representative uVR scenario. Accordingly, all system parameters are aligned with those specified in [7] and are summarized in Table I.

In the considered scenario, the network spans an area of 20 m $\times$ 20 m and is assumed to be highly dense, ensuring the

TABLE I
SIMULATION PARAMETERS

Parameter	Values
f	$[0.2THz, 1THz]$
$K(f)$	$0.0016m^{-1}$
Transmission power	1 W
uVR packet size	10 Mbits
β	20 ms
$\mathcal{W}$	13 GHz
Water vapor percentage	1%

persistent availability of a line-of-sight (LoS) communication link. The computation time associated with both uVR and MT requests follows a hypoexponential distribution, with a mean of 5.25 ms, while all transmitted packets—whether uVR or MT—are fixed in size and equal to 10 Mbits [7].

For both uVR flows, the evaluation assumes a worst-case condition consistent with the perceptual model shown in Fig. 1, where the user perception threshold is set to 20 ms, while the hard QoS deadline remains fixed at 15 ms. The performance results associated with the perception-based (PB) approach are computed by averaging the numerical outcomes over the distribution of perception thresholds reported in Fig. 1 [1].

Figure 3 illustrates the improvement in traffic load handled by the computation subsystem under a reliability constraint of at least 0.90, considering a fixed MT transmission probability of $p_m = 0.01$. The improvements are quantified in terms of the percentage of uVR traffic successfully supported by the system. As expected, the supported traffic load decreases as p_t increases. The figure clearly demonstrates the superior performance of the PB strategy over the brain-agnostic (BA) baseline, which relies solely on QoS deadlines without accounting for user perception.

Moreover, Fig. 3 shows that adopting the EDF scheduling discipline significantly enhances system performance with respect to FIFO, especially when the analysis focuses on uVR flows. Across all tested configurations, the tight match between the martingale-based analytical bounds (MG) and simulation results (SR) confirms the accuracy of the proposed theoretical framework.

Further validation is provided in Fig. 4, where uVR service reliability is plotted as a function of p_m. The results confirm that reliability decreases as p_m increases due to the higher contention at the access channel. Nonetheless, the PB approach consistently outperforms the BA strategy, highlighting the benefits of incorporating human perception in the delay bound formulation.

The influence of transmission time on uVR reliability is examined in Fig. 5. As expected, increasing the time required to return the computation result to the user leads to a decline in service reliability. This is attributed to the higher probability of violating the perception deadline as the round-trip latency grows.

In addition, Fig. 6 analyzes uVR reliability under varying values of the molecular absorption coefficient, adopting the parameters defined in [7]. The results indicate that molecular absorption significantly affects the data rate in future communication systems, thereby degrading the performance of distributed computing architectures. As absorption increases, uVR reliability deteriorates accordingly.

Figures 5 and 6 also confirm the excellent agreement between MG and SR trends. In both cases, only the FIFO discipline is presented to improve plot readability, due to the specific scale of the y-axis. The performance gain achieved by the PB method compared to the conventional BA approach is consistently observable across all figures, validating its effectiveness in enhancing next-generation distributed computing systems.

V. Conclusions

This paper proposed the formulation of an e2e delay bound, expressed in terms of complementary cumulative probability distribution, for uVR traffic in the presence of heterogeneous and concurrent MT flows. The bound has been derived by combining SNC principles with martingale envelopes, and applied to a cascade system composed of access, computation, and transmission subsystems.

A key contribution of the analysis is the integration of human perception constraints within the delay model, in contrast to classical QoS-based approaches. The study considers the coexistence of multiple uVR flows and MT devices, and evaluates the impact of different scheduling strategies, namely FIFO and EDF, on system performance.

Numerical results confirm the accuracy of the proposed analytical framework and highlight the advantages of EDF over FIFO in managing uVR traffic. Furthermore, the integration of perceptual thresholds leads to significant improvements over the BA alternative, demonstrating the potential of human-aware approaches in optimizing resource usage for future distributed computing systems.

References

[1] A. T. Z. Kasgari, W. Saad, and M. Debbah, "Human-in-the-loop wireless communications: Machine learning and brain-aware resource management," *CoRR*, vol. abs/1804.00209, 2018. [Online]. Available: http://arxiv.org/abs/1804.00209

[2] F. Tang, Y. Kawamoto, N. Kato, and J. Liu, "Future intelligent and secure vehicular network toward 6g: Machine-learning approaches," *Proceedings of the IEEE*, vol. 108, no. 2, pp. 292–307, 2020.

[3] W. Saad, M. Bennis, and M. Chen, "A vision of 6g wireless systems: Applications, trends, technologies, and open research problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, 2020.

[4] M. Chen, W. Saad, and C. Yin, "Virtual reality over wireless networks: Quality-of-service model and learning-based resource management," 2018.

[5] T. S. Rappaport, Y. Xing, O. Kanhere, S. Ju, A. Madanayake, S. Mandal, A. Alkhateeb, and G. C. Trichopoulos, "Wireless communications and applications above 100 ghz: Opportunities and challenges for 6g and beyond," *IEEE Access*, vol. 7, pp. 78 729–78 757, 2019.

[6] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, "The roadmap to 6g: Ai empowered wireless networks," *IEEE Communications Magazine*, vol. 57, no. 8, pp. 84–90, 2019.

[7] C. Chaccour, M. N. Soorki, W. Saad, M. Bennis, and P. Popovski, "Can terahertz provide high-rate reliable low-latency communications for wireless vr?" *IEEE Internet of Things Journal*, vol. 9, no. 12, pp. 9712–9729, 2022.

[8] W.-J. Joo and M. L. Brongersma, "Creating the ultimate virtual reality display," *Science*, vol. 377, no. 6613, pp. 1376–1378, 2022. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.abq7011

[9] J. Zhao, L. Qian, and W. Yu, "Human-centric resource allocation in the metaverse over wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 3, pp. 514–537, 2024.

[10] D. Mourtzis, N. Panopoulos, J. Angelopoulos, B. Wang, and L. Wang, "Human centric platforms for personalized value creation in metaverse," *Journal of Manufacturing Systems*, vol. 65, pp. 653–659, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0278612522001959

[11] M. Xu, W. C. Ng, W. Y. B. Lim, J. Kang, Z. Xiong, D. Niyato, Q. Yang, X. Shen, and C. Miao, "A full dive into realizing the edge-enabled metaverse: Visions, enabling technologies, and challenges," *IEEE Communications Surveys and Tutorials*, vol. 25, no. 1, pp. 656–700, 2023.

[12] T. Zhao, Q. Liu, and C. Chen, "Qoe in video transmission: A user experience-driven strategy," *IEEE Communications Surveys and Tutorials*, vol. 19, pp. 285–302, 2017.

[13] Y. Chen, K. Wu, and Q. Zhang, "From qos to qoe: A tutorial on video quality assessment," *IEEE Communications Surveys and Tutorials*, vol. 17, no. 2, pp. 1126–1165, 2015.

[14] M. Fidler and A. Rizk, "A guide to the stochastic network calculus," *IEEE Communications Surveys Tutorials*, vol. 17, no. 1, pp. 92–105, Firstquarter 2015.

[15] F. Ciucu and J. Schmitt, "Perspectives on network calculus: No free lunch, but still good value," *SIGCOMM Comput. Commun. Rev.*, vol. 42, no. 4, pp. 311–322, Aug. 2012. [Online]. Available: http://doi.acm.org/10.1145/2377677.2377747

[16] B. Picano and R. Fantacci, "Human-in-the-loop virtual reality offloading scheme in wireless 6g terahertz networks," *Computer Networks*, vol. 214, p. 109152, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S138912862200264X

[17] P. J. Burke, "Delays in single-server queues with batch input," *Operations Research*, vol. 23, no. 4, pp. 830–833, 1975. [Online]. Available: http://www.jstor.org/stable/169862

[18] E. Del Re and R. Fantacci, "Performance evaluation of input and output queueing techniques in atm switching systems," *IEEE Transactions on Communications*, vol. 41, no. 10, pp. 1565–1575, 1993.

[19] S. Petkoski, A. Spiegler, T. Proix, and V. Jirsa, "Effects of multimodal distribution of delays in brain network dynamics," *BMC Neuroscience*, vol. 16, pp. P109 – P109, 2015.

[20] B. Picano, "End-to-end delay bound for vr services in 6g terahertz networks with heterogeneous traffic and different scheduling policies," *Mathematics*, vol. 9, no. 14, 2021. [Online]. Available: https://www.mdpi.com/2227-7390/9/14/1638

[21] B. Picano and R. Fantacci, "End-to-end delay bound analysis for vr and industrial ioe traffic flows under different scheduling policies in a 6g network," *Computers*, vol. 12, no. 3, 2023. [Online]. Available: https://www.mdpi.com/2073-431X/12/3/62

[22] D. Williams, *Probability with Martingales*. Cambridge University Press, 1991.