Lightweight Distributed Adaptive Resource Manager for 6G Edge Server Slicing

Eya Arouri, Younghwan Yoo

Pusan National Univ

eyaarouri12@pusan.ac.kr, ymomo@pusan.ac.kr

Abstract

Smart factories and enterprise campuses using private edge networks require efficient resource management for mission-critical services. However, many existing slicing solutions for private 5G/6G edge deployments rely on centralized control or learning-based decision mechanisms, causing management overhead and limited predictability. This paper proposes a distributed, rule-based Lightweight Adaptive Resource Manager (LARM) for private 6G edge-cloud environments that enables dynamic inter-slice resource borrowing. Unlike static allocation or URLLC-only borrowing, LARM allows eMBB and mMTC to borrow unused URLLC resources under strict safeguards. Simulation results under a representative industrial traffic scenario show full URLLC SLA compliance while improving eMBB SLA from 65.5% to 89.2%.

Key Words: Slicing, Private edge networks, Resource management, Inter-slice borrowing, Rule-based Manager

Ⅰ. Introduction

The rapid evolution of mobile communication has enabled private and localized edge networks tailored to specific application scenarios [1], [2]. These networks must serve diverse traffic types, including ultra-reliable low-latency communications (URLLC) for safety systems and robotic control, enhanced mobile broadband (eMBB) for video monitoring and AR/VR applications, and massive machine-type communications (mMTC) for sensor networks [1], [3]. Traditional network slicing techniques in 5G systems rely on either strict static reservations or unidirectional borrowing, where only URLLC can preempt resources from eMBB and mMTC, which leads to inefficient resource utilization during load fluctuations even when URLLC demand is low [3]. Although learning-based slicing approaches exist and can be deployed at the edge, they often introduce coordination complexity and reduced predictability for strict URLLC guarantees [5]. To overcome these constraints, this paper proposes a lightweight, distributed manager that combines time-varying allocation, controlled bidirectional borrowing, and multi-layer protection mechanisms. Unlike conventional schemes, the proposed approach allows eMBB and mMTC slices to opportunistically borrow unused URLLC resources under strict safeguards, while preserving hard URLLC guarantees through minimum reservations, borrowing limits, and emergency preemption. LARM operates locally at the edge node, making it suitable for future private multi-edge network deployments [4].

Ⅱ. Lightweight Adaptive Resource Manager

A. System Model

We consider a private 6G edge base station with a total bandwidth R divided into resource blocks (RBs). In this work, the term resource refers exclusively to channel bandwidth, excluding computation, storage, and other infrastructure components. The system supports three network slices: URLLC (slice 1), eMBB (slice 2) and mMTC (slice 3). Each slice $i \in \{1, 2, 3\}$ is characterized by a time-varying resource blocks demand $r_i(t)$, a minimum guaranteed quota $q_{\min,i}$, and a maximum borrowing limit $B_{\max,i}$. The objective function is to maximize the overall bandwidth utilization $U(t) = (\sum_i a_i(t)) / R$ where $a_i(t)$ refers to the allocated resource blocks to a slice i, while guaranteeing strict SLA compliance for the URLLC slice and maintaining adaptability to dynamic, unpredictable traffic patterns.

B. Scheduling Algorithm

Let $a_i^{avail}(t)$ denote the available resource blocks (RBs) for slice i at time t. Slice i may borrow up to $B_{\max,i}$ RBs from other slices j, where each slice j is assigned a minimum reserved quota $q_{\min,j}$. The unserved requests may remain pending for at most $T_i^{\max}$ before being dropped.

The LARM scheduler operates as follows:

1. $r_i(t)$ are first served using a_i^{avail} (t) when sufficient and $a_i(t)$ are released after the service duration Δ_i.

2. If a_i^{avail} (t) are insufficient, slice i may borrow RBs from other slices j, provided that $q_{min,j}$ and $B_{max,i}$ are both respected.

3. For eMBB and mMTC traffic, if $r_i(t)$ is partially served after borrowing, the remaining portion is queued and retried until resources become available, otherwise the request is dropped after waiting time T_i^{max}.

4. If an URLLC request cannot be satisfied through $B_{max,i}$, the resource manager preempts previously borrowed resources from eMBB and mMTC slices to fully meet the URLLC demand.

Ⅲ. Simulation

A. Evaluation Setup

LARM is evaluated using a discrete-event simulator over 100 s with a 20 MHz bandwidth and 15 kHz subcarrier spacing, giving 111 RBs. Initial slice allocations are 30% (URLLC), 50% (eMBB), and 20%(mMTC), with q_{min} of 16, 27, and 11 RBs, respectively, and B_{max} limited to 20 RBs/s. URLLC traffic is generated by multiple devices transmitting small payloads (0.1 – 10 kB) at quasi-periodic intervals, eMBB traffic consists of long-lived sessions with large payloads (500 – 3300 kB), and mMTC traffic is modeled as periodic sensor reports with small jitter and lightweight payloads (0.05 – 5 kB). Unserved eMBB and mMTC requests are queued for at most T^{max}= 1 s and T^{max}= 5 s before being dropped, while URLLC requests are never queued and may trigger emergency preemption. Performance is evaluated using SLA satisfaction and dropped requests per slice, and compared against Static Allocation and URLLC-only Borrowing.

B. Results and Analysis

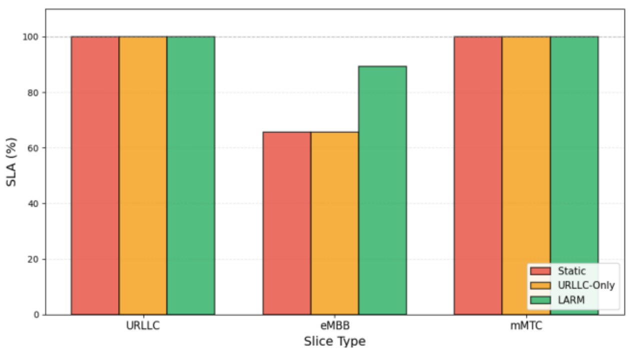

Figure 1. SLA comparison across network slicing methods

Figure 1 demonstrates that all evaluated schemes fully protect URLLC traffic, achieving 100% SLA compliance. This confirms that LARM does not violate mission-critical services.

Table 1. Dropped requests comparison under different allocation schemes.

Slice	Total Requests	Dropped (S /U)	Dropped (LARM)
URLLC	41,863	0/0	0
eMBB	7	2/2	0
mMTC	171	0/0	0

As shown in Table 1, LARM significantly improves eMBB performance compared to static allocation and URLLC-only borrowing by reducing dropped requests and increasing the SLA from 65.5% to 89.2%. The mMTC slice remains fully satisfied in all cases. Overall, these results indicate that controlled bidirectional borrowing improves resource utilization while maintaining strict URLLC SLA.

IV. Conclusion

This paper introduced LARM for private edge networks that enables controlled bidirectional resource borrowing. By combining adaptive allocation with protection mechanisms such as minimum guarantees, borrowing limits, and emergency preemption, the proposed approach maintains full URLLC SLA compliance while improving overall resource utilization compared to the other schemes. Since all decisions are made locally at the edge without centralized control, LARM is well suited for practical private deployments. Future work will extend this study to multi-base station scenarios with coordination between edge nodes.

REFERENCES

[1] M. A. Al-Shareeda et al., “Network Slicing in 6G: A Strategic Framework for IoT in Smart Cities,” Sensors, vol. 24, no. 13, p. 4235, Jun. 2024.

[2] B. de X. Hesselbach et al., “Intelligent Network Slicing in the Multi-Access Edge Computing for 6G Networks,” in Proc. 25th ICTON, Riga,Latvia, Jul. 2023, pp. 1– 5.

[3] Y. Zhang et al., “A Comprehensive Survey on Resource Management in 6G Networks,” IEEE Commun. Surveys Tuts., vol. 26, no. 3, pp. 1850– 1885, 3rd Quart., 2024.

[4] W. O'Brien, A. Dooley, M. Penica, D. Moore, and E. O'Connell, "Performance Benchmarking of Private 5G Networks," Comput. Netw., vol. 258, Art. no. 110846, 2025.

[5] F. J. E. N. Andong and Q. Min, "Federated Multi-Agent Reinforcement Learning for Privacy-Preserving and Energy-Aware Resource Management in 6G Edge Networks," arXiv:2509.10163, Sep. 2025.