

Towards Resilience: Slice-Driven Orchestration of Mobile IoMT in 6G-Edge-Continuum

Muhammad Zeeshan
Centre for Wireless Communications
University of Oulu
Oulu, Finland
muhammad.zeeshan@student.oulu.fi

Hafiz Faheem Shahid
Centre for Wireless Communication
University of Oulu
Oulu, Finland
faheem.shahid@oulu.fi

Erkki Harjula
Centre for Wireless Communication
University of Oulu
Oulu, Finland
erkki.harjula@oulu.fi

Abstract—Emerging healthcare services, such as real-time diagnostics, remote health monitoring and computer-assisted surgery, require stringent Quality of Service (QoS) guarantees while operating under limited and heterogeneous edge resources. Existing Internet of Medical Things (IoMT) orchestration methods primarily focus on communication metrics such as latency and bandwidth, while overlooking edge resource constraints and dynamic adaptation to workload and user mobility. This paper presents a novel resource-slicing-based adaptive orchestration framework that integrates 5G network slicing (NS), resource-aware orchestration, and mobility-aware task offloading into a single adaptive system. IoMT services are assigned to distinct 5G network slices while jointly considering CPU capacity, memory availability and amount of storage, enabling more efficient service placement and execution. In addition, the system supports user mobility and dynamic task offloading to maintain service continuity under changing network conditions. SimPy-based simulation demonstrates that the proposed method enhances service deployment efficiency, reduces request blocking, and improves user acceptance while maintaining reliable performance across diverse IoMT scenarios.

Index Terms—Network Slicing, Service Orchestration, Quality of Service, Edge-Cloud Continuum, 6G.

I. INTRODUCTION

The increasing adoption of Internet of Medical Things (IoMT) technologies is transforming modern healthcare services through applications, such as computer-assisted surgery, wearable monitoring, telemedicine and diagnostic imaging. Different IoMT applications produce heterogeneous workloads with diverse Quality of Service (QoS) requirements, ranging from ultra-low latency for computer-assisted surgery to massive connectivity for patient monitoring and high bandwidth for medical imaging. Efficient service management remains a significant challenge in distributed computing systems [1].

Traditional cloud-centric IoMT applications suffer from high latency and are highly dependent on reliable internet connection. Delays can negatively affect time-critical medical services, while failing Internet connection may completely block the service [2]. While edge computing brings computation and storage resources closer to users end to improve responsiveness, edge nodes usually operate under limited and

heterogeneous resource availability. In practical healthcare environments, changing network conditions, dynamic workloads, and user mobility further complicate service orchestration and resource management across distributed edge environments.

Network Slicing (NS) supports heterogeneous IoMT workloads with guaranteed QoS over a single physical infrastructure in 5G and beyond networks. In healthcare scenarios, ultra-Reliable Low-Latency Communication (uRLLC) slice supports critical-care services, massive Machine-Type Communication (mMTC) slice enables large-scale wearable patient monitoring, and enhanced Mobile Broadband (eMBB) slice supports high-throughput applications including real-time diagnostic imaging [3]. Combining NS with edge computing enables allocation of communication and computational resources for IoMT applications more efficiently [4].

Existing orchestration approaches mainly consider communication-level QoS metrics such as latency and bandwidth, while giving limited attention to computational resource constraints, user mobility, and adaptive service deployment across distributed edge nodes. These gaps affect service continuity and reliability for life-critical IoMT applications, highlighting the need for an adaptive orchestration framework capable of jointly managing network slicing, computational resource allocation, load balancing and mobility-aware task offloading in distributed edge environments. However, a comprehensive IoT service orchestration architecture in a multi-tier computing continuum, including IoT-cloud, MEC, fog computing, and edge-cloud computing, has been analysed in [5].

The main contribution of this paper is the development of a resource-slicing-based adaptive orchestration framework that integrates 5G network slicing (NS) with resource-aware orchestration [6] to support diverse IoMT QoS requirements. In addition, the proposed framework incorporates mobility-aware task offloading to maintain service continuity under user movement and changing network conditions. Finally, SimPy-based simulation demonstrates improvements in service deployment efficiency, resource utilization, and application acceptance ratio compared to a baseline approach. The remainder of this paper is organized as follows: Section II reviews related work, Section III presents the proposed framework, Section IV discusses results, and Section V concludes the paper.

This work was supported by the 6G Flagship (Grant Number 369116) and 6GESS (Profit6 Grant Number 336449) funded by the Research Council of Finland.

II. RELATED WORK

A. Network Slicing in 5G Networks

NS is a key enabling technology in 5G networks, allowing the same physical infrastructure to be divided into logical slices, configured according to the latency, reliability, and bandwidth needs [4]. In healthcare systems, this capability is important because IoMT applications generate highly diverse traffic patterns. For instance, emergency healthcare services require very low latency and reliable communication, whereas wearable monitoring devices mainly require support for massive connectivity. Similarly, medical imaging and remote consultation services depend on high data throughput. As the number of connected IoMT devices increases, maintaining stable QoS and efficient utilization of network resources requires intelligent and adaptive slice management mechanisms [7].

Previous studies have explored different resource allocation approaches for network slicing in 5G systems. Reservation-based methods provide stronger isolation between slices and improve service reliability, but they may reduce overall resource efficiency because resources remain reserved even during low utilization periods. In contrast, share-based approaches improve utilization by allowing statistical sharing of resources among slices, although this can weaken isolation and affect QoS guarantees during congestion [8]. Loven et al. in [9] introduce the concept of AI-driven semantic slicing to optimize the computational resource utilization to deploy IoT applications in the computing continuum with 6G networks.

However, most existing studies mainly focus on communication-level QoS parameters such as bandwidth and latency, with limited attention to joint management of network and computational resources. As a result, there is still a need for orchestration mechanisms that can efficiently coordinate network slicing with edge resource management for IoMT applications.

B. Resource Slicing in Edge Computing

Edge nodes operate with limited processing, memory, and storage capacity, making efficient resource management a critical task for latency-sensitive IoMT applications. Resource slicing (RS) addresses this issue by partitioning network and compute resources into dedicated virtual slices, allocated to specific applications based on their requirements [10]. Recent research has explored combining resource slicing with network slicing in MEC-enabled 5G environments to improve overall system efficiency. These approaches aim to coordinate communication and computational resource allocation so that heterogeneous IoMT workloads can be supported more effectively. While such methods improve scalability and resource utilization, they still face challenges related to dynamic workloads, mobility support, and constrained edge resources [11].

Some studies further propose microservice-based architectures and task offloading strategies to enhance service responsiveness by processing data closer to the edge. This reduces end-to-end latency and enables more flexible utilization of CPU, memory, and bandwidth resources across distributed

nodes [12]. Despite this, most approaches are designed for generic IoT systems and do not fully address the requirements of adaptive orchestration for heterogeneous IoMT services. Joint management of network slicing, resource constraints, and mobility-aware service execution in edge environments remains an open problem.

C. Intelligent Edge Orchestration and Load Balancing

Static resource allocation policies are insufficient for dynamic IoMT workloads, motivating the use of intelligent dynamic orchestration and load balancing in edge computing environments. Existing approaches use AI-based orchestration to distribute workloads between edge and cloud, where delay-sensitive tasks are processed at the edge and compute-intensive tasks are deployed to the cloud data centers, ensuring latency, energy efficiency, and service performance [13]. Recent works also integrate Machine Learning (ML) with Software-Defined Networking (SDN) and NS to enable adaptive node selection and dynamic resource management under changing network conditions. Reinforcement learning methods further improve adaptability by learning optimal allocation policies from historical system states, supporting efficient operation in multi-tenant environments [14].

In parallel, load balancing techniques are used to avoid resource overloading and improve utilization in distributed edge systems. AI-based methods such as swarm intelligence and PSO enable dynamic task distribution across nodes, reducing delay and improving system efficiency [15]. Further studies show that distributed and dynamic load balancing strategies consistently outperform static methods for edge environments [16]. However, most of the existing solutions depend on complex model training, limiting their suitability for real-time IoMT deployments under constrained edge resources. This highlights the need for lightweight and adaptive orchestration approach capable of jointly managing network slicing, resource constraints, and mobility without training dependency.

In summary, existing works address network slicing, resource management, and intelligent orchestration in isolation, without jointly considering all three alongside user mobility in IoMT edge environments.

III. PROPOSED SYSTEM MODEL

A. System Architecture

The proposed system model combines 5G network slicing with distributed edge infrastructure to support adaptive orchestration of heterogeneous IoMT services. It is designed to handle heterogeneous healthcare services with different computation, latency, and reliability requirements under dynamic network conditions and user mobility. The adaptive orchestration framework jointly considers application requirements and real-time resource availability at each BS. Each IoMT application is characterized through application profiling, which defines its QoS, latency, bandwidth, and computational needs. Each BS continuously reports its available resources through resource profiling, capturing the current state of compute and network capacity. Based on these profiles, service requests are mapped

to an appropriate BS and NS that is uRLLC, mMTC, or eMBB. If the nearest BS is overloaded, the orchestrator dynamically redirects the request toward an alternative BS in the same or adjacent coverage region to prevent service degradation and balance load across the network as shown in Fig. 1.

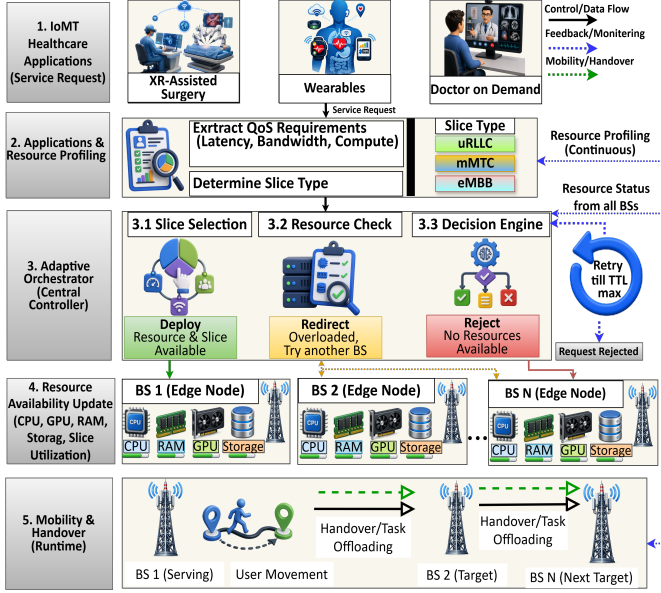


Fig. 1. System architecture for IoMT application orchestration.

To support mobility active sessions are continuously monitored. When a user transitions to a new BS, a handover process is initiated and ongoing services are offloaded to another BS to maintain service continuity. When a specific slice becomes saturated, unused resources from other slices can be temporarily reallocated to maintain service continuity for high-priority applications, reducing request blocking and improving overall system efficiency.

B. IoMT Use Cases

Different IoMT applications are categorized based on their bandwidth, latency, and computational requirements, which enable orchestration based on NS and resource awareness within edge computing environments.

a) *XR-assisted Surgery*: XR-assisted surgery (XS) enables real-time 3D visualization and interaction during surgical procedures. To support reliable operation with strict latency requirements below 20 ms, it is assigned to the uRLLC slice. Substantial processing capacity is required for real-time 3D model rendering and sensor-data processing. Depending on the quality of imaging and XR streams, bandwidth requirements varies between 10 Mbps and 100 Mbps [17].

b) *Wearable-based Health Monitoring*: Wearable health monitoring (WHM) involves in continuous collection of patient data such as heart rate, SpO2, and glucose levels from wearable sensors. It generates low-rate but frequent transmissions from a large number of devices and is mapped to the mMTC slice. The application requires scalable connectivity

with lightweight edge processing. Typically, bandwidth usage remains in range of 1-2 Mbps per device, while delays up to 100 ms are acceptable for continuous monitoring [18].

c) *Doctor-on-Demand*: Doctor on Demand (DoD) provides real time audio and video-based medical consultations between patients and healthcare professionals. it is assigned to the eMBB slice due to its requirement for high BW and stable communication. Bandwidth demand typically varies between 10 and 150 Mbps, with an acceptable latency threshold of 150 ms to maintain smooth and interactive communication [17].

C. Orchestration Algorithm and Workflow

Algorithm 1 describes the deployment procedure for IoMT applications across multiple BSs among different regions. For each application the algorithm identifies a suitable NS and a suitable BS, verifies computational and slice-level resource availability and decides to deploy, offload, or reject the application based on current network conditions.

Algorithm 1 Resource-aware IoMT Service Orchestration in Distributed Base Station Networks

Input:

Service set $\mathcal{W} = \{XS, WHM, DoD\}$
 Slice set $S = \{uRLLC, mMTC, eMBB\}$
 Base Stations $BS = \{bs_0, \dots, bs_X\}$ with
 Capacity(bs) = $\{CPU, RAM, Stor, GPU\}$
 Coverage area = $2000 \times 2000 m^2$,
 $T_{run} = 100 s$, $TTL = Max_{attempts}$

```

1: Get resources from BS  $Res(bs) \leftarrow Capacity(bs), \forall bs \in BS$ 
2: for  $t \leftarrow 0$  to  $T_{run}$  do
3:   for each service  $w \in \mathcal{W}$  do
4:      $N_{attempts} \leftarrow 0$ 
5:      $s \leftarrow AssignSlice(w)$ 
6:     while  $N_{attempts} < Max_{attempts}$  do
7:        $bs^* \leftarrow NearestBaseStation(w, t)$ 
8:       if  $Res(bs^*) \geq ResourceDemands(w)$  and
          CheckSliceAvailability( $bs^*, s$ ) then
9:         Assign  $w \rightarrow (bs^*, s)$ 
10:        Iterate  $Res(bs^*), SliceCap(bs^*, s)$ 
11:        break
12:       else if Alternate  $bs \in BS$  available then
13:         Redirect  $w \rightarrow (bs, s)$ 
14:         break
15:       else
16:          $N_{attempts} \leftarrow N_{attempts} + 1$ 
17:       end if
18:     end while
19:     if  $N_{attempts} = Max_{attempts}$  then
20:       Reject  $w$ 
21:     end if
22:   end for
23: end for

```

Output:

Deployment set D and metrics $\{deployed, rejected, redirected\}$

D. Application and Resource Profiling

Efficient orchestration of IoMT applications relies on two profiling mechanisms. Application profiling captures the requirements of each IoMT application in terms of latency, bandwidth, and computational needs (CPU, GPU, RAM, storage) and maps them to the appropriate 5G network slice as shown in

Table I. Resource profiling tracks real-time compute and communication availability at every BS, maintaining information about its remaining resources and slice-specific bandwidth.

Using both profiles, the orchestrator performs real-time admission and placement decisions. An IoMT application is deployed at a selected BS only if sufficient computational resources are available and the associated network slice can satisfy its QoS constraints. If the nearest BS does not meet these conditions, the orchestrator redirects the request to an alternative BS within the same or neighbouring region. In case of persistent resource scarcity, the request may be delayed or rejected to preserve service quality for running applications.

E. Resource Slicing for IoMT Services

NS in the proposed IoMT framework enables logical separation of the physical network into multiple virtual network slices, each designed to meet specific service requirements. This isolation ensures that heterogeneous applications can run simultaneously without affecting system performance, while maintaining dedicated QoS guarantees. Three main NSs are considered: uRLLC for life-critical services such as XR-assisted surgery, mMTC for large-scale wearable-based monitoring, and eMBB for bandwidth-intensive applications like remote consultation. Each NS is configured with specific latency, bandwidth, and reliability constraints based on the application. The orchestrator uses slice-aware policies jointly with compute resource availability (CPU, RAM, GPU, storage) to assign each application to the most suitable BS and NS.

TABLE I
RESOURCE REQUIREMENTS AND NETWORK SLICE MAPPING FOR IOBT SERVICES.

Service	CPU (Cores)	GPU (Cores)	RAM (MB)	Storage (MB)	BW (Mbps)	Latency (ms)	Slice
Doctor on Demand	6	1-4	4096	2048	10-150	≤ 150	eMBB
XR-assisted Surgery	8	1-8	8192	4096	10-100	≤ 20	uRLLC
Wearable Monitoring	2	-	1024	512	1-2	≤ 100	mMTC

F. Proposed System Configuration and Experimental Setup

The simulation environment is designed in SimPy to evaluate the proposed IoMT orchestration framework in a distributed 5G edge network. The scenario consists of 7 base stations in 2000 by 2000 meters area as shown in Fig. 2, where 500 users move randomly across 3 regions runs for 100 s. The total combined resource capacity across the 7 BSs is: 2300 CPU cores, 2200 GB RAM, 1400 GPU cores, and 1400 GB storage. Each base station has a circular coverage zone. Users inside zone remain connected, while users outside it lose service until they re-enter the BS coverage area. User mobility triggers frequent handovers between base stations, which is used to assess service continuity, especially for delay-sensitive IoMT applications. Each BS is provisioned with distinct computational capacities, reflecting realistic edge deployments, where resources are inherently heterogeneous. Performance is evaluated using metrics such as service acceptance rate, blocking rate, handover frequency, and resource utilization.

1) *Proposed System*: The total available resources across the network are distributed among the base stations, including CPU cores, RAM, GPU capacity, and storage, and are used by the orchestrator for runtime scheduling and offloading decisions. Applications are mapped to appropriate slices based on their QoS requirements. As users move across regions, the orchestrator dynamically handles handovers and, when required, offloads tasks to suitable base stations based on resource availability and network conditions.

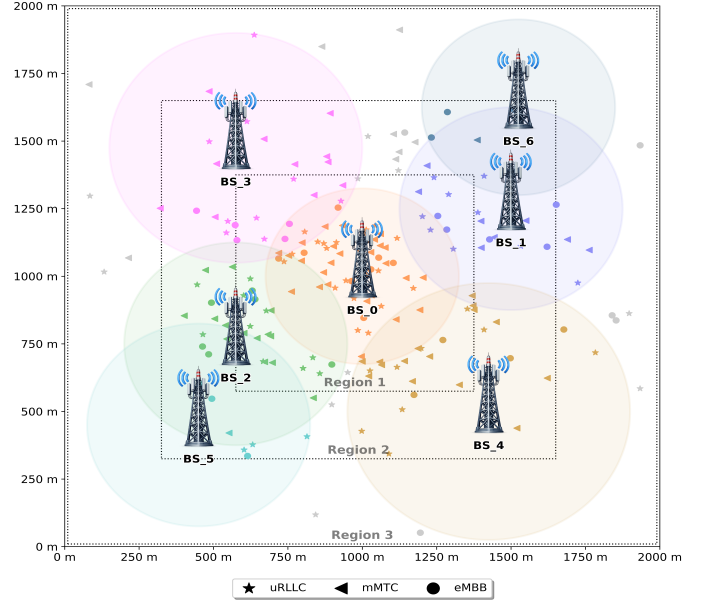


Fig. 2. IoMT users association with nearest base stations across regions in the proposed system.

2) *Baseline System*: Baseline represents an established but non-slice-aware IoMT deployment in which heterogeneous applications compete for shared computing and network resources without network slicing or resource-aware orchestration. The baseline system uses a random deployment strategy without any intelligent orchestration framework and network or resource slicing approach. Applications are assigned to base stations without considering their QoS requirements, computational demand, or network conditions. There is no application profiling or resource awareness in the baseline model, which leads to non-optimal placement decisions and higher chances of network congestion or resource underutilization. However, to ensure fair comparison, both the baseline solution and the proposed framework use identical simulation topology, computational capacity of BSs, user mobility model, and workload characteristics.

IV. RESULTS AND EVALUATION

A. Resource Slicing Behavior in 5G and Beyond Network

The behavior of different 5G network slices demonstrates how the proposed system allocates resources based on application requirements. Fig. 3 shows that the mMTC slice serves the highest number of applications, ranging from 184 to 4457

applications per base station, reflecting its support for massive IoT devices with periodic and lightweight data transmissions. The uRLLC slice serves a moderate number of applications (69–1054), supporting latency-critical services such as XR-assisted surgery, where reliability and delay constraints are stringent. In contrast, the eMBB slice serves fewer applications (48–919) due to the higher demand for data-intensive services, such as doctor-on-demand applications. Overall, the results confirming that the proposed Resource Slicing mechanism effectively isolates traffic, distributes load across base stations, and maintains QoS for heterogeneous IoMT applications.

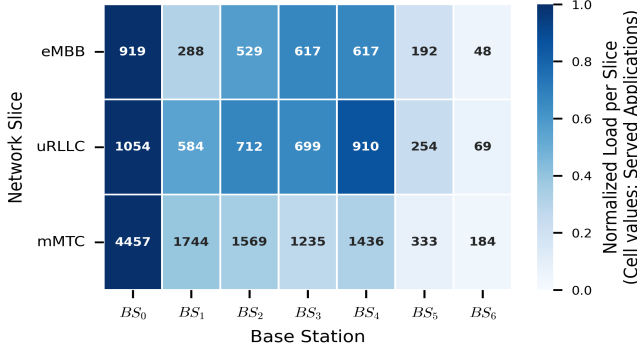


Fig. 3. Cumulative served applications per network slice across base stations.

B. User Mobility Behaviour Across the Regions

The support for user mobility significantly impacts service continuity in the system. As users move across different base stations coverage areas, frequent handover events are triggered to maintain connectivity. Higher handover activity is observed in central base stations located in region 1 due to higher user density and overlapping coverage areas with other BS, while BS located in the region 2 and 3 experience fewer transitions.

TABLE II
BASE STATION HANDOVER EVENTS DURING USER MOVEMENT.

From/To	BS ₀		BS ₁		BS ₂		BS ₃		BS ₄		BS ₅		BS ₆	
	P	B	P	B	P	B	P	B	P	B	P	B	P	B
BS ₀	0	0	31	32	29	28	32	30	38	42	0	0	0	0
BS ₁	18	23	0	0	0	0	0	0	19	12	0	0	6	6
BS ₂	19	15	0	0	0	0	22	16	1	0	10	7	0	0
BS ₃	25	21	0	0	13	13	0	0	0	0	0	0	0	0
BS ₄	25	27	13	9	1	1	0	0	0	0	0	0	0	0
BS ₅	0	0	0	0	2	5	0	0	0	0	0	0	0	0
BS ₆	0	0	4	5	0	0	0	0	0	0	0	0	0	0

P = Proposed system, B = Baseline system

Compared to the baseline, the proposed approach results in more handover counts, as shown in Table II, since decisions are guided by resource availability and service requirements. This leads to improved service continuity for real-time IoMT applications requiring uninterrupted operation during mobility.

C. Computational and Resource Usage Analysis

The proposed framework achieves balanced utilization of computing resources across all base stations. Central nodes handle higher workloads due to increased user density, while BS associated with regions 2 and 3 remain lightly loaded. This distribution reflects realistic edge network behavior.

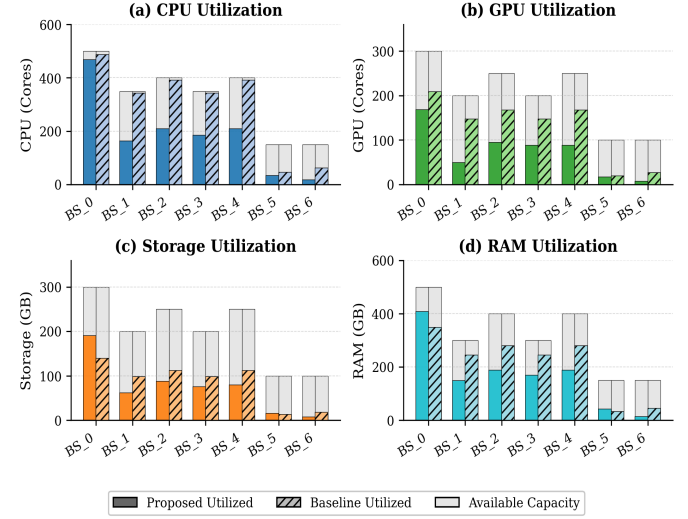


Fig. 4. Comparison of average resource utilization between the proposed and baseline systems for (a) CPU, (b) GPU, (c) Storage, and (d) RAM.

In contrast to using the baseline system, which often leads to uneven resource usage and early congestion at heavily loaded nodes, the proposed orchestration framework distributes tasks based on both slice requirements and available compute capacity (CPU, RAM, GPU, and storage). As demonstrated in Fig. 4, the proposed system avoids overload situations while significantly improving resource efficiency across the network.

D. System Performance Comparative Analysis

The proposed Resource Slicing framework achieves an acceptance ratio of 0.97 vs 0.47 baseline, and reduces blocking from 0.53 to 0.03. As shown in Fig. 5, bandwidth consumption is also reduced due to efficient task placement and avoidance of unnecessary data transmission. Furthermore, the proposed system demonstrates better support for mobility, a higher handover count in the proposed approach is a positive indicator, ensuring seamless service continuity during mobility.

TABLE III
COMPARISON OF NETWORK PERFORMANCE METRICS.

Network Metrics	Proposed	Baseline
Total Users	500	500
Average Connectivity Ratio	0.85	0.59
Average BW Usage (Mbps)	6241	10504
Average Service Acceptance Ratio	0.97	0.47
Average Handover Count	3.08	2.88

Across all evaluated metrics, the proposed Resource Slicing framework surpasses the baseline as shown in Table III,

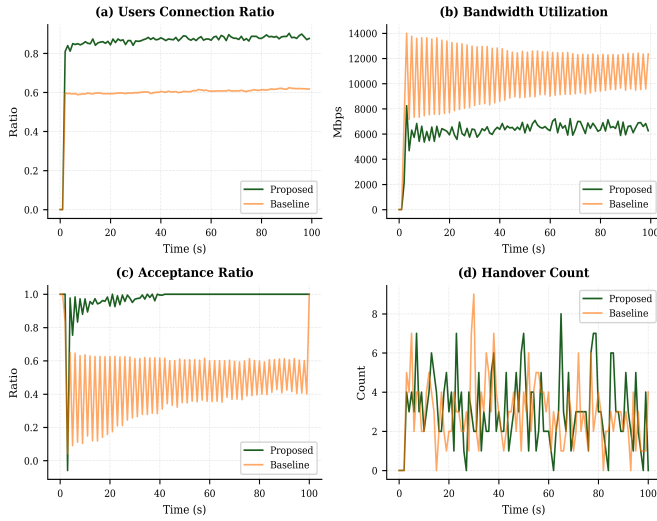


Fig. 5. Performance metrics: (a) Connectivity ratio, (b) Bandwidth utilization, (c) Acceptance ratio, and (d) Handover count.

including service reliability, resource efficiency, and QoS satisfaction. This confirms that combining slice-aware deployment with compute-level resource management improves the performance of distributed IoMT systems in edge environments.

V. CONCLUSION AND FUTURE WORK

In this work, we proposed a novel resource slicing-based adaptive orchestration concept, that integrates network slicing with resource-aware service deployment and mobility-aware task offloading for IoMT edge environments. By jointly considering application QoS requirements, network slice constraints and heterogeneous edge resources, the proposed framework improves load distribution, reduces resource congestion, and enhances service reliability compared to a non-slice-aware baseline approach. These improvements directly benefit healthcare applications. For XR-assisted surgery, lower latency and higher reliability ensure smooth and reliable real-time 3D visualization. For wearable monitoring, improved resource utilization enables scalable patient tracking under high device density. For doctor-on-demand services, higher service acceptance and lower blocking enhance real-time video consultation.

To extend our proposed methodology, future work will compare the proposed framework with state-of-the-art AI-driven orchestration algorithms and real-world IoMT-Edge implementations. However, we aim to address real-world deployment challenges, including network failures, security threats, and energy constraints. AI-assisted and predictive orchestration techniques will be investigated to enable proactive resource allocation, intelligent service migration, and mobility prediction. Furthermore, the framework can further be extended to next-generation 6G networks with ultra-low latency and highly reliable communication requirements.

ACKNOWLEDGMENT

This work was supported by the 6G Flagship (Grant Number 369116) and 6GESS (Profi6 Grant Number 336449) funded by

the Research Council of Finland.

REFERENCES

- [1] U. Islam, M. N. Alatawi, A. Alqazzaz, S. Alamro, B. Shah, and F. Moreira, "A hybrid fog-edge computing architecture for real-time health monitoring in iomt systems with optimized latency and threat resilience," *scientific reports*, vol. 15, no. 1, p. 25655, 2025.
- [2] A. D. Aguru, E. S. Babu, S. R. Nayak, A. Sethy, and A. Verma, "Integrated industrial reference architecture for smart healthcare in internet of things: a systematic investigation," *Algorithms*, vol. 15, no. 9, p. 309, 2022.
- [3] H. F. Shahid, M. Zeeshan, and E. Harjula, "Autonomous orchestration of iomt applications with network slicing in distributed edge computing," *IEEE Access*, vol. 14, pp. 76601–76615, 2026.
- [4] H. F. Shahid and E. Harjula, "Resource slicing through intelligent orchestration of energy-aware iot services in edge-cloud continuum," in *Proceedings of the 14th International Conference on the Internet of Things*, pp. 244–245, 2024.
- [5] H. F. Shahid, B. Akdemir, J. Islam, I. Ahmad, I. Ahmad, and E. Harjula, "Iot service orchestration in edge-cloud continuum with 6g: A review," *IEEE Internet of Things Journal*, vol. 13, no. 10, pp. 20339–20358, 2026.
- [6] H. F. Shahid and E. Harjula, "Resource-aware orchestration of iot applications in edge-cloud continuum with 6g," in *2024 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*, pp. 245–246, 2024.
- [7] T. A. Suleiman and A. Adinoyi, "Telemedicine and smart healthcare—the role of artificial intelligence, 5g, cloud services, and other enabling technologies," *International Journal of Communications, Network and System Sciences*, vol. 16, no. 3, pp. 31–51, 2023.
- [8] A. Banchs, G. de Veciana, V. Sciancalepore, and X. Costa-Perez, "Resource allocation for network slicing in mobile networks," *IEEE Access*, vol. 8, pp. 214696–214706, 2020.
- [9] L. Lovén, H. F. Shahid, L. Nguyen, E. Harjula, O. Silvén, S. Pirttikangas, and M. B. López, "Semantic slicing across the distributed intelligent 6g wireless networks," in *2023 20th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pp. 79–84, 2023.
- [10] S. Husain, A. Kunz, A. Prasad, K. Samdanis, and J. Song, "Mobile edge computing with network resource slicing for internet-of-things," in *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*, pp. 1–6, IEEE, 2018.
- [11] F. Bahramisirat, M. A. Gregory, and S. Li, "Multi-access edge computing resource slice allocation: A review," *IEEE Access*, vol. 12, pp. 188572–188589, 2024.
- [12] J. Hwang, L. Nkenyereye, N. Sung, J. Kim, and J. Song, "Iot service slicing and task offloading for edge computing," *IEEE Internet of Things journal*, vol. 8, no. 14, pp. 11526–11547, 2021.
- [13] H. F. Shahid, J. Islam, I. Ahmad, and E. Harjula, "Optimizing resource-aware service orchestration in edge-cloud continuum," in *2025 IEEE Intelligent Mobile Computing (MobileCloud)*, pp. 44–50, 2025.
- [14] C. Natalino, M. R. Raza, A. Rostami, P. Ohlen, L. Wosinska, and P. Monti, "Machine learning aided orchestration in multi-tenant networks," in *2018 IEEE Photonics Society Summer Topical Meeting Series (SUM)*, pp. 125–126, IEEE, 2018.
- [15] H. F. Shahid and C. Pahl, "Enhanced particle swarm optimisation and multi objective optimization for the orchestration of edge cloud clusters," 09 2019.
- [16] H. Pydi and G. N. Iyer, "Analytical review and study on load balancing in edge computing platform," in *2020 Fourth international conference on computing methodologies and communication (ICCMC)*, pp. 180–187, IEEE, 2020.
- [17] E. Selem, M. Fatehy, and S. M. Abd El-Kader, "E-health applications over 5g networks: challenges and state of the art," in *2019 6th International Conference on Advanced Control Circuits and Systems (ACCS) & 2019 5th International Conference on New Paradigms in Electronics & Information Technology (PEIT)*, pp. 111–118, IEEE, 2019.
- [18] S. O. Ajakwe, C. I. Nwakanma, D.-S. Kim, and J.-M. Lee, "Key wearable device technologies parameters for innovative healthcare delivery in b5g network: A review," *IEEE Access*, vol. 10, pp. 49956–49974, 2022.