Dynamic Traffic Steering for Networked Robotics Using 3GPP-Compliant Application Functions

Binh V. Duong*, Yushan Yang†, Ricardo Pousa*, Frank H.P. Fitzek*‡, Giang T. Nguyen†‡
*Deutsche Telekom Chair of Communication Networks, TU Dresden, Germany
†Haptic Communication Systems, TU Dresden, Germany
‡Centre for Tactile Internet with Human-in-the-Loop (CeTI), TU Dresden, Germany
E-mails: {binh.duong | yushan.yang | ricardo.pousa | frank.fitzek | giang.nguyen}@tu-dresden.de

***Abstract*—Low-latency communication in 5G networks enables advanced applications in networked robotics, such as collaborative visual Simultaneous Localization And Mapping (vSLAM), where robots share real-time sensor data to build accurate maps and localize themselves. To reduce onboard energy consumption, robots can offload compute-intensive tasks—such as feature extraction and map fusion—to nearby Multi-access Edge Computing (MEC) servers. However, existing 5G traffic steering mechanisms often ignore the dynamic compute load at MEC sites, leading to service congestion and degraded Quality of Service (QoS). This paper proposes a novel traffic steering scheme that jointly considers real-time network conditions and MEC compute availability to optimize service instance selection. Leveraging the 3GPP-defined Application Function (AF), our system dynamically associates mobile clients with the most suitable service endpoint, minimizing end-to-end latency and improving load distribution. We outline an experimental evaluation plan including key performance metrics, a 5G-enabled testbed, and representative robotic workloads. The proposed scheme is based on Release 18 and adaptable to future beyond-5G architectures.**

***Index Terms*—Networked Robotics, MEC, vSLAM, 3GPP**

I. INTRODUCTION

Visual simultaneous localization and mapping (vSLAM) enables robots to build maps and localize themselves using real-time video and 3D point cloud data. High-resolution images, depth information, and stereo vision enhance scene understanding but require low-latency communication, high bandwidth, and reliable compute resources [1, 2]. Performing these tasks entirely onboard is often impractical due to high computational and energy demands. To address this, Multi-access Edge Computing (MEC)[1] brings cloud-like resources closer to robots by deploying compute nodes at the network edge, such as base stations and local servers. MEC significantly reduces end-to-end latency, enabling time-sensitive vSLAM tasks like feature extraction, loop closure, and map fusion to be offloaded with minimal delay. However, not all data streams are equally critical. Smart systems must prioritize low-latency control and perception data over non-urgent telemetry. For instance, vSLAM feedback loops must follow the fastest available path, while secondary streams can adapt in quality based on network conditions. 5G supports this differentiation through high data rates, low-latency communication, and a software-defined infrastructure that enables dynamic control of both the data and control planes [3]. These capabilities make 5G and MEC jointly critical for deploying scalable and responsive vSLAM in networked robotic systems.

While traffic steering in the 5G user plane has been extensively studied [4, 5], existing approaches fall short in supporting dynamic, latency-sensitive applications such as 3D collaborative mapping. In these scenarios, robots rely on service instances hosted at MEC sites, e.g., for real-time mapping or control—which must be reached via optimized 5G paths to ensure low latency and high availability. The performance of these services directly impacts the responsiveness and reliability of robotic systems. To support such applications, 3GPP [6] defines traffic steering mechanisms including dynamic session breakout, which selects routes based on real-time network conditions, and pre-established breakout, which uses static paths. However, both mechanisms focus solely on network path selection and neglect the compute load at MEC servers. As a result, traffic can be routed to overloaded instances, leading to degraded Quality of Service (QoS), including increased latency, missed connections, and inefficient resource utilization.

An optimal service instance for 3D collaborative mapping must be capable of handling multiple robots and executing compute-intensive tasks within strict latency bounds. Suitability can be evaluated using metrics such as CPU load, memory usage, queue length, and network conditions [7], which vary by service type. However, current 5G systems lack a standardized mechanism to dynamically select such instances based on real-time MEC workload information. The key challenge lies in integrating MEC resource awareness into 5G traffic steering to jointly optimize network paths and computational availability. This requires dynamic session routing that considers both network state and MEC server load—factors like energy efficiency, CPU utilization, and task queue depth. While 3GPP has outlined architectural support for service instance selection [6], practical realization depends on secure, programmable interfaces between 5G core functions and MEC platforms. In response, this paper proposes a traffic steering framework that leverages the 3GPP-defined Application Function (AF)[8] to enable compliant and secure coordination between the 5G core and MEC. Our architecture comprises: (i) a Release 18-compliant 5G system[9] for low-latency data streaming, (ii) a MEC-based execution environ-

---

[1] https://www.etsi.org/technologies/multi-access-edge-computing

ment for offloading robot workloads, and (iii) a cloud-native AF controller that selects the optimal MEC service instance for each robot in real time. The proposed design is standards-aligned and adaptable to future beyond-5G systems, e.g., 6G.

II. RELATED WORK

A. Traffic Steering for Optimal Service Instances

The advent of 5G networks has introduced advanced capabilities, including Ultra-Reliable Low-Latency Communication (URLLC). A key element of the 5G Service-Based Architecture (SBA) is the Application Function (AF), which enables policy-based service selection and traffic steering for applications deployed on MEC infrastructure. The AF has been extensively studied in 5G-enabled mobile networks, where it plays a crucial role in optimizing service routing and directing data flows to the most suitable compute instances. The authors in [4] proposed a user plane architecture to accelerate traffic steering in mobile networks, while [5] introduced a computing-aware mobile data plane architecture to improve service instance efficiency. Although these studies have advanced traffic management and service optimization, a dedicated architecture design for service metrics and user plane architectures tailored for robotics remains an open challenge.

B. MEC for Networked robotic Workload Distributions

Networked robotic systems have been widely adopted across various domains, such as autonomous exploration, and industrial automation. These systems generate vast amounts of sensor data, such as point clouds and video streams, all of which require efficient processing to enable real-time mapping and decision-making [10]. Traditional local and cooperative mapping approaches rely on onboard processing or decentralized data sharing. However, in large-scale deployments, these methods face significant challenges, including computational bottlenecks, network congestion, and synchronization issues [11]. To overcome these challenges, MEC has been integrated with networked robotics mapping, enabling the offloading of computationally intensive tasks such as vSLAM [12]. While MEC offers dynamic workload allocation, a key challenge remains: selecting the optimal MEC service instance for processing networked robotics data. In large-scale deployments, multiple MEC instances can be available, each with varying levels of computing power, network congestion, and proximity to robots [13].

Existing studies on 5G traffic steering and MEC-enabled robotic workload distribution focus on either network-aware or compute-aware optimization but lack an approach that integrates real-time network conditions with MEC resource availability. This disconnect between network providers and MEC service providers leads to suboptimal service instance selection, increasing latency and inefficiencies in robotic applications. This gap highlights the need for collaboration between MEC and the 5G Application Function to enhance service instance hosting and selection. By leveraging AF, robots can be dynamically assigned to the most suitable MEC service instance, ensuring efficient processing while maintaining low-latency communication.

III. 3GPP-COMPLIANT DYNAMIC TRAFFIC STEERING FOR OPTIMAL MEC SERVICES

Figure 1 describes the proposed system architecture, where the cloud-native AF Controller connects to the 5G Core through Service-Based Interfaces (SBI) and monitors metrics in MEC to determine optimal service instances.

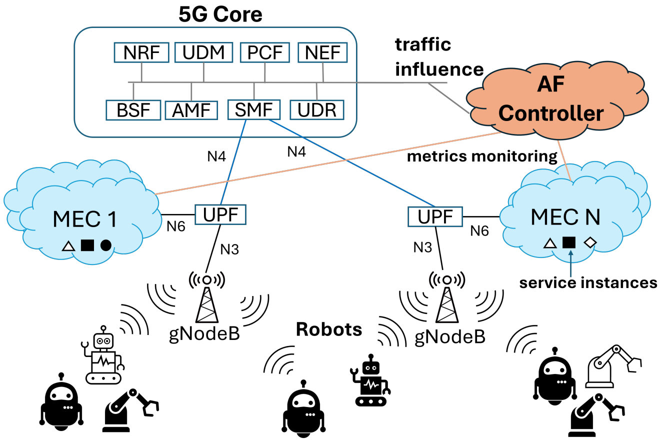

Fig. 1. Proposed system architecture – AF controller orchestrates dynamic traffic steering from UE to optimal MEC service instances.

A. Application Function Controller

Application Function is a network function in 5G through which external application service providers are able to interact with 5G Core to optimize the connection between their application and the devices served. The optimization actions available are Quality of Service (QoS) updates and Protocol Data Unit Session Anchor (PSA) alteration. These actions can be applied per User Equipment (UE) on an end-to-end QoS basis, such that the MEC cloud resources hosting service instances of the application are used efficiently while obeying the application needs.

The AF Controller subscribes to Packet Data Unit (PDU) session events at the Session Management Function (SMF) [8] with specified conditions (e.g., the service instances to which robots are connected, robots' locations, etc.). When conditions are met at the SMF, the AF Controller monitors the collected metrics to identify the optimal service instances for the robots. The AF Controller then triggers a traffic influence, requesting the SMF a PDU session modification process to redirect traffic to the optimal service instance.

This process will take place in two envisioned scenarios: UE mobility induced handovers and MEC cloud resource scarcity. For the UE handover case, the AF subscription to SMF session events notifies it on user plane management events, prompting the recalculation of traffic routes from UE to service instances. In the case of MEC cloud resource scarcity, where application QoS can be infringed, the traffic influence will be initiated from the AF into the 5G Core.

B. MEC Service Instances for Map Fusion Service

In the proposed system, local computation remains responsible for critical real-time tasks, such as obstacle avoidance and individual robot mapping, which require instantaneous decision-making without network dependencies. Each robot processes onboard sensor data to construct local occupancy maps and execute collision avoidance algorithms. These capabilities enable robots to navigate autonomously in their immediate environment while ensuring robust operation even in network-constrained or latency-sensitive scenarios. However, to achieve global environmental awareness where the target service is vSLAM [12, 14], individual maps must be aggregated and fused into a unified representation - a task that exceeds the computational capacity of a single robot. To address this challenge, MEC servers are utilized for networked robotics map fusion, allowing local maps from multiple robots to be synchronized and merged into a globally consistent environmental model. By leveraging the high computational power of MEC, sensor data from multiple robots can be processed in parallel, enhancing the accuracy and efficiency of global mapping.

IV. EVALUATION METHODOLOGY

To assess the effectiveness of our proposed system architecture, we plan to conduct evaluations based on key performance aspects, including latency, service instance selection efficiency, and 3D collaborative mapping performance. The evaluation aims to answer the following questions: *i)* How efficiently does the AF identify and steer traffic to the optimal MEC service instance? *ii)* What is the impact of AF-based traffic steering on network and computing performance? *iii)* How does the proposed system architecture influence real-time mapping accuracy and latency? In the following we elaborate on the scenarios considered for this study.

Network topology: The network topology comprises multiple MEC nodes, each corresponding to a Kubernetes node, enabling dynamic service orchestration and efficient workload distribution. Robots, acting as User Equipment (UE), connect to the MEC environment via a 5G network using free5GC[2] as the 5G Core and UERANSIM[3] as gNodeB. We deploy vSLAM as the primary service across MEC nodes, with each instance running as a Kubernetes (K8s)[4] microservice.

Metrics: We will evaluate the following metrics: *i)* round-trip time (RTT), measuring the time for a network request to travel to its destination and return; *ii)* PDU session modification latency, representing the time needed to redirect traffic to an optimal MEC service instance; and *iii)* AF traffic influence rule creation time, quantifying the delay in generating and enforcing traffic influence rules within the 5G Core for MEC service selection.

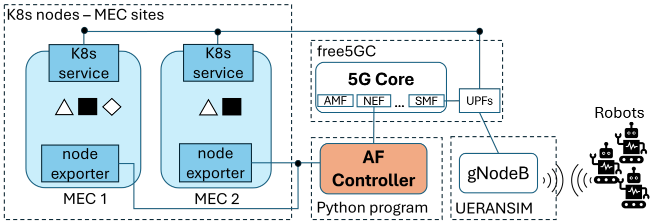

Fig. 2. Testbed design for implementation and evaluation.

Benchmarked algorithm: We adopt a heuristic algorithm, as optimization is beyond this paper's scope and left for future work. The algorithm selects the optimal service instance by comparing CPU and memory usage, applying thresholds to prevent overload. For example, a service instance is deemed suboptimal if either CPU or memory usage exceeds 70%.

Traffic generation: In our testing environment, multiple UEs continuously send request traffic to service instances to measure RTT. Various MEC nodes with different computing capacities and network conditions validate service selection efficiency. The AF, a custom program, leverages Prometheus[5] APIs to identify optimal service instances, supporting different algorithm implementations. Service instance computing status: MEC nodes with varying capacities influence service performance under computational load. AF monitors service instances' status and triggers PDU session modifications in the 5G network to steer traffic to the optimal service. Figure 2 illustrates our system architecture.

Hypothesis: Without our proposal, increasing traffic degrades QoE over time due to suboptimal service quality. Current 5G procedures cannot dynamically modify PDU sessions based on computing resources at service instances or MEC sites. In contrast, our approach enables AF to dynamically trigger PDU session modifications when service quality declines. For example, a service instance may still function but with increased latency, which the standard 5G procedure would not address. Our solution ensures optimal service selection, reducing latency and benefiting ultra-low-latency applications like robot management.

V. SUMMARY

This paper presents a novel traffic steering scheme for 5G-enabled networked robotics, optimizing service instance selection by considering both real-time network conditions and MEC compute availability. While existing 3GPP 5G traffic steering mechanisms disregard dynamic MEC resource availability, our approach leverages the 3GPP-defined AF to intelligently route traffic to optimal service instances, minimizing latency and enhancing service efficiency. As an ongoing study, this work focuses on detailing the proposed method and evaluation plan, which includes key performance metrics, a 5G-enabled testbed, and representative robotic workloads, ensuring adaptability to future beyond-5G architectures.

---

[2] https://free5gc.org/
[3] https://github.com/aligungr/UERANSIM
[4] https://kubernetes.io/
[5] https://prometheus.io/

ACKNOWLEDGMENT

Funded by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) as part of Germany's Excellence Strategy – EXC 2050/1 – Project ID 390696704 – Cluster of Excellence "Centre for Tactile Internet with Human-in-the-Loop" (CeTI) of TU Dresden, the German Federal Ministry of Education and Research (BMBF) project 6G-ANNA, grant agreement number 16KISK102 and the Federal Ministry of Education and Research of Germany in the programme of "Souverän. Digital. Vernetzt." - Joint project 6G-life - projectID: 16KISK001K.

R. Pousa thanks the Horizon Europe under Marie Skłodowska-Curie grant agreement No. 101073265 (EWOC). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them.

REFERENCES

[1] Ben Kehoe et al. "A Survey of Research on Cloud Robotics and Automation". In: *IEEE Transactions on Automation Science and Engineering* 12.2 (2015), pp. 398–409. DOI: 10.1109/TASE.2014.2376492.

[2] Yulan Guo et al. "Deep Learning for 3D Point Clouds: A Survey". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.12 (2021), pp. 4338–4364. DOI: 10.1109/TPAMI.2020.3005434.

[3] Paul Schwenteck et al. "6G Perspective of Mobile Network Operators, Manufacturers, and Verticals". In: *IEEE Networking Letters* 5.3 (2023), pp. 169–172. DOI: 10.1109/LNET.2023.3266863.

[4] Satoru Matsushima et al. *Mobile User Plane Architecture for Distributed Mobility Management*. Internet-Draft draft-mhkk-dmm-mup-architecture-01. Work in Progress. Internet Engineering Task Force, Oct. 2024. 23 pp.

[5] Tran Minh Ngoc and Younghan Kim. *Computing Aware Traffic Steering Consideration for Mobile User Plane Architecture*. Internet-Draft draft-dcn-dmm-cats-mup-04. Work in Progress. Internet Engineering Task Force, Nov. 2024. 18 pp.

[6] 3GPP. *5G; 5G System Enhancements for Edge Computing; Stage 2 (3GPP TS 23.548 version 18.8.0 Release 18)*. Tech. rep. 2024.

[7] Cheng Li et al. *A Framework for Computing-Aware Traffic Steering (CATS)*. Internet-Draft draft-ietf-cats-framework-05. Work in Progress. Internet Engineering Task Force, Feb. 2025. 26 pp.

[8] 3GPP. *5G; Procedures for the 5G System (5GS) (3GPP TS 23.502 version 18.8.0 Release 18)*. Tech. rep. 2025.

[9] 3GPP. *5G; System architecture for the 5G System (5GS) (3GPP TS 23.501 version 18.8.0 Release 18)*. Tech. rep. 2025.

[10] Samih El Nakib et al. "Networking of multi-robot systems: issues and requirements". In: *Int. J. Sen. Netw.* 43.2 (Jan. 2023), pp. 88–98. ISSN: 1748-1279. DOI: 10.1504/ijsnet.2023.134307. URL: https://doi.org/10.1504/ijsnet.2023.134307.

[11] Gerasimos Damigos et al. "Communication-Aware Control of Large Data Transmissions via Centralized Cognition and 5G Networks for Multi-Robot Map merging". In: *J. Intell. Robotics Syst.* 110.1 (Jan. 2024). ISSN: 0921-0296. DOI: 10.1007/s10846-023-02045-4. URL: https://doi.org/10.1007/s10846-023-02045-4.

[12] Peter Sossalla et al. "DynNetSLAM: Dynamic visual SLAM network offloading". In: *IEEE Access* 10 (2022), pp. 116014–116030. DOI: 10.1109/ACCESS.2022.3218774.

[13] Subhramoy Mohanti et al. "L-NORM: Learning and Network Orchestration at the Edge for Robot Connectivity and Mobility in Factory Floor Environments". In: *IEEE Transactions on Mobile Computing* 23.4 (2024), pp. 2898–2914. DOI: 10.1109/TMC.2023.3266643.

[14] Danping Zou, Ping Tan, and Wenxian Yu. "Collaborative visual SLAM for multiple agents: A brief survey". In: *Virtual Reality & Intelligent Hardware* 1.5 (2019), pp. 461–482.