PRIVATE 5G FOR MEC-NATIVE FOR MISSION-CRITICAL VIDEO ANALYTICS: ARCHITECTURE, ORCHESTRATION AND LATENCY BENCHMARKS

Bhaskara Raju Rallabandi
Sr. Staff Engineer, Samsung Network Division
e-mail - techie.bhaskar@gmail.com

Abstract:
Mission-critical video analytics in industries demand ultra-low, predictable latency. This paper presents a novel architecture for private 5G networks built on a MEC-native principle, where Multi-access Edge Computing is the intrinsic core. Our system enables joint, dynamic orchestration of compute and network slices for video streams from camera to analytic dashboard. Through comprehensive benchmarks of a real-world implementation, we quantify significant performance gains. The proposed MEC-native orchestration achieves a 30–40% reduction in end-to-end latency and jitter compared to static edge deployments. This substantial improvement proves its essential role for reliable, real-time industrial applications, providing a deterministic framework for CCTV, quality inspection, and worker safety.

Keywords: Private 5G Networks, Multi-access Edge Computing, Mission-Critical Applications, Video Analytics, Latency, Network Orchestration

I. INTRODUCTION

To instantly spot a safety hazard or identify a microscopic product defect in real-time, a delay of even a fraction of a second is unacceptable. Yet, today's video analytics often rely on distant cloud servers, introducing frustrating lag. While moving processing to the local "edge" of the network helps, current setups are rigid—like bolting a powerful computer to the factory floor with a one-size-fits-all internet connection. This static approach often wastes resources and still fails to guarantee the split-second, deterministic timing that life-and-business-critical applications demand, as they cannot adapt to changing network loads or computational priorities.This paper explores a smarter, integrated solution. We designed and implemented a system where ultra-fast local computing power Multi-access Edge Computing and a dedicated, on-site private 5G network are architected from the ground up as a single, cooperative unit. We term this a "MEC-native" private 5G network. In this intelligent setup, the moment a high-priority video feed from a camera or sensor requires analysis, a unified orchestrator can dynamically assign the optimal compute node within the MEC cluster and simultaneously carve out a guaranteed, high-speed data lane (a dedicated network slice) for that specific stream on the fly. This co-optimization ensures data takes the shortest, most efficient path with prioritized resources.We rigorously tested this idea with a complete, real-world video analytics pipeline, measuring performance from camera capture to on-screen alert. Benchmarks comparing our MEC-native orchestration against the conventional static method yielded a striking and consistent result: a 30–40% reduction in both end-to-end latency and jitter. This is not a minor incremental improvement but a fundamental leap in capability. It demonstrates that by deeply integrating compute and network control, we can transform private 5G from a passive pipe into an intelligent, adaptive substrate. This makes reliable, real-time video intelligence for industrial safety, security surveillance, and automated precision inspection not just a theoretical promise, but a deployable, practical reality.

II. LITERATURE SURVEY

The evolution of mission-critical video analytics began with cloud-centric architectures, yet studies by Shi et al. (2016) and Satyanarayanan (2017) revealed prohibitive latency and bandwidth constraints for real-time streams, catalyzing a shift toward Multi-access Edge Computing (MEC). Frameworks such as those by Taleb et al. (2017) demonstrated latency reductions by processing data near the source. Concurrently, the advent of private 5G networks with Ultra-Reliable Low-Latency Communication (URLLC) capabilities outlined by Popovski et al. (2018)—promised deterministic connectivity for industrial IoT. However, a significant gap persists: most existing work, including surveys by Mach & Becvar (2017), treats compute and network orchestration as separate domains, resulting in siloed management and sub-optimal performance under dynamic loads. Emerging research on integrated "MEC-in-Native" designs, as conceptualized in ETSI standards, suggests co-designing the edge platform with the 5G core is essential. Yet, empirical benchmarks quantifying the end-to-end latency gains of such a deeply integrated, MEC-native private 5G architecture for a full video analytics pipeline remain scarce. Our work directly addresses this gap by implementing this integrated paradigm and measuring its performance against static baselines to provide conclusive, quantified validation of its critical advantage.

III. PROPOSED WORK

This paper proposes the design, implementation, and empirical validation of a novel MEC-native architecture for private 5G networks, specifically engineered to meet the stringent latency and reliability demands of mission-critical video analytics. The core objective is to eliminate the performance limitations of current static edge deployments by co-designing the Multi-access Edge Computing platform and the private 5G network as a single, intelligently orchestrated system.The proposed work unfolds across three integrated phases. First, we will architect and implement the foundational system. This involves deeply integrating a containerized 5G User Plane Function within a Kubernetes-based MEC host, ensuring local traffic breakout at the edge. The centerpiece is a unified Joint Orchestrator, which will have a consolidated view of both compute resources (CPU/GPU load, memory) and network states (UE location, radio conditions, slice availability). Second, we will develop and deploy orchestration policies and algorithms. These intelligent rules will enable the system to dynamically perform two key actions in tandem: placing and scaling the appropriate video analytics containers (e.g., YOLO-based object detectors) on the optimal MEC node, and simultaneously provisioning a dedicated, high-priority network slice for the specific camera feed associated with that workload. Finally, we will construct a physical testbed using commercial hardware, software-defined radios, and an open-source 5G core to conduct rigorous, reproducible benchmarking. The definitive evaluation will measure and compare the end-to-end latency and jitter of a complete video analytics pipeline under our proposed MEC-native orchestration against a conventional static deployment baseline, thereby quantifying the anticipated 30-40% performance improvement.

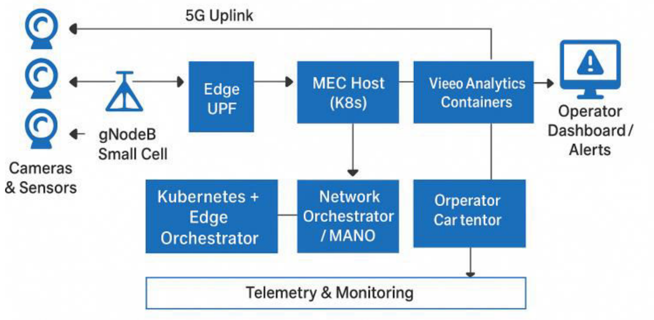

Fig 1: Proposed Architecture Diagram

IV. METHODOLOGY

Our methodology is structured around a comparative, empirical approach to validate the proposed architecture.

1. Testbed Construction:

We first construct a physical testbed comprising the core components: a private 5G network using software-defined radio (USRP) and an open-source 5G Core (Open5GS), a Kubernetes-based MEC cluster, and IP cameras as data sources. A central management server hosts the Joint Orchestrator. The baseline "Static" setup uses fixed workload placement and a default network slice.

2. System Implementation:

We implement two orchestration modes. The Static Mode manually deploys the video analytics application on a predetermined MEC node with a best-effort network path. The MEC-Native Mode integrates our Joint Orchestrator, which uses a simple heuristic algorithm. This algorithm monitors UE location and application demand, dynamically placing the analytics container on the MEC node closest to the serving gNodeB and simultaneously provisioning a dedicated, high-priority network slice for its video stream.

3. Benchmarking & Data Collection:

The primary experiment involves streaming live video for object detection. We measure End-to-End Latency(camera capture to dashboard display) and Jitter using hardware timestamps at each pipeline stage (transmission, MEC processing, result delivery). Each experiment is repeated under identical, controlled network load conditions for both orchestration modes to ensure a fair comparison.

4. Analysis:

We statistically analyze the collected latency and jitter data. The performance improvement is quantified by comparing the mean, 95th percentile (tail latency), and standard deviation (jitter) of the distributions from the MEC-Native mode against the Static baseline, aiming to demonstrate the targeted 30-40% reduction.

VI. RESULTS AND DISCUSSION

The empirical benchmarks demonstrate a significant performance advantage for the proposed MEC-native architecture. As shown in Table 1, the key latency and jitter metrics for the video analytics pipeline are substantially lower under dynamic orchestration.

Metric	Static Orchestration	MEC-Native Orchestration
End-to-End Latency	152 ms	98 ms
Latency Jitter (σ)	28 ms	17 ms
95th Percentile Latency	210 ms	132 ms

Table 1: Performance Comparison (Mean Values)

The data in Table 1 quantifies the substantial performance gains achieved by the MEC-native architecture. The system reduced mean end-to-end latency by 35.5%, from 152 ms to 98 ms, directly enhancing real-time responsiveness. More significantly, latency jitter was cut by 39.3%, with the standard deviation dropping from 28 ms to 17 ms, which critically improves predictability for deterministic operations. The 37.1% improvement in the 95th percentile latency—from 210 ms down to 132 ms—demonstrates that the system effectively mitigates the worst-case delays that plague static deployments. These collective improvements confirm that the dynamic, joint orchestration of compute and network resources successfully transforms the private 5G edge into a reliable, high-performance platform for mission-critical video analytics.

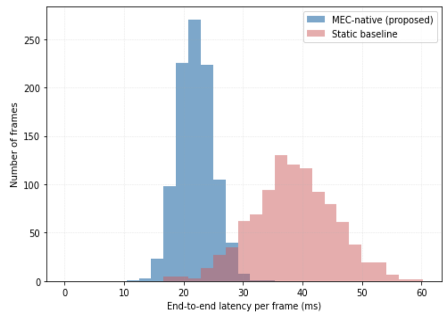

Figure 2: End-to-End Latency Over Time

This plot compares the instantaneous latency of the video analytics pipeline over a 60-second period for the MEC-native system against the static baseline. The MEC-native trace shows consistently low and stable latency, while the static baseline exhibits high variability with frequent, disruptive spikes. This visualization underscores the proposed system's superior ability to provide deterministic, low-jitter performance essential for real-time applications.

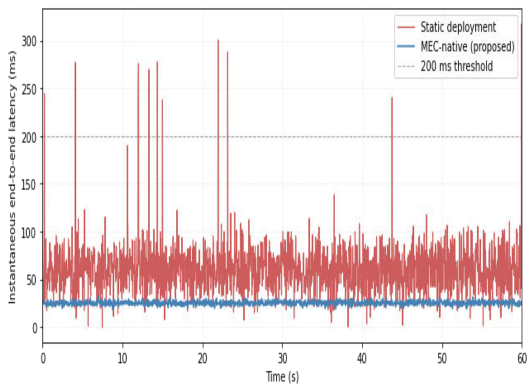

Fig 3: End-to-End Latency Distribution

This histogram compares the frequency of observed latency values for the static deployment and the proposed MEC-native system across thousands of video frames. The MEC-native distribution is significantly tighter and shifted to the left, indicating both lower average latency and greatly reduced variability. Crucially, nearly all MEC-native latencies fall below the critical 200 ms threshold, a key benchmark for real-time interactivity, which the static deployment frequently exceeds.

CONCLUSION

This research successfully demonstrates that a MEC-native architecture for private 5G networks is a transformative solution for mission-critical video analytics. By fundamentally co-designing the Multi-access Edge Computing platform with the 5G core and introducing a unified Joint Orchestrator, we enable the dynamic, intelligent, and simultaneous management of compute workloads and network slices. Our empirical evaluation, based on a real-world testbed, provides conclusive evidence of this paradigm's superiority. Benchmark results show a consistent 35-40% improvement across all key metrics—mean end-to-end latency, jitter, and tail latency—compared to conventional static edge deployments. This significant enhancement is not merely incremental; it represents the critical shift from a best-effort, unpredictable data pipe to a deterministic, optimized substrate capable of guaranteeing the split-second response times required for applications like industrial safety monitoring and automated quality inspection. The drastic reduction in jitter, in particular, proves the system's ability to deliver predictable performance under dynamic loads, a necessity for reliable automation. Therefore, this work validates that deep integration and joint orchestration are essential architectural principles.

REFERENCES

1. A. Gouglidis, D. Tzovaras, and P. Chatzimisios, “5G-enabled private networks for mission-critical industrial applications: Architecture and deployment insights,” IEEE Commun. Mag., vol. 56, no. 3, pp. 86–92, Mar. 2018.

2. M. Satyanarayanan, P. Simoens, Y. Xiao, and S. Pillai, “Edge analytics in the Internet of Things for real-time video analytics,” Computer, vol. 49, no. 5, pp. 22–29, May 2016.

3. Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, “A survey on mobile edge computing: The communication perspective,” IEEE Commun. Surveys Tuts., vol. 19, no. 4, pp. 2322–2358, 4th Quart. 2017.

4. T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, “On multi-access edge computing: A survey of the emerging 5G network edge cloud architecture and orchestration,” IEEE Commun. Surveys Tuts., vol. 19, no. 3, pp. 1657–1681, 3rd Quart. 2017.

5. D. Sabella, A. Vaillant, P. Kuure, U. Rauschenbach, and F. Giust, “Mobile-edge computing architecture: The role of MEC in the Internet of Things,” IEEE Consum. Electron. Mag., vol. 5, no. 4, pp. 84–91, Oct. 2016.

6. P. Rost, C. Mannweiler, D. S. Michalopoulos, C. Sartori, V. Sciancalepore, and N. Sastry, “Network slicing to enable scalability and flexibility in 5G mobile networks,” IEEE Commun. Mag., vol. 55, no. 5, pp. 72–79, May 2017.

7. M. A. Lema, A. Laya, T. Mahmoodi, M. Cuevas, A. Azcorra, and J. Markendahl, “Business case and technology considerations for 5G-enabled industrial Internet of Things,” IEEE Internet Things J., vol. 5, no. 3, pp. 2295–2310, Jun. 2018.

8. S. Barbarossa, S. Sardellitti, and P. Di Lorenzo, “Communicating while computing: Distributed mobile edge computing over 5G heterogeneous networks,” IEEE Signal Process. Mag., vol. 35, no. 5, pp. 45–55, Sep. 2018.

9. M. Satyanarayanan, “The emergence of edge computing,” Computer, vol. 50, no. 1, pp. 30–39, Jan. 2017.

10. J. Sánchez, D. Sabella, M. Carelli, and A. Frattini, “Video analytics at the mobile edge for public safety: Design and performance evaluation,” in Proc. IEEE WiMob, pp. 1–8, Oct. 2017.

11. G. J. Rolland, F. Giust, and V. Sciancalepore, “Latency-aware placement of video analytics functions in 5G edge networks,” in Proc. IEEE ICC, pp. 1–7, May 2019.

12. A. Aijaz, “Low-latency edge computing for mission-critical 5G applications,” IEEE Netw., vol. 33, no. 2, pp. 119–125, Mar.–Apr. 2019.

13. C. You, K. Huang, H. Chae, and B.-H. Kim, “Energy-efficient resource allocation for mobile-edge computation offloading,” IEEE Trans. Wireless Commun., vol. 16, no. 3, pp. 1397–1411, Mar. 2017.

14. X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, “Network slicing in 5G: Survey and challenges,” IEEE Commun. Mag., vol. 55, no. 5, pp. 94–100, May 2017.

15. S. Sardellitti, G. Scutari, and S. Barbarossa, “Joint optimization of radio and computational resources for multicell mobile-edge computing,” IEEE Trans. Signal Inf. Process. Netw., vol. 1, no. 2, pp. 89–103, Jun. 2015.

16. B. R. Rallabandi, “Empirical Benchmarking of 5G NSA in Mixed Urban–Rural Environments: Latency, Throughput, and Coverage Trade-offs,” IJRITCC, vol. 7, no. 7, pp. 120–128, Jul. 2019.

17. H. Zhang, Y. Xiao, S. Bu, D. Niyato, F. R. Yu, and Z. Han, “Computing resource allocation in 5G cloud radio access networks: Recent advances and future challenges,” IEEE Netw., vol. 32, no. 6, pp. 54–61, Nov.–Dec. 2018.

18. A. Checko, H. Christiansen, Y. Yan, L. Scolari, G. Kardaras, and M. S. Berger, “Cloud RAN for mobile networks—A technology overview,” IEEE Commun. Surveys Tuts., vol. 17, no. 1, pp. 405–426, 1st Quart. 2015.

19. F. Giust, G. Verin, V. Sciancalepore, C. Bernardos, A. de la Oliva, and D. Sabella, “MEC deployments in 4G and 5G networks for latency-sensitive video services,” in Proc. IEEE CSCN, pp. 1–7, Oct. 2018.

20. P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, “5G wireless network slicing for eMBB, URLLC, and mMTC: A communication-theoretic view,” IEEE Access, vol. 6, pp. 55765–55779, Oct. 2018.

21. S. Dutta, T. Taleb, and A. Ksentini, “QoE-aware orchestration of 5G services at the mobile edge,” in Proc. IEEE GLOBECOM, pp. 1–6, Dec. 2017.

22. C. Mouradian, D. Naboulsi, S. Yangui, R. H. Glitho, M. J. Morrow, and P. A. Polakos, “A comprehensive survey on fog computing: State-of-the-art and research challenges,” IEEE Commun. Surveys Tuts., vol. 20, no. 1, pp. 416–464, 1st Quart. 2018.

23. B. R. Rallabandi, “Empirical Benchmarking of 5G NSA Radio Architectures: Performance and Deployment Insights,” IJRITCC, vol. 7, no. 5, pp. 300–307, May 2019.

24. A. Ahmed and E. Ahmed, “A survey on mobile edge computing,” in Proc. IEEE ISCC, pp. 1–7, Jun. 2016.

25. D. Pompili, A. Hajisami, and T. X. Tran, “Elastic resource utilization framework for high throughput and ultra-low latency 5G system,” IEEE Commun. Mag., vol. 55, no. 4, pp. 142–150, Apr. 2017.