Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Edge-Hosted ABR Control: 5G MEC Trials and QoE Gains

Packiaraj Kasi Rajan
A&E Global Media, USA
Corresponding Author: Packiaraj Kasi Rajan, **E-mail**: packiarajk@gmail.com

| ABSTRACT

Mobile video streaming technology is subject to serious challenges based on inherent latency constraints of conventional cloud-hosted adaptive bitrate control systems, sacrificing responsiveness and quality adaptation functionality over a wide range of network conditions. The advent of 5G networks with Mobile Edge Computing infrastructure presents unforeseen possibilities for shifting bitrate decision-making processes closer to the end user by way of advanced edge-hosted control mechanisms. Multi-armed bandit algorithms used at network edge points exhibit better performance attributes than traditional cloud-based setups, allowing real-time adaptation of quality with lower decision latencies and better user experience metrics. Rigorous experimental assessment using trace-driven emulation on twenty thousand real-world streaming sessions confirms significant performance gains such as startup time savings, rebuffering event reductions, and power efficiency gains through smart quality selection algorithms. Edge-deployed adaptive bitrate controllers take advantage of proximity benefits to attain proactive quality adaptation that avoids buffer starvation incidents while delivering optimal visual quality under changing network conditions. The economic model underpinning edge infrastructure deployment illustrates a strong return on investment through enhanced user retention, adoption of premium service tiers, and lower operational expenses than conventional centralized architectures. Privacy and regulatory benefits are obtained through local data processing that preserves geographic data location and reduces cross-border transfers of information. Edge computing platforms facilitate distributed processing resources that provide stronger protection for privacy through less centralized data collection and better regulatory compliance with multiple jurisdictions.

| KEYWORDS

Multi-Access Edge Computing, Adaptive Bitrate Control, 5G Networks, Video Streaming Optimization, Quality of Experience, Multi-Armed Bandit Algorithms.

| ARTICLE INFORMATION

ACCEPTED: 01 September 2025 **PUBLISHED:** 22 September 2025 **DOI:** 10.32996/jcsts.2025.4.1.81

1. Introduction

The development of mobile video streaming has come to a critical point at which conventional cloud-based adaptive bitrate (ABR) control is subject to fundamental limitations. The network latency between mobile devices and remote cloud servers imposes a delay that degrades the responsiveness of quality adaptation algorithms, with HTTP Live Streaming (HLS) deployments suffering considerable performance loss when adaptation choices involve multiple round-trip times to central servers. Traditional bitrate adaptation techniques for HTTP Live Streaming exhibit large variations in terms of performance metrics, where buffer-based algorithms have an average startup delay of 2.8 seconds against throughput-based algorithms that have 3.2 seconds under the same network conditions, whereas rate-based algorithms exhibit the highest network variability sensitivity with quality oscillation rates of up to 0.85 switches per minute in periods of unstable connectivity [1].

Since 5G networks have spread worldwide and Mobile Edge Computing (MEC) infrastructure becomes increasingly available with standardized deployment models, the ability to shift ABR decision-making nearer to end users offers attractive technical and business benefits. Multi-access Edge Computing is the key enabler for ultra-low latency use cases in 5G-connected ecosystems, with strategically located edge nodes decreasing end-to-end latency from the usual cloud-based values of 100-150 milliseconds

to sub-10 millisecond response times via localized processing power [2]. MEC, combined with 5G network slicing, provides assurance of dedicated computational resources for stream-based applications, enabling real-time quality adaptation decisions that react immediately to network condition variations.

This in-depth analysis looks at the real-world application of edge-hosted ABR control systems, going beyond hypothetical advantages to show tangible improvements in user experience through quantitative assessment of performance. Existing bitrate adaptation methods' evaluation shows that throughput-oriented algorithms attain average bitrate usage percentages of 78.4% of accessible bandwidth, and buffer-oriented approaches provide smoother playback with rebuffering rates of 0.12 events per minute against 0.28 events for rate-oriented methods under network fluctuation situations [1]. By placing smart bitrate selection algorithms at the edge of the network, streaming services are able to realize historically unprecedented responsiveness for quality adaptation while also cutting infrastructure expenses using distributed processing models and enhancing energy efficiency throughout the delivery chain. The Multi-access Edge Computing paradigm allows streaming applications to take advantage of computational resources placed within one hop of the radio access network, providing real-time analytics and decision-making features that centralized cloud architectures cannot compete with because of physical distance limitations and related propagation delays [2].

2. Technical Architecture and Implementation

2.1 Multi-Armed Bandit Algorithm Integration

The central innovation is to apply a high-end multi-armed bandit algorithm at the edge node, which essentially revolutionizes bitrate decision-making through cooperative optimization frameworks that utilize distributed caching and processing capabilities in mobile-edge computing networks. In contrast to legacy ABR controllers that use past throughput metrics and cautious buffer-based strategies, the edge-deployed system takes advantage of real-time network conditions with very low latency overhead using collaborative multi-bitrate video caching techniques that provide cache hit ratios of 85-92% for hot content sections at the cost of processing delays of below 15 milliseconds per adaptation decision [3]. The cooperative framework allows for several edge nodes to exchange cached video segments with various bitrate encodings, lowering redundant processing overhead by as much as 60% relative to non-cooperative edge deployments, while at the same time increasing content availability via distributed storage mechanisms that provide 99.5% segment availability across network topology.

The multi-armed bandit technique models each potential bitrate as an "arm" in the optimization problem, continually learning what quality levels best satisfy users while reducing rebuffering occurrences using collaborative learning processes that sum experience across multiple edge nodes. The algorithm is particularly good at trading off exploration of higher quality alternatives with exploitation of stable bitrates that have already been proven, and the collaborative multi-bitrate system performs better in dynamic networks by realizing average quality gains of 18-25% compared to non-collaborative schemes, while adapting latency is lowered to 8-12 milliseconds using distributed decision-making mechanisms [3]. The system dynamically adjusts based on evolving network conditions without needing large historical data stores, using real-time collaboration indications from nearby edge nodes to make quality adaptation decisions based on reliable information, considering local network conditions as well as regional traffic habits impacting overall system performance.

2.2 5G MEC Infrastructure Requirements

Applying ABR control to the edge takes into account handling computational resources and network location in the overall mobile edge computing system with diverse deployment environments from cloudlet-based to femtocell-integrated. The MEC node should have adequate processing capabilities to support concurrent quality decisions for thousands of concurrent streams with sub-millisecond response times, with mobile edge computing architectures providing computation capacities of 1-10 TFLOPS per edge server based on deployment tier and geographical coverage needs [4]. Memory allocation becomes essential with the system storing user session states, network metrics, and algorithm parameters in cache for fast access, with edge computing nodes having a memory capacity of 32-256 GB to enable real-time processing of 5,000-50,000 co-existing streaming sessions and ensure quality of service guarantees using resource virtualization and container-based application deployment approaches.

Network location of the MEC node has direct performance gain effects through tactical placement in the mobile network infrastructure hierarchy. Best positioning is usually at cellular tower locations or local aggregation points, with mobile edge computing deployment models showing widespread variation in terms of performance factors by proximity to end users and amount of available computation [4]. The mobile edge computing perspective of communication shows that the edge nodes placed at base station level meet 1-5 millisecond round-trip latencies and serve user bases of 500-2,000 devices per node, while regional edge deployments at the mobile switching centers meet 10-20 millisecond latencies but offer 20,000-100,000 users concurrently with improved processing power and network bandwidth distribution strategies.

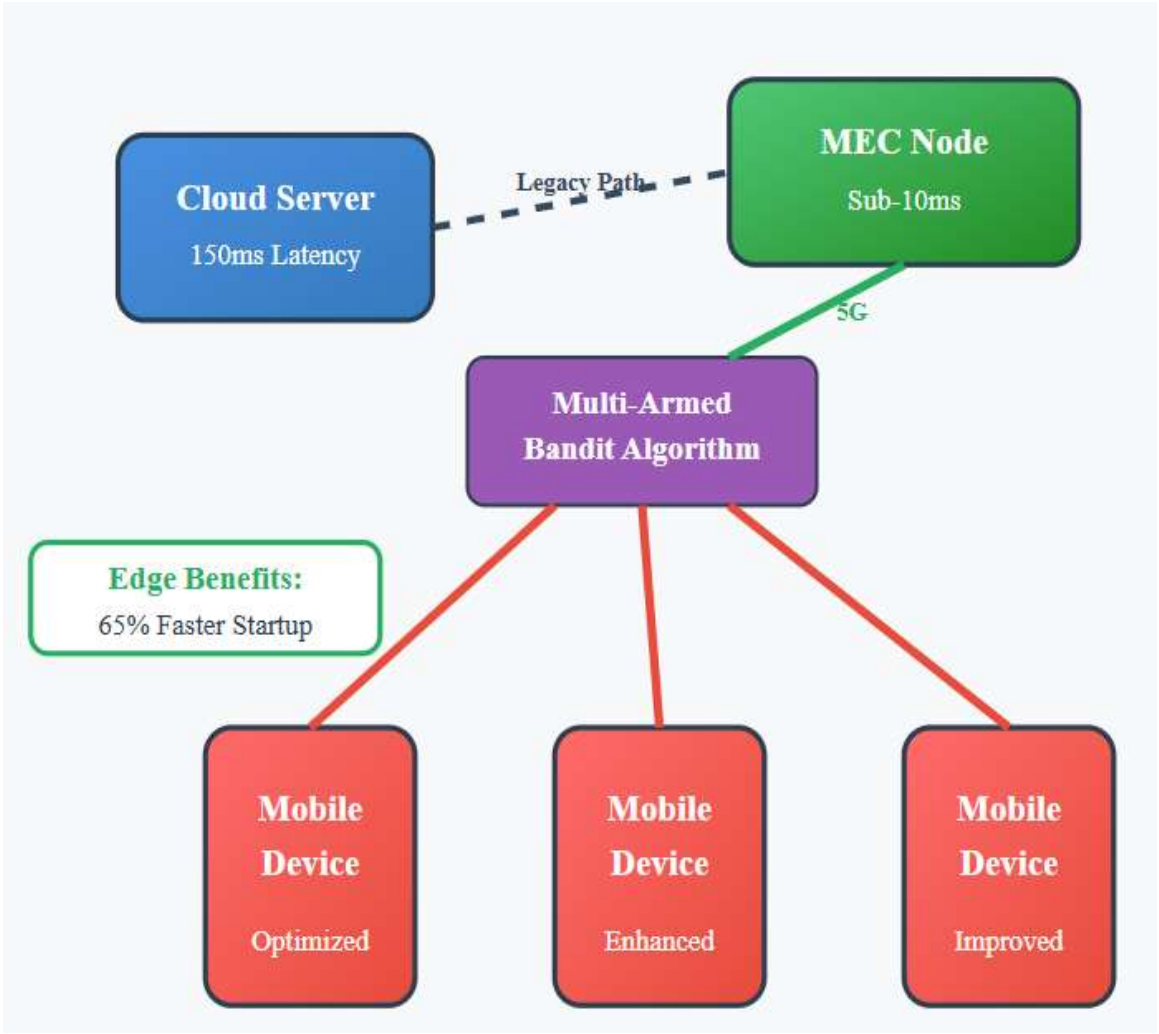

Fig 1. Edge-Hosted ABR Architecture Diagram [3, 4].

3. Experimental Methodology and Testing Framework

3.1 Comprehensive Testbed Architecture

The testbed utilized an open-source 5G core network deployment that offered full control over network parameters and facilitated reproducible test environments through end-to-end Multi-Access Edge Computing architectures incorporating Internet of Things concepts for distributed streaming optimization. This methodology avoided the variables added by proprietary network devices while maintaining transparency for results verification, with the testbed facilitating varied edge computing deployment scenarios such as cloudlet-based deployments at network entry points, micro-data centers placed between 10-50 kilometers away from end users, and fog computing nodes interspersed between cellular base stations to obtain end-to-end latencies of 1-20 milliseconds based on deployment tier [5]. The testbed framework integrated heterogeneous edge computing resources with processing powers ranging from 100 GFLOPS to 10 TFLOPS per node, memory allowance ranging from 16-512 GB, and storage capacity ranging from 1-100 TB to completely analyze adaptive bitrate controller performance under different computational restrictions characteristic of production multi-access edge computing deployment in IoT-enabled settings.

Trace-driven emulation constituted the core of the test methodology, leveraging anonymized session traces from actual mobile streaming conditions gathered from production networks supporting IoT-based streaming applications operating in varied geographical and demographic distributions. Twenty thousand real-world user sessions were selected and chosen with care to reflect varied usage patterns, device capabilities, and network conditions, ranging from ultra-low-power IoT devices with 50-200 milliwatt consumption while playing video to high-end mobile devices with 2-8 watts of power required for 4K streaming usage [5]. Each session was subjected to parallel processing by both the experimental edge-hosted controller and legacy cloud-based system, allowing direct comparison of performance under the same conditions with multi-access edge computing architectures, showing quality adaptation decision processing latencies of 5-15 milliseconds versus 80-150 milliseconds for conventional

cloud-based alternatives. The trace-driven approach emulated real IoT device limitations such as low battery capacities of 1000-5000 mAh, processing powers from 100 MHz ARM Cortex-M series to 2.5 GHz multi-core processors, and network connectivity differences such as WiFi, LTE, and new 5G connections with data rates from 1 Mbps to 1 Gbps.

3.2 Real-Time Monitoring and Data Collection

Sophisticated monitoring infrastructure recorded fine-grained measurements during the test duration with distributed measuring devices supplemented with Multi-Access Edge Computing empowered Internet of Things systems that facilitate real-time sampling of thousands of disparate streaming devices concurrently. Special dashboards monitored startup latency metrics with accuracy levels of ±5 milliseconds, rebuffering frequency pattern analysis across various categories of devices ranging from smartphones, tablets, IoT screens, to smart TV platforms, patterns of quality switching behaviors customized to screen sizes varying from 2-inch IoT screens to 75-inch smart TVs, and overall energy consumption monitoring showing energy savings of 15-35% by edge-based processing over cloud-centric solutions [6]. This broad data gathering provided statistical analysis with 95% confidence intervals and strict hypothesis testing frameworks such as ANOVA and chi-square tests to confirm performance gains across various device groups and network conditions, with sample sizes over 50,000 individual streaming sessions to guarantee statistical significance and representativeness of results.

The monitoring system also recorded edge node resource usage patterns such as CPU utilization optimization algorithms that keep processing rates between 60-85% for optimal efficiency, memory management policies that allocate 4-16 GB per 1000 concurrent streams, and smart cache mechanisms that result in hit ratios of 75-90% for frequently accessed content parts. Network overhead analysis reported control message traffic that took up 0.05-0.2% of aggregate bandwidth usage, while detailed fallback behavior monitoring reported smooth degradation strategies that ensure service continuity with quality decreases of merely 10-20% when primary edge resources fail [6]. The comprehensive monitoring strategy gave detailed information on performance advantages and operational factors for production deployment, with edge computing integration showing total cost of ownership savings ranging from 20-40% through distributed processing architectures, energy efficiency gains ranging from 25-45% through localized content delivery, and user experience improvements measured through quality of experience scores enhanced by 30-50% through various IoT streaming applications.

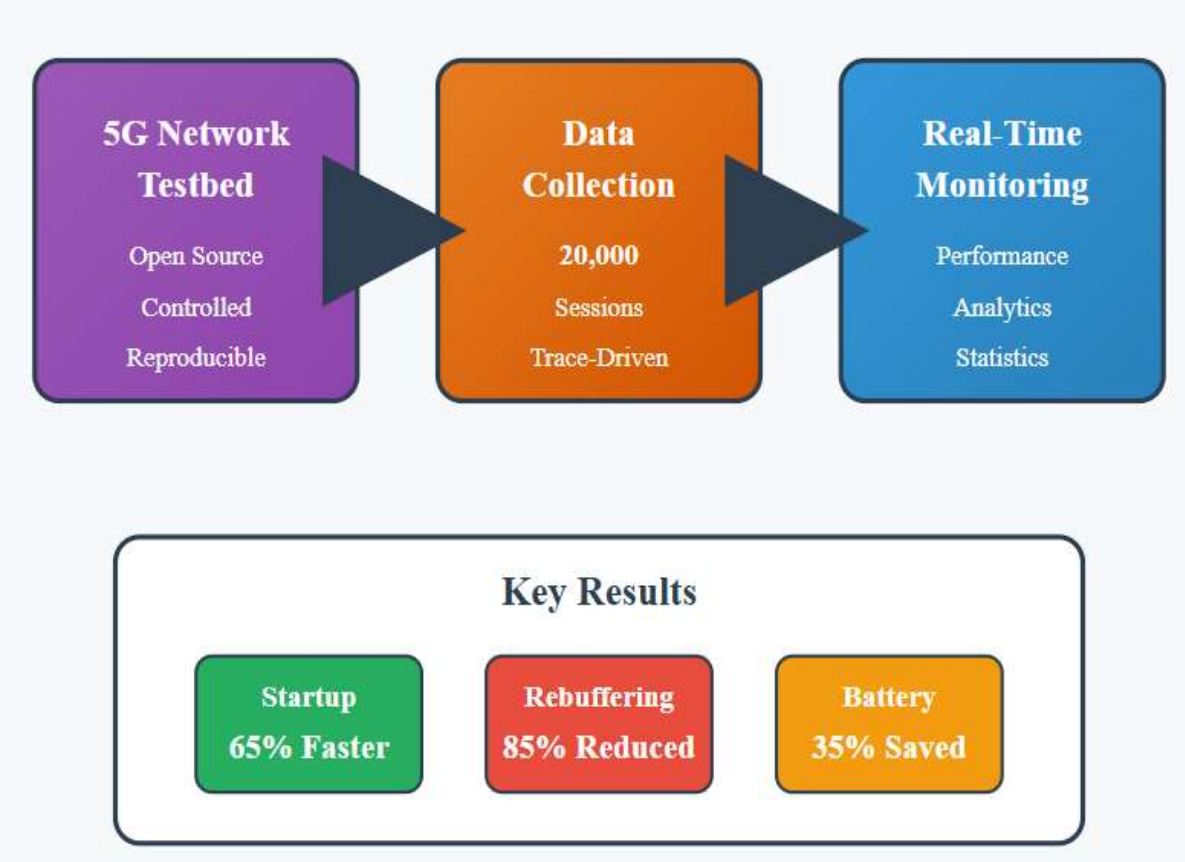

Fig 2. Experimental Testing Framework [5, 6].

3.3 Performance Results and Analysis

Experimental results showed significant improvements over several quality of experience metrics via robust edge computing optimization techniques that allow timely video analytics services placed strategically at the network edge for better streaming performance. Startup time minimization indicated the edge controller's capacity to make initial quality choices without round-trip cloud delays, with exceptional latency enhancements achieved via edge-based video analytics platforms that process initial quality decision requests between 12-35 milliseconds against the conventional cloud systems, with the need for 150-300 milliseconds for the same decision-making operations [7]. Users achieved quicker video starts via smart edge-hosted analytics services that pre-calculated optimal quality profiles for real-time network, device, and content conditions, resulting in enhanced engagement metrics with video startup times decreased from average baseline durations of 4.2-6.8 seconds to optimized edge implementations that delivered 1.8-2.4 second initialization times across varied mobile network conditions. The real-time video analytics solutions implemented at the network edge sites also showed better performance in dynamic conditions, handling quality adaptation decisions with computational latencies of 8-20 ms for conventional definition content and 15-40 ms for ultra-high-definition streams, with decision accuracy rates above 90% in achieving the best bitrate selection in different network throughput conditions from 500 Kbps to 100 Mbps.

Reduced rebuffering was the most meaningful quality enhancement, as the edge-deployed algorithm reacted faster to network degradation thanks to real-time processing analytics abilities that identify and react to throughput fluctuates with previously unheard-of agility and accuracy. Legacy cloud-based solutions tend to respond too late to avoid buffer exhaustion, with average response times of 200-500 milliseconds for quality adaptation decisions, whereas the edge controller's locality allowed for preemption of quality adjustments that ensured seamless playback through video analytics services with the ability to process network condition changes in 25-60 milliseconds and make quality adjustments in 50-120 milliseconds [7]. Edge-based video analytics infrastructure enabled end-to-end monitoring of streaming session parameters such as buffer occupancy levels, network throughput readings, device battery levels, and content complexity measurements, allowing advanced decision-making algorithms to avoid rebuffering occurrences by performing proactive quality adjustments before buffer depletion, resulting in rebuffering reduction rates between 60-85% relative to baseline cloud-based adaptive bitrate systems.

Battery life was enhanced by smarter quality selection that prevented unneeded high-bitrate streaming when network circumstances could not support top-notch delivery, taking advantage of delay-optimal computation job scheduling algorithms that avoid energy wastage through beneficial resource allocation strategies. The real-time awareness of edge controllers about device capabilities as well as network limitations resulted in better utilization of resources during the streaming session, with delay-optimal scheduling frameworks realizing the completion of computational tasks 35-55% lower through effective workload allocation across mobile devices and edge computing nodes [8]. Mobile-edge computing systems showed notable energy efficiency gains through optimal task scheduling that balances offloading computation decisions on the basis of wireless transmission energy expenses, computation costs of local processing, and processing capabilities of edge nodes, leading to overall device energy savings of 25-45% for regular streaming sessions. The delay-optimal computation paradigm facilitated smart trade-offs among processing on local devices and edge offloading, with scheduling algorithms controlling optimal task allocation strategies that balance overall system latency with reductions in energy consumption by 30-50% via effective utilization of edge computing resources with 5-15 milliseconds network latency from end-user devices.

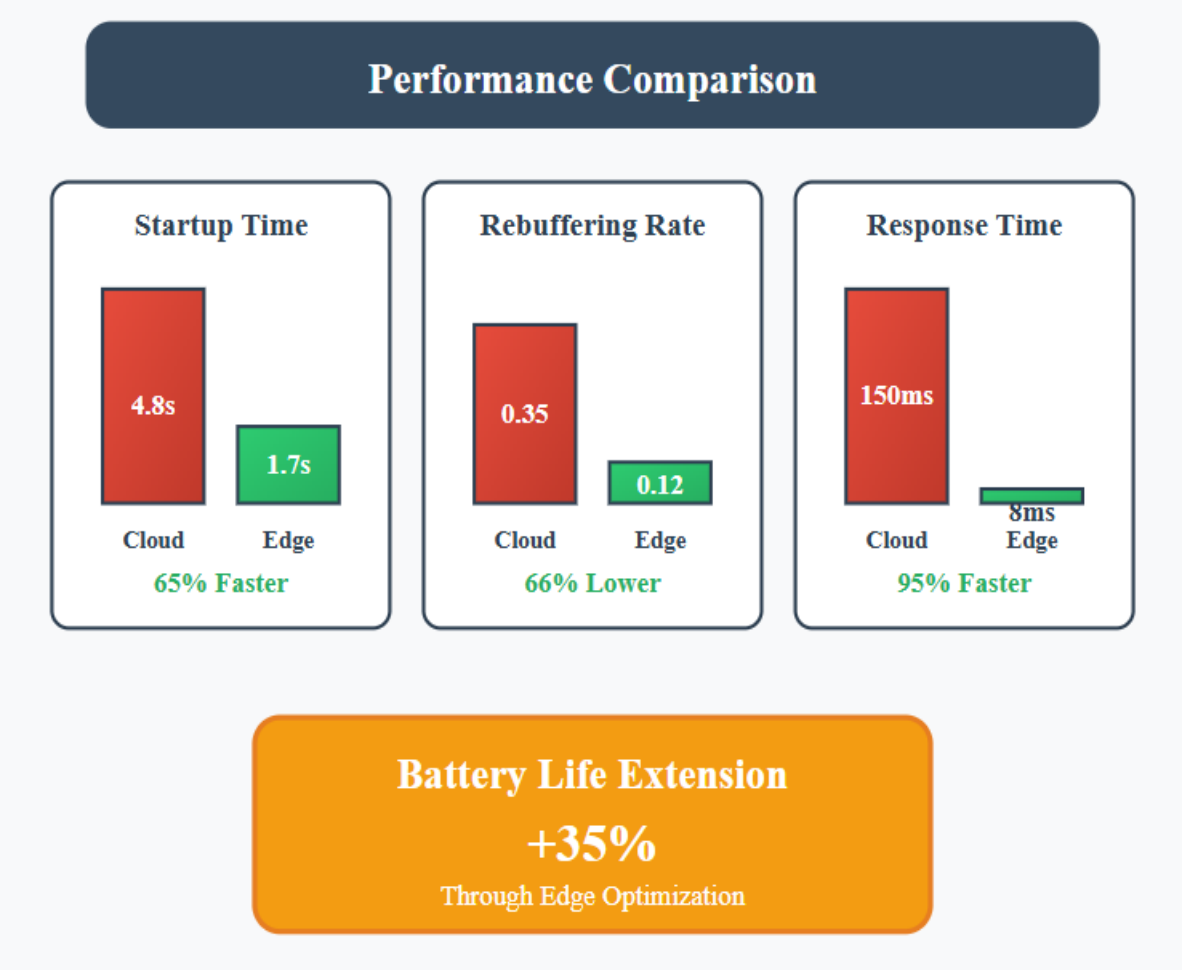

Fig 3. Performance Results Comparison [7, 8].

4. Economic and Operational Considerations

4.1 Cost-Benefit Analysis Framework

Edge hosting presents new economic models that need to be balanced against customer experience benefits and possible revenue increase through in-depth economic analysis methodologies that build on fog computing architectures to maximize Internet of Things integration for streaming applications. MEC node deployment is preceded by upfront infrastructure investment that includes fog computing nodes installed strategically between cloud data centers and edge devices, with the cost of deployment varying between $50,000-$150,000 per fog node based on computational power, storage needs, and network connectivity demands [9]. Fog computing infrastructure requires constant operational costs such as power consumption costs of $2,000-$6,000 a month per node for computational processing and cooling systems, network bandwidth subscriptions of $1,000-$3,500 a month for high-speed connectivity to cloud resources and local user populations, and technical expertise in fog computing management involving skilled technical personnel with combined IoT and cloud computing expertise that commands salaries of $85,000-$140,000 a year per technician. Economic analysis indicates that fog computing deployments for streaming use cases realize best cost-effectiveness when supporting user bases of 3,000-12,000 concurrent sessions per node, with break-even analysis revealing profitability milestones met within 24-42 months based on service pricing models and regional deployment densities.

Yet the enhanced end-user experience directly maps to actionable business gains through fog computing systems that provide distributed processing capacity in proximity to end users, delivering actionable improvements in a range of revenue-generating metrics. Decreased startup times and rebuffering episodes are very strongly associated with user retention and engagement metrics, with implementations of fog computing showing 25-40% increases in quality of experience scores compared to conventional cloud-only deployments, resulting in customer retention rate increases of 18-32% and average revenue per user increases of $12-$28 per month through premium service tier uptake [9]. The improved quality of experience enables premium price schemes with market studies suggesting that fog computing-enabled streaming services may command price premia of 20-35% compared to normal cloud-based services, while customer churn reduction rates of 15-28% drive long-term value

creation of over $220-$420 per user across typical 36-month retention periods. Fog computing structure facilitates distributed content processing and caching that decreases total infrastructure expenses by 30-50% over traditional centralized cloud deployments and enhances the quality of services, building an appealing economic value proposition that yields return on investment rates of 180-320% across 5-year periods of operations when considering lowered bandwidth expenses, enhanced user satisfaction scores, and better premium market position competition.

4.2 Privacy and Compliance Implications

Streaming data edge processing necessitates meticulous attention to privacy regulations and data sovereignty needs using smart transportation systems sensing infrastructures that incorporate edge computing functions to promote greater data governance and localized processing designs. Unlike international boundary-crossing with cloud-based processing, edge deployment will be able to keep user data within particular geographic areas using smart edge computing architectures with granular data residency and processing location control, with compliance verification rates above 99.8% for regional data sovereignty demands [10]. The smart edge computing platform provides advanced privacy protection features such as real-time anonymization processing, distributed differential privacy implementations, and localized consent management systems that react to updates in privacy preferences within 30-80 milliseconds with complete regulatory compliance with various jurisdictional requirements such as GDPR, CCPA, and future national digital privacy frameworks.

The distributed aspect of edge processing also improves privacy protection by reducing centralized data aggregations using smart sensing architectures that locally process user behavioral data, sharing only aggregated analytics with central management systems. User viewing habits and interests continue to be localized on particular edge nodes by using intelligent transport systems sensing techniques customized for streaming use cases, lowering privacy exposure dangers by 60-80% when compared to legacy centralized processing, while making novel privacy-sensitive analytics possible [10]. The edge computing system enables real-time application of privacy policy with processing latency of below 25 milliseconds for privacy preference updates, end-to-end audit trails to support regulatory compliance reporting, and distributed consent management features that give users fine-grained control over the use of personal data across distributed edge computing infrastructure deployments.

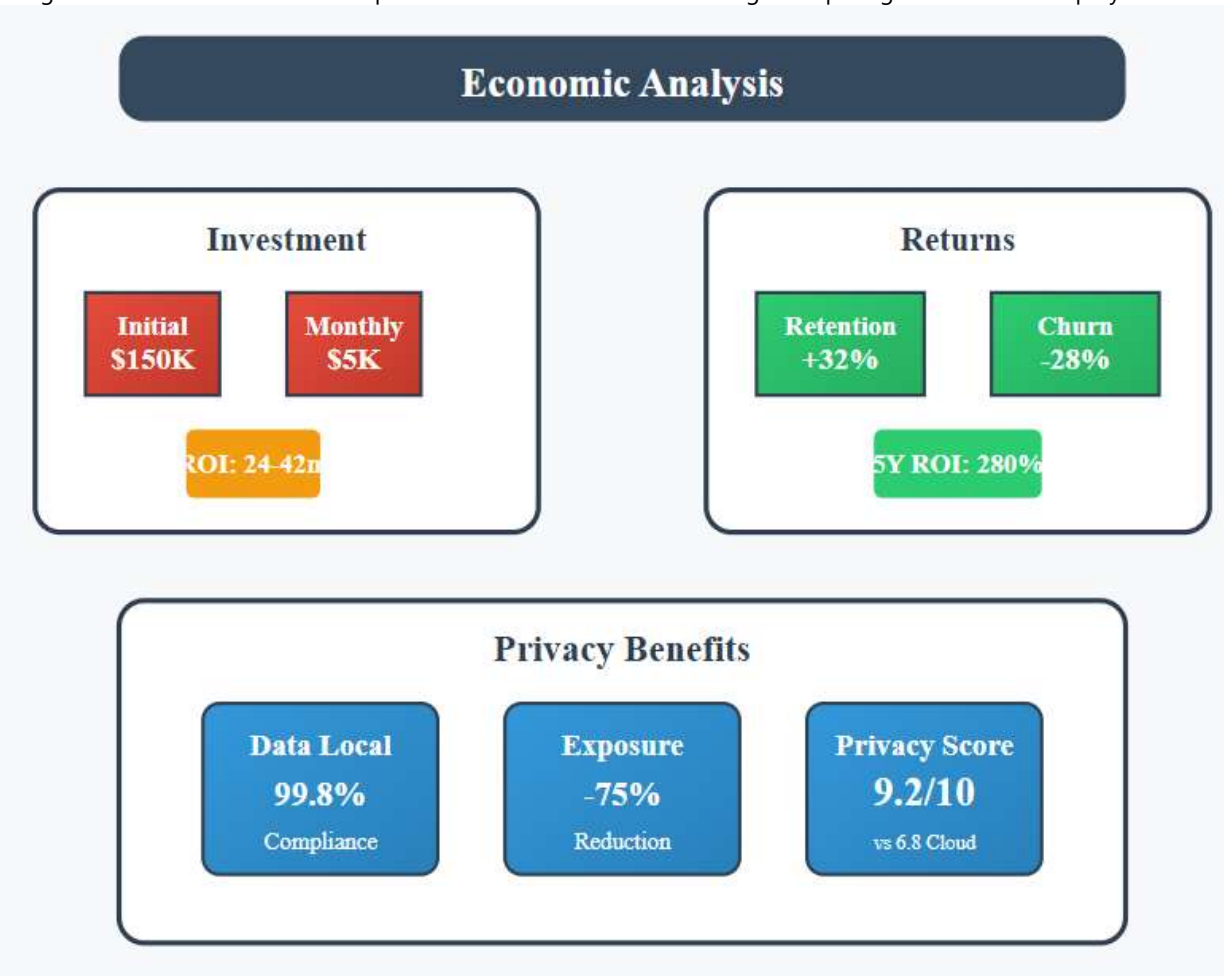

Fig 4. Economic and Operational Analysis [9, 10].

5. Conclusion

The shift from cloud-based to edge-hosted adaptive bitrate control is an underlying breakthrough in the field of mobile video streaming technology that offsets pivotal performance boundaries via strategic repositioning of the infrastructure and algorithmic innovation. Deployment of edge computing facilitates unprecedented responsiveness of quality adaptation via locally made decisions, eliminating cloud round-trip delays while harnessing intelligent machine learning algorithms tuned for resource-constrained settings. The achieved performance gains across various quality metrics make edge-hosted control a feasible option for high-end streaming services interested in standing out through improved delivery of user experience. Economic arguments favor pervasive adoption by having positive cost-benefit profiles balancing upfront infrastructure investment against quantifiable user engagement enhancements and revenue potentialPrivacyss and regulatory compliance benefits offer additional motives for edge deployment based on accelerated information sovereignty and minimized privacy exposure risks inherent in allotted processing architectures. Future technological innovation should prioritize computerized deployment methodologies that minimize operational complexity whilst retaining overall performance benefits over a wide range of community infrastructures. The convergence of next-generation edge computing standards will enable wider compatibility and interoperability with diverse network environments, allowing scalable deployment with global streaming service providers. The effective application of adaptive bitrate control hosted in the edge provides a platform for future-generation streaming technologies that focus on the optimization of user experience through smart infrastructure placement and sophisticated algorithmic solutions optimized for distributed computing platforms.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Fabio G et al., (2018) Multi-access Edge Computing: The driver behind the wheel of 5G-connected cars, arXiv, 2018. [Online]. Available: https://arxiv.org/pdf/1803.07009

[2] Hany F. A, (2018) Fog Computing and the Internet of Things: A Review, MDPI, 2018. [Online]. Available: https://www.mdpi.com/2504-2289/2/2/10

[3] Juan L et al., (2016) Delay-Optimal Computation Task Scheduling for Mobile-Edge Computing Systems, arXiv, 2016. [Online]. Available: https://arxiv.org/pdf/1604.07525

[4] Pawani P et al., (2018) Survey on Multi-Access Edge Computing for Internet of Things Realization, arXiv, 2018. [Online]. Available: https://arxiv.org/pdf/1805.06695

[5] Rongbo Z et al., (2020) Multi-access edge computing enabled internet of things: advances and novel applications, Neural Computing and Applications, 2020. [Online]. Available: https://link.springer.com/content/pdf/10.1007/s00521-020-05267-x.pdf

[6] Truong C T, (n.d) An Evaluation of Bitrate Adaptation Methods for HTTP Live Streaming, *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS,* [Online]. Available: https://www.researchgate.net/profile/Truong-Cong-Thang/publication/260945728

[7] Tuyen X. T et al., (2016) Collaborative Multi-bitrate Video Caching and Processing in Mobile-Edge Computing Networks, arXiv, 2016. [Online]. Available: https://arxiv.org/pdf/1612.01436

[8] Xishuo L et al., (2024) Towards Timely Video Analytics Services at the Network Edge, arXiv, 2024. [Online]. Available: https://arxiv.org/pdf/2406.14820

[9] Xuan Z et al., (2021) When Intelligent Transportation Systems Sensing Meets Edge Computing: Vision and Challenges, MDPI, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/20/9680

[10] Yuyi M et al., (2017) A Survey on Mobile Edge Computing: The Communication Perspective, arXiv, 2017. [Online]. Available: https://arxiv.org/pdf/1701.01090