Optimizing Digital Communication and Data Processing: A Survey of Network Congestion, Frame Synchronization, Latency Compensation, VRR, Edge Computing, Cloud Infrastructure, and Adaptive Streaming

www.surveyx.cn

Abstract

This survey paper provides a comprehensive analysis of technological frameworks and methodologies designed to optimize digital communication and data processing, focusing on Network Congestion, Frame Synchronization, Latency Compensation, VRR, Edge Computing, Cloud Infrastructure, and Adaptive Streaming. As digital ecosystems evolve, managing network congestion and ensuring frame synchronization are crucial for maintaining high-quality multimedia services. Latency compensation, facilitated by edge computing, reduces response times by processing data closer to its source, while cloud infrastructure supports scalable resource allocation. Variable Refresh Rate technology enhances visual performance in dynamic environments, and adaptive streaming adjusts video quality based on network conditions to ensure smooth playback. The integration of edge computing with cloud infrastructure addresses latency challenges, optimizing resource management and enhancing service delivery. Emerging trends highlight the importance of machine learning in optimizing resource allocation and enhancing Quality of Experience (QoE). Future research should focus on developing adaptive resource management strategies and hybrid computing solutions to address the complex challenges of modern digital ecosystems, ensuring efficient and reliable service delivery in increasingly connected environments.

1 Introduction

1.1 Contextual Background

The digital communication landscape is undergoing a significant transformation due to the rapid proliferation of Internet-connected devices and the widespread adoption of the Internet of Things (IoT). This evolution necessitates substantial advancements in network infrastructure to manage increasing data volumes and optimize resource allocation [1]. Transitioning from centralized cloud architectures to Mobile Edge Computing (MEC) is essential for addressing high latency and suboptimal Quality of Experience (QoE) challenges associated with Mobile Cloud Computing (MCC) [2]. In this context, network slicing in MEC-enabled 5G networks enhances resilience and service continuity, enabling operators to create tailored virtual networks for specific service requirements [3].

Integrating MEC with vehicular networks, as explored in Vehicular Edge Computing (VEC), is crucial for addressing the growing demands of communication, computation, and storage in Intelligent Transportation Systems (ITS) [2]. In multimedia services, optimizing edge and cloud infrastructure is vital for improving service delivery and addressing existing knowledge gaps [4]. Furthermore, the surge in IoT devices has escalated network traffic, necessitating efficient communication and data reduction strategies within the IoT-edge-cloud continuum [5].

The advent of 5G networks introduces new challenges, particularly in achieving low latency across Radio Access Networks (RAN), core networks, and caching solutions [6]. The rise of wearable communications highlights the need for a novel communication architecture capable of supporting high capacity, low latency, and massive connectivity in future 5G systems [7]. In this dynamic environment, traffic prediction and energy consumption emerge as critical factors driving the optimization needs of mobile networks [8].

The Industrial Internet of Things (IIoT) aims to enhance productivity, reduce costs, and improve safety in manufacturing through connected devices [9]. However, deploying data stream processing applications in the Cloud-Fog continuum presents challenges due to stringent latency and bandwidth requirements [10]. The limitations of existing methods, which often overlook the internal fabric of edge computing infrastructures, further complicate the effort to minimize end-to-end latency [5].

Additionally, the challenge of real-time data transmission in the metaverse, especially for connected autonomous vehicles (CAVs), underscores the need for optimized digital communication frameworks capable of supporting high data rates and low latency [11]. These factors collectively emphasize the urgency of optimizing digital communication and data processing to meet the evolving demands of modern digital ecosystems, necessitating advancements in algorithm efficiency and scalability.

1.2 Relevance to Current Trends

The rapid evolution of digital communication technologies calls for continuous improvements in both data processing and network infrastructure to accommodate contemporary application demands. The proliferation of 5G networks and their integration with services like YouTube Edge exemplify the increasing need for enhanced latency and bandwidth metrics critical for real-world applications [12]. As mobile edge computing (MEC) environments evolve, refined cache update strategies are essential for improving end-user QoE, addressing the dynamic nature of client demands [13].

The growing multicast nature of data streams further highlights the necessity for advancements in multicast service control, optimizing packet processing and routing to meet the complex requirements of modern digital ecosystems [14]. The widespread adoption of protocols like QUIC in video streaming applications emphasizes the importance of understanding fallback behaviors to ensure seamless performance across diverse network conditions [15].

The surge in live video streaming consumption, particularly during the pandemic, has underscored the critical role of Internet Service Providers (ISPs) in monitoring and maintaining QoE for live streams, driving technological innovation in this area [16]. Moreover, advancements in data processing for large-scale machine learning applications are crucial to meet current efficiency requirements, enabling the handling of massive data volumes [17].

Edge computing frameworks are increasingly relevant, offering significant improvements in efficiency and reductions in energy costs, particularly regarding mobile network traffic prediction [18]. These trends collectively underscore the imperative for ongoing innovation in digital communication and data processing technologies to support the evolving landscape of modern digital applications.

1.3 Challenges and Motivations

The digital communication and data processing landscape is fraught with challenges that necessitate innovative strategies for efficient service delivery. A primary challenge is the dynamic nature of resource demand, requiring efficient provisioning and placement of virtual resources to adapt to changing network conditions [19]. The interdependence of network latency and computing performance complicates this issue, as existing methods often fail to jointly optimize these dimensions, resulting in sub-optimal response times [20]. The computational complexity of current methods exacerbates these issues, leading to slow processing times and inaccuracies [21].

In vehicular edge computing, the high mobility of vehicles results in dynamic network topologies, compounded by limited computational and storage resources, harsh channel environments, and security concerns [2]. The NP-hard nature of the Virtual Network Function (VNF) placement problem poses significant challenges, especially when coupled with the resource limitations of satellites, which restrict their ability to support large-scale IoT services [22]. Additionally, the heterogeneity of the Cloud-Fog continuum complicates resource allocation and application deployment, particularly for latency-sensitive applications [23].

The necessity for substantial reductions in end-to-end latency in 5G networks is critical to support applications such as autonomous driving, telemedicine, and the tactile Internet [24]. Ensuring reliable and efficient communication and computation between far edge devices and cloud services while maintaining low latency and high adaptability further underscores the need for advanced solutions [25]. The diverse communication requirements of wearable devices, which vary by application, present additional challenges in achieving the requisite data rates, latency, and reliability [26].

The curse of dimensionality is another key obstacle, leading to increased model complexity and decreased generalization performance in traditional methods [8]. The tail latency problem remains a significant bottleneck in interactive web applications, adversely affecting application run-time in data-intensive online scenarios [6]. Delays caused by multiple flows and excessive on-device buffering further diminish overall service quality in the Mobile Device Cloud [1].

Efficiently managing network resources during peak traffic times and ensuring Quality of Service (QoS) for diverse new services are critical challenges [3]. These challenges collectively motivate the development of advanced methods to optimize digital communication and data processing, ensuring that modern applications can meet the demands of an increasingly connected world, where efficient and reliable service delivery is paramount.

1.4 Research Scope and Objectives

This survey aims to provide a comprehensive analysis of technological frameworks and methodologies designed to optimize digital communication and data processing, focusing on Network Congestion, Frame Synchronization, Latency Compensation, VRR, Edge Computing, Cloud Infrastructure, and Adaptive Streaming. The scope includes an in-depth examination of fog and edge computing principles and architectures, emphasizing resource management strategies crucial for optimizing performance in latency-sensitive applications. The integration of edge computing as a pivotal solution for managing video frame offloading between local devices and edge servers is investigated, enhancing real-time processing capabilities [2].

The survey explores edge video processing methodologies relevant to multimedia Internet of Things (M-IoT) systems, particularly in smart cities and Internet-of-Vehicles scenarios [2]. It seeks to orchestrate NextG media services efficiently over distributed compute networks, ensuring optimal function/data placement, flow routing, and resource allocation [25]. Furthermore, the feasibility of drone computing offloading in real-world 5G networks is assessed, contrasting the performance of multi-access edge computing (MEC) and cloud solutions for follow-me drone services [24].

The survey also investigates the integration of MEC and software-defined networking (SDN) to enhance resource utilization and management, facilitating collaborative data processing among autonomous vehicles and servers [20]. Additionally, leveraging reconfigurable intelligent surfaces (RIS) to optimize task offloading is highlighted, minimizing delays in edge computing systems [23]. The challenges of measuring model performance in complex scenarios where traditional metrics fall short are addressed, utilizing tailored benchmarks for such purposes [26].

The overarching objective is to propose innovative solutions that address critical challenges in digital communication, ensuring that modern applications can meet the demands of an increasingly connected world. By employing a novel evolutionary algorithm that dynamically adapts the learning rate based on performance feedback, this survey seeks to advance the state-of-the-art in network optimization [25]. A cross-layer solution utilizing multi-agent techniques is proposed to optimize latency by dynamically adjusting network parameters through a hierarchical and independent learning approach [25]. Collectively, these objectives underscore the commitment to advancing the field by addressing critical challenges and proposing innovative solutions for optimizing data processing and network infrastructure.

1.5 Structure of the Survey

This survey is organized to systematically explore the key technological concepts and solutions essential for optimizing digital communication and data processing. It commences with an **Intro-**

AI-generated, for reference only.

duction section, which sets the stage by providing a contextual background and discussing the relevance of current trends, challenges, motivations, and research objectives. Following this, the **Background and Definitions** section elucidates the core concepts, including Network Congestion, Frame Synchronization, Latency Compensation, VRR, Edge Computing, Cloud Infrastructure, and Adaptive Streaming, highlighting their roles in digital communication.

The survey then delves into specific thematic sections, starting with **Network Congestion**, where challenges and innovative solutions such as machine learning-based approaches and active queue management techniques are discussed. This is followed by a detailed examination of **Frame Synchronization**, emphasizing its significance in applications like video streaming and gaming. The section on **Latency Compensation** explores methods to mitigate transmission delays, particularly in edge and cloud computing environments.

Subsequently, the survey addresses **Variable Refresh Rate (VRR)**, analyzing its impact on visual performance in gaming and video playback. The discussion then transitions to **Edge Computing**, focusing on its benefits and challenges in processing data closer to the source. The subsequent section delves into the pivotal role of **Cloud Infrastructure**, particularly focusing on its integration with edge computing technologies. This integration is essential for enhancing data processing and storage capabilities by leveraging the proximity of edge devices to reduce latency, improve bandwidth efficiency, and enable real-time analytics. The synergy between cloud infrastructure and edge computing supports time-sensitive applications, such as those in smart city management and disaster response, while addressing the growing demands for context-aware and scalable solutions amid increasing data generation from IoT devices. Additionally, this integration facilitates better resource management and privacy policy enforcement, ultimately leading to more efficient and responsive cloud services [27, 28, 29, 7, 30].

The survey culminates with a section on **Adaptive Streaming**, exploring its dynamic nature in adjusting video quality based on network conditions. Finally, the **Conclusion** synthesizes the key findings, emphasizes the interconnectedness of the discussed technologies, and suggests future research directions and applications.The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Background and Definitions

Network congestion impedes digital communication, leading to slowdowns and degraded performance, particularly over unreliable wireless links. Traditional congestion control methods, such as those based on the Transmission Control Protocol (TCP), require optimization to ensure efficient data transmission [21]. Effective congestion management is crucial for maintaining optimal performance and service quality in digital ecosystems [31]. In IoT and 6LoWPAN networks, robust mechanisms for congestion detection, avoidance, and control are vital for managing data flow [32].

Frame synchronization is critical for maintaining consistent video frame timing in streaming and gaming applications. It aligns video frames with display refresh rates, reducing visual artifacts and enhancing the viewing experience, particularly with variable refresh rate (VRR) technologies [33, 34, 35, 36]. Precise timing mechanisms are essential for achieving seamless synchronization and improving user experience.

Latency compensation is vital in real-time applications like virtual reality and online gaming. Edge and cloud computing environments employ latency-aware techniques to minimize delays between data requests and responses, enhancing user experience [37]. Edge computing, by processing data closer to the source, significantly reduces latency and improves response times [31]. The three-layer architecture in Vehicular Edge Computing (VEC), comprising vehicular terminals, Road-Side Units (RSUs), and cloud servers, exemplifies effective latency management in dynamic environments [2].

Variable Refresh Rate (VRR) technology is crucial for aligning display refresh rates with content frame rates, reducing screen tearing and enhancing visual smoothness in gaming and dynamic video playback. VRR works alongside adaptive streaming techniques, dynamically adjusting video quality and frame rates based on real-time network conditions to ensure a smooth viewing experience despite fluctuating bandwidth [33, 38, 39].

Edge computing, a distributed paradigm, brings computation and data storage closer to data sources, reducing latency and network congestion [40]. It leverages virtualized, heterogeneous resources accessible with low latency from the network's edge [41]. In 5G networks, Mobile Edge Computing (MEC) enhances user experience by managing dedicated cache space within the radio access network, improving caching services for content providers [42]. The integration of distributed radiance fields with MEC supports video compression and real-time updates for metaverse applications, highlighting edge computing's significance in modern digital ecosystems [11].

Cloud infrastructure provides scalable online resources essential for large-scale data processing and storage. Its integration with edge computing enhances data processing and storage efficiency, meeting the dynamic needs of contemporary digital applications [43]. In 5G networks, cloud infrastructure supports resource management and Quality of Service (QoS) provisioning, critical for modern digital environments [3].

Adaptive streaming dynamically adjusts video quality based on network conditions, ensuring smooth playback and maintaining a high Quality of Experience (QoE). This technology is crucial for media delivery, as it adapts to varying bandwidth and network performance. Content Delivery Networks (CDNs) play a significant role in optimizing video content delivery across distributed networks [38]. The limitations of traditional CDNs in handling increasing mobile video traffic necessitate innovative content delivery methods to ensure timely user access [44].

These core concepts collectively optimize digital communication and data processing, addressing challenges posed by increasing data volumes and the demand for low-latency, high-quality service delivery. The orchestration of these technologies within a Three-Tier Edge-Cloud framework emphasizes dynamic resource allocation and task management, fundamental to enhancing overall communication efficiency [37].

3 Network Congestion

Category	Feature	Method
Challenges in Traditional Congestion Control	Proactive Congestion Management Network Resource Optimization	INCVS[33], NUIA[10] SMC[45]
Machine Learning-Based Approaches	Predictive Modeling Concurrent Data Handling	TSANC[46], Iris[47] DPPF[17]
Flow Differentiation Techniques	Dynamic Resource Allocation	OMSCC[14], D2L-MTP[18]
Active Queue Management (AQM) Techniques	Protocol and Network Strategies Learning and Adaptation Approaches Congestion and Feedback Mechanisms Performance Evaluation Methods	CHOKeD[48], EACCM[49] KRED[50], ADPG[51], QTCP[52] FF[53], RPM[54] PFC-RCM[55]
Innovative Congestion Control Protocols	Independent and Efficient Operations Performance Optimization Techniques Dynamic Traffic Management Specialized Network Control	MVFST[56], MSHAP[57] GHS[58], ABH1TC[59] INCAB[60], MPDCCA[61], IRIS[62] Dart[63]

Table 1: The table provides a comprehensive overview of various methodologies employed in network congestion management. It categorizes these methodologies into traditional congestion control challenges, machine learning-based approaches, flow differentiation techniques, active queue management (AQM) techniques, and innovative congestion control protocols, highlighting specific features and methods associated with each category. This detailed classification aids in understanding the advancements and applications in optimizing network performance and resource allocation.

Addressing network congestion requires exploring methodologies that enhance performance and resource management. This section examines traditional congestion control mechanisms, their limitations, and the need for innovative solutions to meet evolving traffic demands. As illustrated in **??**, the hierarchical classification of network congestion management strategies encompasses traditional challenges, machine learning-based approaches, flow differentiation techniques, active queue management methods, and innovative congestion control protocols. Table 1 presents a detailed classification of network congestion management strategies, illustrating the diverse methodologies employed to enhance performance and resource management. Each primary category is further divided into subcategories, detailing specific methodologies and their applications in optimizing network performance and resource allocation. Additionally, Table 4 illustrates the comparative features of various network congestion control strategies, emphasizing the distinctions between traditional methods, machine learning-based approaches, and flow differentiation techniques. The following subsection details specific challenges faced by traditional approaches, paving the way for an analysis of advancements in this field.

3.1 Challenges in Traditional Congestion Control

Method Name	Packet Prioritization	Resource Management	Adaptability Requirements
INCVS[33]	Irap Packet Prioritization	Minimal Resource Usage	Varying Congestion Scenarios
MVFST[56]	-	Bandwidth And Link	Varying Network Conditions
NUIA[10]	Optimal Time Intervals	Network Resource Availability	Fluctuating Network Resources
SMC[45]	-	Link Rate Schemes	Varying Network Conditions

Table 2: Comparison of congestion control methods highlighting their packet prioritization strategies, resource management techniques, and adaptability requirements under varying network conditions. The table illustrates the diverse approaches and limitations of INCVS, MVFST, NUIA, and SMC in addressing traditional congestion control challenges.

Traditional congestion control techniques face significant hurdles, particularly in packet prioritization and efficient routing under high traffic loads. Inadequate prioritization of crucial Intra Random Access Point (IRAP) packets often results in quality degradation during packet loss [33]. The synchronous nature of current reinforcement learning (RL) frameworks exacerbates these issues by introducing decision-making delays, leading to under-utilization of available bandwidth [56]. The rising demand for mobile services, coupled with limited 5G network resources, poses additional challenges, potentially leading to congestion and performance degradation [3]. This gap underscores the necessity for novel approaches to manage message flows effectively, as previous methods have proven insufficient [10]. Furthermore, optimal placement of aggregation switches within network topologies remains complex, necessitating careful consideration of link rates, load distribution, and resource constraints [45]. The inadequacy of existing benchmarks in capturing the diverse impacts of congestion across applications complicates the identification of effective metrics for predicting performance degradation [64]. Efficient routing and congestion avoidance are vital, especially in bandwidth- and energy-constrained environments like IoT networks [32]. Table 2 provides a comparative analysis of various congestion control methods, emphasizing their packet prioritization, resource management, and adaptability to fluctuating network conditions. These challenges highlight the need for advanced, adaptable congestion control mechanisms capable of dynamically responding to the complexities of modern digital communication networks.

3.2 Machine Learning-Based Approaches

Machine learning has revolutionized congestion control by enabling adaptive mechanisms that respond dynamically to changing network conditions. Time-series analysis of Round Trip Time (RTT) data predicts congestion scenarios, facilitating proactive management strategies that maintain efficient data flow [46]. In video streaming, statistical-learning models like Iris regulate video bit rates, optimizing queue loads and enhancing user experience by reducing buffering [47]. The Distributed Deep Learning for Mobile Traffic Prediction (D2L-MTP) method exemplifies the application of deep learning architectures, such as CNNs and RNNs, to process data locally at base stations, improving congestion control and reducing latency [18]. Machine learning frameworks categorize research on resource management and Quality of Service (QoS) in 5G networks, focusing on advanced technologies like network slicing, enabling tailored solutions to specific congestion challenges [3]. The Dynamic Parallel Processing Framework (DPPF) enhances congestion management by allowing parallel processing of data chunks, improving throughput and reducing delays [17]. The application of non-linear techniques and bifurcation theory to analyze feedback effects on stability represents a significant advancement in congestion control, providing insights into system behavior under varying conditions [65]. Benchmarking efforts have introduced structured evaluations of Pre-Congestion Notification (PCN) algorithms, offering insights into their performance across diverse network conditions [66]. Non-blocking reinforcement learning agents further enhance network performance by allowing continuous data transmission while the agent computes its next action, thus minimizing decision-making delays [56]. Integrating machine learning techniques into congestion control strategies creates a comprehensive and adaptive framework for addressing the intricate challenges of network congestion. These approaches utilize advanced methods such as reinforcement learning, clustering, and classification to enhance real-time decision-making, ultimately improving resource utilization and network performance. By differentiating between types of packet losses—congestive

AI-generated, for reference only.

versus non-congestive—machine learning-based solutions can optimize throughput, particularly for bandwidth-intensive applications like video streaming, significantly enhancing overall network efficiency [33, 67, 68].

3.3 Flow Differentiation Techniques

Flow differentiation techniques are essential for managing network congestion, enabling prioritization and classification of data flows based on specific requirements. These techniques are crucial for optimizing resource allocation and ensuring high network performance, particularly in 5G environments experiencing a surge in mobile traffic and diverse QoS requirements. With projections indicating a potential 1000-fold increase in mobile traffic, efficient management of limited resources is imperative to prevent congestion and performance degradation. These techniques enhance service discovery and mitigate latency issues associated with high volumes of service requests, improving user experience across varying traffic patterns and QoS demands [69, 10, 3]. Advanced scheduling algorithms dynamically adjust data flow priorities based on real-time network conditions and application demands, ensuring critical packets receive higher priority and reducing congestion likelihood [19]. Implementing flow differentiation in multicast service chains has optimized packet processing and routing, enhancing service delivery in complex digital ecosystems [14]. Integrating machine learning models into flow differentiation strategies enables adaptive classification of data flows, leveraging historical data to predict and proactively respond to congestion scenarios, which is crucial for maintaining seamless data transmission and reducing latency [18]. Utilizing machine learning in network slicing further exemplifies the potential of these techniques to tailor network resources to specific service requirements, optimizing performance and reducing congestion [3]. In energy-sensitive environments like IoT networks, flow differentiation techniques are vital for optimizing resource utilization and minimizing energy consumption by prioritizing data flows based on energy requirements, contributing to sustainable network operations [32]. Flow differentiation techniques provide a sophisticated framework for managing network congestion by prioritizing data flows and optimizing resource allocation. Techniques such as per-flow scheduling and Pre-congestion Notification (PCN) enable traffic pattern monitoring and bottleneck identification, facilitating informed decisions regarding flow admission and rerouting. Utilizing technologies like Cisco's NetFlow and implementing active queue management systems such as CHOKeD ensures fairness among traffic types, maintains QoS, and mitigates congestion effects, ultimately enhancing network performance [48, 70, 66, 21, 71]. These techniques are essential for supporting the diverse demands of modern digital applications and maintaining high service quality across complex network environments.

3.4 Active Queue Management (AQM) Techniques

Active Queue Management (AQM) techniques are crucial in mitigating network congestion by proactively managing packet queues and optimizing bandwidth allocation. Among these, CHOKeD stands out as a stateless AQM protocol that dynamically adjusts the number of packets drawn from the queue for comparison based on the current queue size, ensuring fairness in bandwidth allocation among both responsive and unresponsive flows [48]. The TCP Additive Increase Multiplicative Decrease (AIMD) algorithm, while effective, can cause queue length oscillations during congestion, resulting in substantial delay jitter for end-to-end applications. Addressing these oscillations is vital for maintaining application performance consistency [50].

In this context, ADPG, utilizing deterministic policies, ensures stability and convergence in multi-agent, partially observable environments, contributing to more stable network performance [51]. QTCP represents an innovative approach that dynamically adjusts the congestion window (cwnd) using a Q-learning algorithm, maximizing throughput and minimizing latency, reflecting the trend of integrating machine learning techniques into AQM strategies [52]. FASTFLOW combines multiple congestion signals to provide a more accurate and timely response to network conditions, enhancing fairness and overall performance [53].

The implementation of PFC and RCM in OMNeT++ simulations emphasizes the importance of modeling and evaluating AQM techniques to assess their effectiveness in real-world scenarios [55]. Additionally, buffering delay-tolerant data traffic at the Mobile Edge Computing (MEC) layer until network congestion subsides or traffic deadlines approach represents an edge-based strategy to alleviate congestion [49]. Another noteworthy approach is RPM, which provides immediate feedback to senders by marking in-flight ACKs, allowing for quicker adjustments to congestion control without modifying the end-host stack, thus facilitating faster response times [54].

Furthermore, combining adaptive control, backstepping techniques, and H-infinity (H1) control offers a robust strategy for effectively tracking desired queue lengths in TCP/AQM systems [59]. Benchmarking efforts evaluating AQM algorithms such as FRED, BLUE, SFB, and CHOKe against established baselines like RED and Drop Tail provide valuable insights into their effectiveness in managing congestion [72]. These evaluations are crucial for identifying the most efficient AQM strategies to ensure optimal network performance and service quality.

Collectively, these AQM techniques underscore the critical role of proactive queue management in addressing the complexities of modern digital communication networks. Figure 2 illustrates the categorization of Active Queue Management (AQM) techniques, highlighting stateless protocols, learning-based approaches, and simulation/benchmarking efforts to manage network congestion effectively.

3.5 Innovative Congestion Control Protocols

Method Name	Design Approach	Performance Metrics	Technological Integration
MVFST[56]	Asynchronous Design	Throughput And Delay	Not Mentioned
GHS[58]	Graph-based Heuristic	Service Execution Time	-
INCAB[60]	Bloom Filter	Completion Time	P4 Language
MPDCCA[61]	Multi-path Algorithms	Aggregated Throughput	-
IRIS[62]	Standalone Approach	Packet Identification Accuracy	-
Dart[63]	Divide-and-specialize	Latency And Throughput	-
ABH1TC[59]	Adaptive Control	Tracking Error	Machine Learning
MSHAP[57]	Heuristic Solutions	End-to-end Latency	Hardware Acceleration

Table 3: Overview of innovative congestion control protocols detailing their design approaches, performance metrics, and technological integrations. This table highlights the diversity in methodologies and their respective impacts on network performance, showcasing advancements in asynchronous design, heuristic solutions, and adaptive control mechanisms.

Innovative congestion control protocols are vital for addressing the complexities of modern digital communication networks characterized by high data traffic and diverse application requirements. Table 3 provides a comprehensive summary of various innovative congestion control protocols, illustrating their design strategies, performance outcomes, and the extent of technological integration. The asynchronous design of MVFST RL represents a significant advancement, enabling RL agents to operate without waiting for policy updates, thus improving congestion control efficiency by allowing continuous data transmission and reducing delays associated with traditional synchronous methods [56]. Heuristic solutions in vehicular fog computing demonstrate enhanced resource utilization and service resilience compared to traditional methods, particularly in video processing applications, showcasing potential for improved network performance [58]. The In-network Congestion-Aware Load Balancing (INCAB) system achieves a 31.97% reduction in average flow completion time relative to stateless solutions, underscoring the effectiveness of congestion-aware systems in optimizing load distribution [60]. Multi-path dual congestion control algorithms enhance stability and throughput by leveraging multiple paths for efficient traffic distribution, reducing congestion and improving network reliability [61]. The In-router Identification Scheme for Selective Discard of Video Packets (IRIS) introduces a novel technique for managing video traffic through header analysis, leading to reduced congestion and improved service quality [62]. The Dart protocol achieves a 60% reduction in tail latency in small-scale tests and a 79% improvement in simulations, along with a 58% increase in throughput compared to existing methods, effectively addressing RDMA congestion control challenges [63]. The application of the PPC technique to TCP/AQM systems allows for improved tracking of queue lengths and disturbance rejection, enhancing the stability and performance of congestion control mechanisms [59]. Benchmarking efforts, such as CC-Bench1 and CC-Bench2, provide structured methodologies for evaluating Internet congestion control schemes, facilitating the development of more effective protocols [73]. In vehicular edge computing, key research directions include enhancing QoS, scalability, economic profit models, and improving security and privacy mechanisms [2], all critical for developing robust congestion control protocols to meet future digital communication demands. Furthermore, systematic evaluations of hardware-accelerated transport methods highlight their potential in reducing model-serving latency compared to traditional TCP-based approaches, emphasizing the benefits of hardware acceleration in improving congestion control efficiency [57]. The integration of machine learning techniques into congestion control protocols presents promising research avenues, with adaptive mechanisms utilizing predictive analytics and enhanced routing protocols being crucial for managing congestion in increasingly complex IoT networks [32]. Collectively, these innovative protocols and methodologies highlight the importance of adaptive, intelligent solutions in managing network congestion, ensuring efficient and reliable service delivery in complex network environments.

Feature	Challenges in Traditional Congestion Control	Machine Learning-Based Approaches	Flow Differentiation Techniques
Optimization Technique	Not Specified	Adaptive Mechanisms	Scheduling Algorithms
Application Focus	Packet Prioritization	Video Streaming	Resource Allocation
Performance Enhancement	Resource Management	Reduced Latency	Reduced Congestion

Table 4: This table provides a comparative analysis of network congestion control methodologies, highlighting the unique features and applications of traditional, machine learning-based, and flow differentiation techniques. It outlines the optimization techniques, application focus, and performance enhancements associated with each approach, offering insights into their effectiveness in addressing network congestion challenges.

4 Frame Synchronization

Frame synchronization is integral to digital communication, particularly in multimedia applications like video streaming and gaming. It ensures optimal performance and user experience by directly impacting visual quality and network efficiency.

4.1 Significance of Frame Synchronization

In digital communication, frame synchronization is crucial, especially for video streaming and gaming, where consistent frame timing is essential for high-quality user experiences. Aligning video frames with display refresh rates reduces visual artifacts such as screen tearing, enhancing playback and gameplay smoothness. This is vital in fast-paced environments like FPS games, where variable frame timing can significantly affect perceived smoothness [35]. Effective synchronization also optimizes network performance by prioritizing video packets based on their characteristics, reducing congestion and improving service quality [62]. This capability is essential in complex digital ecosystems with diverse traffic patterns and QoS demands.

4.2 Techniques for Achieving Frame Synchronization

Frame synchronization is achieved through various hardware and software techniques that align video frames with display refresh rates. Advanced buffering strategies store video frames temporarily to sync with the monitor's refresh cycle, addressing screen tearing and stuttering in high frame rate variability applications [35]. Adaptive synchronization algorithms adjust frame delivery timing based on real-time network conditions and display capabilities, optimizing video packet flow by prioritizing key frames and reducing delays [62]. Machine learning enhances these algorithms by predicting network congestion and adjusting synchronization parameters to improve user experience [18].

Hardware advancements like Variable Refresh Rate (VRR) technology enable displays to adjust refresh rates dynamically to match content frame rates, minimizing input lag and providing smoother visuals, especially in gaming and video playback [35]. Reconfigurable intelligent surfaces (RIS) can further enhance synchronization by improving signal quality and reducing latency in wireless communications [23]. Integrating advanced buffering strategies, adaptive algorithms, VRR technology, and RIS creates a comprehensive framework for frame synchronization. Edge-based video analytics, content delivery networks (CDNs), and in-network computing further enhance video frame delivery precision, improving the quality and reliability of modern digital applications [33, 36, 69].

As illustrated in Figure 3, which categorizes the key techniques for achieving frame synchronization into software and hardware methods, as well as integrated frameworks, it becomes evident how these approaches work in tandem. The software techniques, including advanced buffering, adaptive algorithms, and machine learning, are essential for optimizing synchronization. On the hardware side, technologies such as variable refresh rate and reconfigurable intelligent surfaces play a crucial role. Furthermore, the integration of edge-based video analytics, content delivery networks, and in-network computing showcases a comprehensive approach to optimizing video frame synchronization. The first technique, "Swarm Leader and Worker Communication in a Distributed System," highlights a network architecture where Edge Nodes are interconnected via a Bridge Network, crucial for maintaining synchronization. The second technique, "Computer Setup Parameters," outlines hardware and software configurations influencing synchronization, including display resolution, G-SYNC (VRR), VSync, and latency metrics, all critical for optimizing performance.

4.3 Technologies Supporting Frame Synchronization

Technologies facilitating frame synchronization are essential for seamless video content delivery across digital platforms. Advanced video codecs like High Efficiency Video Coding (HEVC) optimize compression and decompression, ensuring synchronization between video frames and display refresh rates while reducing latency [62]. VRR technology allows displays to adjust refresh rates dynamically to match incoming video signals, significantly reducing screen tearing and stuttering [35], particularly beneficial in gaming and high-definition playback.

Edge computing resources enhance frame synchronization by processing video data closer to the source, minimizing latency and ensuring timely delivery in applications with stringent requirements, such as virtual reality and live streaming [18]. Additionally, RIS technology improves signal quality and reduces latency in wireless networks, supporting effective frame synchronization [23]. The integration of advanced video codecs, VRR, edge computing, and RIS creates a robust framework for frame synchronization, particularly in latency-sensitive applications like smart cities, satellite networks, and the Internet of Vehicles. This approach leverages edge computing to process large volumes of video data in real-time, reducing bandwidth consumption and latency while enhancing video analytics accuracy through innovative algorithms that optimize resource management and offloading decisions [74, 36, 34, 75].

4.4 Applications in Video Streaming

Frame synchronization is crucial in video streaming, ensuring seamless content delivery and playback, vital for high-quality user experiences. Consistent timing of video frames with display refresh rates prevents visual artifacts like screen tearing and stuttering, particularly in adaptive streaming scenarios where video quality adjusts based on network conditions to optimize Quality of Experience (QoE) [38]. Advanced video codecs like HEVC facilitate synchronization by efficiently compressing and decompressing video data, reducing latency and aligning frames with display refresh cycles for smooth playback [62]. Edge computing supports this by processing video data closer to users, minimizing delays in live streaming applications [18].

As illustrated in Figure 4, the advancements in video streaming applications encompass key technologies such as frame synchronization, VRR technology, and RIS. Frame synchronization ensures seamless content delivery through advanced codecs and edge computing, while VRR technology enhances video playback by dynamically adjusting refresh rates, thereby reducing input lag and improving visuals. Additionally, RIS technology contributes to better signal quality and reduced latency in streaming environments [35].

Incorporating these technologies, VRR technology in streaming platforms allows displays to adjust refresh rates dynamically to match video content frame rates, further enhancing the viewing experience [35]. Moreover, RIS can improve signal quality and reduce latency in wireless networks, contributing to effective frame synchronization in video streaming [23].

4.5 Applications in Gaming

Frame synchronization is critical for enhancing the gaming experience by aligning display refresh rates with game content frame rates, minimizing visual artifacts like tearing and stuttering. VRR technology in gaming monitors enables dynamic refresh rate adjustments to match gaming consoles and PCs, resulting in smoother gameplay and reduced input lag [35]. In fast-paced gaming environments, precise frame delivery timing ensures visual output aligns with player actions, enhancing game responsiveness, especially in competitive gaming where minor delays can impact performance [35].

Advanced techniques, such as adaptive frame pacing algorithms, maintain gameplay smoothness by adjusting frame delivery based on real-time analysis of game engine performance and system capabilities, optimizing frame flow and minimizing delays [35]. Deploying edge computing resources in gaming networks significantly reduces latency by processing game data closer to players, improving synchronization of game frames and enhancing the overall gaming experience, particularly in cloud gaming platforms where low latency and high-quality graphics are essential [18].

5 Latency Compensation

Addressing latency in digital communication systems requires strategic latency compensation methods. This section delves into latency-aware mapping in edge computing, a pivotal factor for optimizing performance in latency-sensitive applications. By examining mechanisms and methodologies, it becomes clear how proximity to end-users and intelligent resource allocation can effectively mitigate latency issues. The subsequent subsection will explore the nuances of latency-aware mapping in edge computing environments, emphasizing its role in enhancing application responsiveness and user experience.

5.1 Latency-Aware Mapping in Edge Computing

Effective latency management in edge computing is essential for optimizing performance in latency-sensitive applications like smart cities and video streaming. Traditional cloud solutions often incur high latency due to the distance between clients and data centers, necessitating decentralized solutions such as EdgeKV [76]. Edge computing mitigates these limitations by situating computational resources closer to end-users, thus enhancing application responsiveness.

Key strategies for managing latency include strategically placing Virtual Network Functions (VNFs) within the edge infrastructure. Utilizing closeness centrality enables optimal VNF placement, improving availability and minimizing service disruption in dynamic network environments [77]. Additionally, distributed caching mechanisms within edge platforms help reduce latency, although resource constraints at edge nodes can impede the efficient delivery of ultra-HD video content [77].

Hardware-accelerated communication methods, such as RDMA and GDR, demonstrate how bypassing the CPU for data transfers can significantly lower latency and increase throughput in model-serving applications [57]. This underscores the potential of hardware acceleration in boosting edge computing efficiency.

Risk-sensitive learning frameworks optimize task fetching and offloading decisions in vehicular networks by considering variability in network conditions and resource availability. These frameworks enable efficient latency management by allowing users to make offloading decisions based on their risk tolerance for latency outages, incorporating analyses of user impatience and queue status [77].

Amplification of tail latency in distributed edge computing environments severely impacts user experience and processing efficiency, necessitating adaptive learning methods to address these challenges. Competition among users for limited computing resources can lead to uneven load distribution and increased response times, highlighting the need for effective latency management strategies [57].

5.2 Collaborative Optimization Techniques

Collaborative optimization techniques are vital for latency compensation, particularly where multiple network components must collaborate to reduce delays and enhance performance. Scalable Video Coding (SVC) enables distributed caching of video layers at Mobile Edge Computing (MEC) platforms, increasing the likelihood of successful offloading and reducing latency by caching content closer to users [78].

Proactive flow admission and rerouting strategies, such as PFARS, exemplify collaborative techniques that mitigate congestion propagation. By continuously monitoring network conditions and making informed flow admissions, PFARS ensures optimal latency levels by utilizing network links only when they can handle the load [71].

In Unmanned Aerial Vehicle (UAV) control, the PACED-5G method integrates edge computing with 5G networking to predict and compensate for potential delays, leveraging 5G's low-latency capabilities to enhance UAV operations' responsiveness and reliability [79].

A comprehensive survey categorizes latency compensation techniques into four groups: feedback, prediction, time manipulation, and world adjustment, providing a structured framework for implementing collaborative optimization strategies [80]. This classification is visually represented in Figure 5, which illustrates the hierarchical classification of collaborative optimization techniques in network latency management, emphasizing latency compensation, resource allocation, and network integration strategies.

The Optimal-Transport-Based Reinforcement Learning (RL) approach optimizes offloading decisions by combining reinforcement learning with optimal transport principles, facilitating efficient resource allocation and minimizing latency through dynamic adaptation to changing network conditions [81].

Pipelining scheduling operations in LTE and DOCSIS networks enable proactive grant allocation, improving latency performance by ensuring timely and efficient resource allocation [77].

Recent research underscores the need for comprehensive integration of diverse network components and methodologies to manage latency effectively, particularly for next-generation applications in the Internet of Things and mission-critical services. By addressing complex cross-layer interactions and employing advanced routing and computation offloading strategies, these techniques enhance overall network performance and significantly improve user experience in modern digital applications, especially in real-time communication and content delivery [33, 82, 83, 80, 84].

5.3 Dynamic Resource Allocation Strategies

Dynamic resource allocation strategies are crucial for managing latency in contemporary digital communication systems, where network conditions and workload demands are constantly evolving. Integrating AI and machine learning into Multi-access Edge Computing (MEC) systems presents promising avenues for enhancing resource management and facilitating MEC deployment across diverse environments [37]. These advancements aim to optimize energy consumption, enhance security measures, and develop predictive resource management mechanisms.

The Dynamic Parallel Processing Framework (DPPF) exemplifies dynamic resource allocation by optimizing processing speed through real-time data analysis, allowing for efficient resource adjustment to meet processing demands [17]. In video services, optimizing caching, transcoding, and resource allocation interactions is crucial for meeting client requirements and adapting to network conditions, thereby minimizing latency [85].

The Parsimonious Edge Computing (PEC) method effectively manages microservice resources in edge computing environments while meeting performance targets, underscoring the importance of efficient resource allocation in maintaining service quality and reducing latency [86]. Game theory applications in the PGRA method address the VNF placement problem, employing decentralized resource allocation algorithms to enhance efficiency in complex network environments [87].

In VR applications, optimizing bandwidth allocation significantly reduces average end-to-end latency, facilitating immersive social interactions [88]. Tail-learning, which integrates Laplace transform techniques with reinforcement learning, introduces an innovative approach to derive a theoretical upper bound for response time, enabling improved decision-making in scheduling [89].

Dynamically adjusting scheduling timescales based on real-time conditions remains critical, as existing methods often lead to high operational costs and suboptimal performance [90]. Developing robust algorithms and lightweight data processing frameworks is essential, particularly in fog and edge computing environments where GPU resource management is key [91].

In-network congestion-aware load balancing systems like INCAB ensure even load distribution by dynamically adjusting service instance weights based on congestion levels [60]. Collectively, these strategies highlight the importance of dynamic resource allocation in managing latency, ensuring efficient and responsive service delivery in complex digital ecosystems.

5.4 Latency Compensation in Real-Time Applications

Latency compensation in real-time applications is critical for delivering seamless user experiences, particularly in scenarios requiring instantaneous data processing and response. The MiGrror method significantly reduces downtime and migration time, making it suitable for latency-sensitive applications [5]. This method exemplifies the importance of efficient migration strategies in minimizing service disruption and ensuring continuity.

Urgent Edge Computing (UEC) emerges as a decentralized alternative to traditional Urgent Computing, emphasizing real-time processing and immediate computations [7]. UEC's decentralized approach facilitates rapid data handling and decision-making, essential for applications requiring quick responses, such as emergency services and real-time analytics.

In connected autonomous vehicles (CAVs), the RF-based compression method demonstrates substantial data savings and high reconstruction quality, enhancing scalability and realism in metaverse applications [11]. This compression technique underscores the potential for improved latency compensation by reducing data load and facilitating faster data transmission and processing.

Advanced scheduling techniques, such as the deficit-based airtime fairness scheduler, significantly improve latency and throughput in WiFi networks, benefiting real-time applications like VoIP and web browsing by ensuring fair resource allocation and reduced contention [92]. The Iris method further exemplifies the role of adaptive strategies in managing network resources, achieving a 25% increase in video bitrate under varying network conditions [47]. These strategies highlight the importance of adaptability in maintaining high-quality service.

For data stream processing, the CODA method achieves lower processing times and reduced streaming traffic, showcasing its effectiveness in resource allocation and latency management [23]. This is crucial in environments with high data processing demands, where efficient resource utilization directly impacts performance.

In vehicular networks, dynamically scaling and placing service instances based on mobility patterns enhances latency management, emphasizing the need for adaptive infrastructure placement to optimize service delivery [58]. The integration of optimal transport and reinforcement learning in the OTRL method exemplifies sophisticated task allocation and computation offloading strategies, enabling adaptive optimization of latency compensation [81].

The MPTCP-enabled Mobile Device Cloud effectively manages congestion and enhances service quality, demonstrating its potential for real-time applications in mobile environments [1]. By leveraging multipath communication, this approach reduces latency and improves user experiences, emphasizing the importance of strategic resource allocation and advanced computing paradigms in supporting real-time applications. Collectively, these techniques underscore the significance of dynamic adaptation and resource efficiency in ensuring responsive and reliable service delivery in latency-sensitive scenarios.

As shown in Figure 6, latency compensation is critical in real-time applications, significantly impacting performance and user experience. The figures illustrate the importance of optimizing flow completion times and managing latency distributions across various algorithms and protocols. The first figure presents a comparative analysis of flow completion times using different algorithms such as FASTFLOW, MPRDMA, BBR, Swift, and EQDS + FASTFLOW, highlighting FASTFLOW's efficiency in handling data flows. The second figure examines latency distribution across four distinct protocols: Fast FIFO, Fast FQ-CoDel, Slow FIFO, and Slow FQ-CoDel, underscoring the variability in latency performance and the need for effective management strategies to ensure seamless real-time application performance. Together, these examples underscore the pivotal role of latency compensation in enhancing responsiveness and reliability in real-time systems [53, 92].

5.5 Challenges and Future Directions

Latency compensation in digital communication systems faces challenges, particularly in resource management and the integration of advanced technologies. A significant challenge is developing adaptive network architectures that can dynamically adjust to varying latency demands. Emerging technologies like fog computing and enhanced edge caching show promise for supporting real-time applications, yet their integration into existing systems remains complex. The complexity of Virtual Network Function (VNF) placement and the need for robust failure detection mechanisms further complicate latency management. Future research should prioritize relaxing reliability assumptions in Network Functions Virtualization (NFV) and Mobile Edge Computing (MEC) contexts while investigating innovative VNF migration strategies. This focus is essential for enhancing system resilience and optimizing resource utilization, especially for low-latency, high-availability applications such as video streaming and IoT services. By addressing challenges related to latency, availability, and resource allocation, future studies can bridge the gap between theoretical frameworks and practical implementations, ultimately improving edge computing performance and reliability [93, 69, 94].

In real-time applications, implementing adaptive mechanisms for managing computation and transmission times is crucial, as these systems must effectively respond to network conditions, prioritize data packets, and maintain high Quality of Service (QoS). For instance, in-network computing can significantly reduce packet loss during video streaming in congested networks, while flexible congestion control algorithms like FASTFLOW can rapidly adjust to varying traffic conditions in data centers, ensuring fairness and performance. Proactive management strategies in Edge Computing environments can help monitor and optimize task execution to uphold QoS requirements, demonstrating the critical role of adaptive mechanisms in enhancing overall system performance in dynamic conditions [33, 53, 95]. Extending current models to more complex queueing structures can improve the efficiency of latency compensation techniques. Integrating machine learning techniques offers promising avenues for enhancing the adaptability of controllers in diverse network conditions, essential for maintaining optimal performance. Future research should also explore hardware experiments to validate analytical insights and investigate the impact of queue feedback in more complex network environments.

The application of time-series analysis and forecasting in network environments highlights the need for improved data interpolation methods and enhanced identification of network links. Utilizing parallel processing to speed up computations could further optimize latency management strategies. Investigating further optimizations in Dart's congestion control mechanisms, particularly its performance in large-scale, complex data center environments, presents significant opportunities for future research, especially given Dart's innovative divide-and-specialize approach that effectively addresses receiver congestion and enhances response times in RDMA-based networks [34, 69, 63, 96, 45]. Integrating emerging technologies like AI and machine learning into 5G networks is another promising area for enhancing latency reduction strategies and overall network performance.

In video streaming, improving the model's adaptability to evolving streaming technologies and enhancing the accuracy of Quality of Experience (QoE) predictions for various network conditions is critical. Future research could focus on optimizing communication overhead and exploring predictive congestion control techniques, such as PredicTor, in larger, more complex network topologies. To effectively address the challenges posed by increasing demand for network resources, particularly in video streaming and IoT applications, significant enhancements are required in both network characterization tools and methodologies for assessing computational power. These improvements are essential to bridge the gap between theoretical frameworks and practical implementations, ensuring optimized performance and reduced latency in edge data processing [97, 36, 69].

Evaluating latency compensation techniques in diverse gaming environments and exploring emerging trends in game design and network technologies will be essential for advancing the field. Enhanced optimization of statistical-learning-based models, coupled with their deployment across diverse network environments, has the potential to significantly improve congestion management strategies. By leveraging machine learning techniques such as reinforcement learning, these models can adapt to the dynamic characteristics of modern network scenarios—including data centers, 5G, and hybrid networks—thus offering more effective responses to varying congestion levels. This adaptability addresses traditional congestion control challenges and differentiates between congestive and non-congestive packet losses, ultimately leading to improved network performance and resource utilization [47, 69, 67, 68]. Future research could explore enhancements in algorithms to better handle extreme variability in task demands and investigate integrating additional machine learning techniques for further optimization.

Future research will also focus on improving synchronization between computation and ranging operations, exploring full-scale deployments in real-world settings, and enhancing system scalability and reliability. Further optimizations in Bandwidth Report (BWR) implementation and its application in different network environments are areas for exploration. Enhancing QoS for varied applications and developing secure and efficient data management protocols are crucial for future vehicular network developments. To enhance the effectiveness of edge computing architectures, integrating more powerful computing nodes while addressing critical constraints such as energy efficiency and privacy is essential. This involves leveraging advanced technologies and methodologies to optimize processing capabilities at the network edge, ensuring applications can operate with reduced latency and improved responsiveness, particularly for real-time services reliant on data from numerous IoT devices. Furthermore, implementing robust privacy policies at the edge can safeguard user data before transmission to the cloud, enhancing trust and compliance in edge computing environments [27, 28, 97, 96, 30].

Collectively, these challenges and future research directions highlight the need for innovative solutions to optimize latency compensation, ensuring efficient and reliable service delivery in increasingly complex digital ecosystems. Future research could explore additional metrics and congestion scenarios to refine the understanding of network impacts on application performance. Furthermore, enhancements to algorithms for better adaptability in highly dynamic network conditions and the integration of machine learning techniques for improved decision-making are promising avenues for future research. Additionally, future research should focus on improving model robustness through domain randomization and exploring advanced reward normalization strategies to enhance performance across varied environments [56].

6 Variable Refresh Rate (VRR)

The demand for high-quality visual experiences in digital applications underscores the significance of Variable Refresh Rate (VRR) technology. This section outlines the fundamental aspects of VRR, setting the stage for a detailed analysis of its impact across various contexts. The following subsection will focus on Benchmarking Variable Frame Timing (VFT), which is crucial for assessing VRR performance. This assessment highlights the effects of frame timing fluctuations on user experience, particularly in gaming, emphasizing VRR's role in delivering smooth and responsive visuals.

6.1 Benchmarking Variable Frame Timing (VFT)

Benchmark	Size	Domain	Task Format	Metric
VFT-Bench[35]	640	Gaming	Targeting Performance	Time to Completion, Smoothness Rating
CE-LD[83]	9,000	Cloud Computing	Latency Measurement	Latency Reduction, Digital Divide Index
HPC-CB[98]	7,700,000	Network Congestion	Congestion Measurement	PTS
FaaS-Perf[99]	4,890	Cloud Computing	Performance Evaluation	Execution Time, Wait Time
ECS-Bench[100]	1,000	Edge Computing	Performance Evaluation	Latency, Resource Utilization
ISB[101]	50,000	Image Segmentation	Segmentation	IoU, F1-score
UAV-MEC[102]	12,000	Drone Computing	Service Delay Evaluation	Latency, Service Delay
ECLT[103]	1,000	Energy Consumption	Request-Response Communication	Energy Consumption, Latency

Table 5: Overview of benchmark datasets used in various domains, detailing their respective sizes, domains, task formats, and evaluation metrics. This table serves as a comprehensive reference for understanding the scope and focus of different benchmarks in fields such as gaming, cloud computing, and image segmentation.

Benchmarking Variable Frame Timing (VFT) is essential for evaluating VRR's impact on user experience, especially in gaming. VFT benchmarking identifies how frame timing variations influence the perceived smoothness and performance of games, particularly first-person shooters (FPS), where rapid visual updates are vital [35]. By analyzing VFT, developers can determine optimal VRR configurations that enhance motion fluidity and minimize visual artifacts like tearing and stuttering.

Table 5 provides a detailed overview of benchmark datasets utilized across various domains, highlighting their characteristics and evaluation metrics pertinent to the study of Variable Frame Timing (VFT) and its broader applications. Figure 7 illustrates the hierarchical structure of benchmarking VFT in gaming, focusing on its impact on user experience, VRR optimization, and the development of adaptive techniques for enhanced gaming performance. This visual representation complements the discussion by providing a framework that highlights the interconnectedness of these elements. Understanding this structure is crucial for grasping how frame timing variability and VRR adjustments influence gaming experiences across diverse hardware and network conditions. Insights from VFT benchmarking inform the development of adaptive synchronization techniques that dynamically adjust frame delivery based on real-time analysis, ensuring critical visual information is rendered with minimal delay and maximum fidelity.

6.2 Machine Learning in VRR Optimization

Machine learning plays a significant role in optimizing VRR technology through adaptive synchronization techniques that improve user experiences across various applications. By integrating machine learning algorithms, real-time analysis of frame timing variability is facilitated, allowing for dynamic adjustments of display refresh rates in accordance with content frame rates. This capability is especially beneficial in augmented reality (AR), where user task collaboration optimizes resource allocation and visual performance [104].

The systematic approach of VFT benchmarking provides insights into perceptual smoothness and performance metrics related to VRR technology [35]. Evaluating these factors enables the training of machine learning models to predict optimal VRR settings that reduce visual artifacts, ensuring a seamless viewing experience. The ability to differentiate between constant and variable frame rates further enhances model precision, allowing for tailored optimization strategies suited to specific user scenarios.

Machine learning in VRR optimization demonstrates the potential of intelligent systems to dynamically adjust display refresh rates, enhancing motion fluidity and overall quality of experience (QoE) across platforms. Techniques such as deep reinforcement learning (DRL) have been employed to optimize video bitrate and computation distribution in VR environments, yielding significant improvements in metrics like peak signal-to-noise ratio (PSNR) and reduced rebuffering times. Additionally, continual reinforcement learning frameworks facilitate real-time resource allocation in mobile edge computing, ensuring VR content streaming meets user demands while minimizing latency and maximizing QoE [105, 89, 35, 39].

6.3 Dynamic Environments and Resource Optimization

VRR technology is crucial for adapting to dynamic environments and optimizing resource allocation, especially in applications where frame timing variability can significantly impact user experience. In gaming and multimedia, VRR adjusts display refresh rates to align with content frame rates, reducing screen tearing and stuttering while enhancing visual performance [35]. This adaptability is vital in environments where frame rates fluctuate due to varying computational loads and network conditions.

Datasets capturing diverse frame time sequences across refresh rates provide insights into user responses under different VFT conditions, highlighting VRR's role in maintaining fluid motion and consistent visual quality [35]. By utilizing this data, VRR systems can optimize resource utilization, efficiently allocating computational resources to maintain desired visual fidelity, thus enhancing energy efficiency and performance.

Moreover, VRR technology minimizes the computational overhead associated with rendering frames at fixed refresh rates, enabling more effective processing power allocation, reducing energy consumption, and extending battery life in portable devices. Insights from VFT benchmarking further contribute to developing adaptive synchronization techniques that dynamically adjust frame delivery based on real-time analysis, ensuring critical visual information is rendered promptly and with high fidelity [35].

The capability of VRR technology to dynamically adjust to varying frame timings and optimize resource allocation is vital for enhancing user experiences in modern digital applications, especially in immersive environments like virtual reality (VR) and augmented reality (AR). Utilizing advanced techniques such as deep reinforcement learning and edge computing, VRR not only reduces latency and rebuffering times but also significantly improves overall QoE, establishing it as a crucial component for delivering high-performance, interactive applications in real-time [35, 106, 39, 105, 104]. By aligning display refresh rates with content frame rates, VRR enhances both performance and efficiency, making it indispensable in optimizing digital communication and data processing systems.

7 Edge Computing

7.1 Efficient Data Representation and Video Quality

Edge computing significantly enhances data representation and video quality by situating computational resources near data sources, thereby reducing latency and improving user Quality of Experience

(QoE). This proximity is crucial for computation-intensive applications like real-time video analytics and traffic monitoring, where prompt data processing is essential [107]. By leveraging local processing capabilities, applications can efficiently manage tasks, maintaining high-quality video services even under variable network conditions.

Edge infrastructures support effective resource management and real-time video processing, ensuring sustained high-quality video output. For instance, the Parsimonious Edge Computing (PEC) framework offers a lightweight resource management approach, minimizing energy consumption while maintaining service quality [86]. Additionally, the Optimal-Transport-Based Reinforcement Learning (RL) method adapts to dynamic task requirements, optimizing resource use across cloud and edge servers to enhance video quality [81].

Resource scheduling in edge computing is complex due to resource constraints and the need for dynamic task allocation. The Dynamic Distributed Scheduler optimizes resource use in dynamic environments, ensuring efficient task scheduling for enhanced video processing [108]. Moreover, matching-based service offloading schemes improve performance and video quality by effectively assigning services to edge servers based on invocation frequency and reusability [109].

Fully distributed algorithms based on Lyapunov control theory manage task offloading and resource allocation in multi-hop MEC networks, optimizing video data processing and delivery [110]. These strategies collectively enable edge computing to manage video data effectively, providing high-quality services across diverse network environments.

7.2 Resource Management and Scheduling

Effective resource management and scheduling are pivotal in edge computing for optimizing resource allocation and ensuring efficient task execution across distributed networks. The Edge Federation Integrated Platform (EIP) addresses challenges from the independent operation of edge infrastructure providers, which limits global resource allocation optimization [111]. Frameworks like Edge Federation promote cooperation and information sharing, enhancing resource management.

Adaptive scheduling strategies, such as those in the EdgeTimer framework, improve profit and performance by aligning scheduling timescales with dynamic demand fluctuations [90]. Similarly, the Priority-based Fair Scheduling (EFS) framework ensures equitable resource distribution by considering client access and job priorities [112].

In mobile edge computing, integrating computation and storage capabilities within the Radio Access Network (RAN) enhances context awareness and reduces latency [30]. The Distributed Application-Aware Task Scheduling (DAATS) approach addresses the lack of centralized load balancing among cloudlets, ensuring diverse QoE requirements are met [113].

The Deduplicator optimizes resource use by identifying and deduplicating similar computational tasks offloaded by user devices [114]. Furthermore, the RETO framework enhances reliability and minimizes bandwidth consumption by offloading tasks from IoT applications to edge clouds [93].

Moro et al. propose a techno-economic market model for resource allocation, balancing cost-effectiveness with performance needs [115]. The shared Mobile Data Center (MDC) infrastructure integrates dedicated resources with real-time resource allocation to meet varying demands [116].

The COPM mechanism enhances resource allocation efficiency by determining optimal decision pairs of access points (APs) and edge servers [117]. ENTS manages task scheduling by considering job profiles, resource statuses, and task interdependencies [118]. However, Opportunistic Edge Computing (OEC) faces limitations due to potential resource disconnections and variability in availability [119].

Architectures that enable computation deduplication and reuse leverage caching at multiple layers to enhance resource management [96]. The PEC method uses a PID controller to monitor queue lengths, adjusting microservices to prevent resource wastage [86]. The Dynamic Distributed Scheduler addresses scheduling inadequacies in dynamic environments [108]. WHISTLE efficiently offloads frequently invoked services from the cloud to edge servers, optimizing service delivery [109].

Effective resource management and scheduling necessitate adaptive frameworks, cooperative models, and innovative technologies to address challenges related to dynamic demand, decentralized operations, and diverse application requirements. These strategies collectively enhance the efficiency and reliability of edge computing services, leveraging paradigms such as Opportunistic Edge Computing and Urgent Edge Computing to meet the demand for low-latency access and time-sensitive applications. By optimizing resource placement, enhancing scalability, and ensuring robust security, these approaches meet the dynamic requirements of modern digital ecosystems, including critical applications in disaster response, smart city management, and IoT sensor data processing [7, 27, 28, 119].

7.3 Latency Reduction and QoS Enhancement

Reducing latency and enhancing Quality of Service (QoS) in edge computing heavily relies on intelligent resource management and strategic deployment of computational resources. A significant challenge is the redundant computation of services invoked multiple times with similar inputs, leading to inefficient resource use and hindering the ability to meet stringent latency requirements [109]. Innovative approaches are required to minimize redundancy and optimize resource utilization.

Integrating computation deduplication and reuse mechanisms at the network edge presents a promising solution. By storing and reusing previously executed tasks for similar new tasks, edge computing platforms can significantly reduce computational redundancy, enhancing efficiency. This approach minimizes unnecessary processing and improves latency performance and QoS by optimizing the allocation and utilization of limited edge resources, effectively addressing tail latency challenges in dynamic applications like multimedia streaming and online gaming. Adaptive learning techniques and precise resource management strategies mitigate head-of-line blocking, ensuring efficient execution of distributed tasks across edge servers, ultimately enhancing user experience and resource efficiency [120, 83, 86, 36, 89].

Implementing decentralized architectures and efficient data processing techniques is crucial for maintaining low latency and high QoS in dynamic network environments. These advanced architectures facilitate real-time adaptations to changing network conditions and user demands, employing techniques like Urgent Edge Computing (UEC) for ultra-low latency and flexible congestion control algorithms such as FASTFLOW, which optimize bandwidth estimation and congestion management. This ensures that computational tasks, especially in time-sensitive applications like disaster response and video streaming, are executed with minimal delay while maintaining high QoS through proactive monitoring and intelligent task scheduling across distributed nodes [69, 33, 95, 53, 7]. The strategic coordination of resources across multiple computing layers, as demonstrated by the EdgeFlow framework, optimizes task execution and reduces latency, enhancing the overall user experience in edge computing environments.

Establishing robust fault tolerance mechanisms and designing effective pricing and incentive models are essential for increasing participant engagement and optimizing resource management in edge computing environments. As decentralized resources become critical for addressing latency-sensitive applications, these mechanisms enhance reliability by balancing bandwidth consumption with application performance and facilitating strategic resource allocation based on user contributions, promoting a scalable and efficient edge computing infrastructure [27, 93, 119, 91]. By tackling high computational complexity and data processing inefficiencies, these strategies ensure that edge computing platforms can meet the diverse requirements of modern digital applications, particularly those with stringent latency constraints.

8 Cloud Infrastructure

The digital era underscores the pivotal role of cloud infrastructure, extending beyond resource provision to drive technological advancements and applications. This section explores cloud infrastructure's multifaceted dimensions, emphasizing its significance in modern digital communication systems. By examining the interplay between cloud solutions and emerging technologies, we uncover how these infrastructures enhance operational efficiency and address key challenges in data processing and service delivery.

8.1 Role and Importance of Cloud Infrastructure

Cloud infrastructure is essential to digital communication, providing scalable and flexible resources for diverse applications, from data-intensive processes to real-time services. By integrating with edge computing, cloud solutions mitigate traditional latency issues by placing computational resources nearer to data sources, enhancing responsiveness and efficiency [4]. This synergy benefits low-latency applications, such as IoT and UAVs [26].

Advancements in resource management strategies further optimize network performance and reduce communication latencies, improving resource utilization in cloud environments [17]. For instance, the EdgeFlow framework optimizes task offloading and resource allocation in edge computing, enhancing service delivery [121]. The CODA method allocates microservices to Cloud-Fog resources based on mutual preferences, balancing processing time and bandwidth, crucial for efficient service delivery in dynamic ecosystems [23].

Cloud infrastructure also supports scalable QoE measurement techniques, such as ReCLive, enabling real-time classification without client-side data [16]. These advancements highlight cloud infrastructure's vital role in ensuring high-quality service delivery across various digital communication platforms.

8.2 Integration with Edge Computing

Integrating cloud infrastructure with edge computing enhances digital communication systems by merging the cloud's extensive resources with edge computing's proximity advantages. This combination is crucial for low-latency, high-bandwidth applications, like the Metaverse and real-time analytics, where decentralizing computational tasks significantly improves performance [122].

Edge computing extends cloud infrastructure by providing localized processing power, reducing the load on centralized data centers. Modular architectures utilizing containerization enable seamless application deployment and management across cloud and edge environments [100]. Such architectures facilitate flexible scaling and efficient resource utilization, addressing traditional scaling limitations in platforms like Kubernetes [86].

Strategic edge server placement is critical for optimizing resource distribution and minimizing latency. Real-world dataset evaluations from public Wi-Fi networks underscore the importance of considering both dense and sparse access point deployments [123], ensuring optimal edge resource positioning to meet diverse network demands.

SDN plays a pivotal role in managing network resources during live migrations, enabling dynamic resource allocation between cloud and edge nodes [124]. MEC architectures further enhance resource management and service delivery by offering various integration strategies with existing cloud infrastructures [37].

OEC introduces an innovative model aggregating resources from end-users, increasing edge computing solutions' flexibility and scalability [119]. This approach underscores edge computing's potential to leverage distributed resources, optimizing processing capabilities and reducing reliance on centralized cloud resources.

8.3 Challenges and Solutions in Cloud-Edge Integration

Integrating cloud and edge computing presents challenges that must be addressed to fully leverage this hybrid architecture. A significant hurdle is the substantial initial investment required for establishing MDCs. Resource sharing across enterprises raises data privacy concerns, necessitating robust security measures to protect sensitive information [116].

Load imbalances among edge servers can lead to resource utilization inefficiencies, compounded by the need for effective mechanisms to dynamically distribute hash value spaces, ensuring computational tasks are evenly spread across resources [96]. The absence of standardized development platforms for fog computing hampers rapid fog-based application development and testing, creating barriers to innovation and deployment [125].

Integrating multiple EIPs within an Edge Federation model poses additional challenges. Ensuring data privacy and managing the integration process requires sophisticated stakeholder coordination [111]. Existing methods may not account for all variations in service requirements and network conditions, potentially impacting performance in dynamic environments [115].

Several solutions can address these challenges. Advanced security protocols and encryption techniques can alleviate data privacy concerns during resource sharing. Developing standardized platforms and frameworks for fog and edge computing can facilitate rapid application development and deployment, reducing integration time and costs. Dynamic load balancing algorithms, exemplified by the Deduplicator, enhance computational task distribution across edge servers, optimizing resource utilization through computation reuse while improving overall system performance. The Deduplicator identifies and deduplicates similar tasks, managing server resource usage and achieving up to 20% higher computation reuse compared to traditional methods. Priority-based fair scheduling techniques ensure equitable resource allocation among clients while prioritizing job importance, crucial in resource-constrained edge environments [114, 112].

Fostering collaboration among EIPs and cloud service providers can streamline the integration process, ensuring data privacy and security while enabling seamless interoperability. Strategic investments in technology and collaborative efforts are essential for overcoming cloud-edge integration challenges, optimizing resource allocation, and enhancing performance by leveraging edge computing's proximity to reduce latency, improve bandwidth, and enforce privacy policies. These efforts consider infrastructure scalability and sustainability through innovative frameworks like OEC [27, 28, 119].

8.4 Impact on Data Processing and Storage

Integrating cloud infrastructure with edge computing significantly enhances data processing and storage, improving efficiency, scalability, and responsiveness. Cloud infrastructure provides essential resources for managing large-scale data processing and storage tasks, facilitating the seamless handling of vast information generated by modern applications. The MEC orchestrator's orchestration capabilities are vital for managing resources and making informed caching decisions, enhancing overall data processing performance [78].

The FDK streamlines fog-based application development by efficiently managing resource allocation and facilitating transitions between physical and emulated environments, critical for optimizing data processing in fog computing settings [125]. Understanding transient effects and resource contention in FaaS systems is essential for performance optimization. Insights from empirical investigations into these factors can inform future system design, ensuring data processing meets diverse applications' demands [99].

When integrated with cloud infrastructure, edge computing significantly reduces energy consumption for LTE terminal nodes, particularly in connection-oriented scenarios where latency affects throughput. This reduction highlights edge computing's advantages over traditional cloud solutions, enhancing data processing efficiency [103]. The DCRL framework further exemplifies edge computing's impact on real-time data processing, improving processing rates and detection accuracy in video analysis systems by effectively addressing challenges posed by fluctuating network conditions [75].

9 Adaptive Streaming

The evolution of adaptive streaming technologies has revolutionized video content delivery over the internet, addressing challenges posed by fluctuating network conditions and diverse user demands. As the need for seamless, high-quality video experiences grows, understanding the mechanisms that facilitate this adaptability is crucial. The subsequent subsection, "Introduction to Adaptive Streaming," delves into the core principles and operational frameworks of adaptive streaming, highlighting its role in optimizing Quality of Experience (QoE) for users across varied environments.

9.1 Introduction to Adaptive Streaming

Adaptive streaming dynamically adjusts video quality in real-time based on current network conditions, ensuring minimal interruptions and high QoE despite bandwidth fluctuations. It balances factors such as bitrate, segment switches, and stalls, which are critical for user satisfaction [126]. Unlike traditional models that predefine video quality, Adaptive BitRate (ABR) streaming allows real-time quality adjustments, scaling with network conditions, particularly beneficial in unstable mobile environments [**?**]. However, ABR streaming in multi-user contexts can suffer from coordination challenges, leading to inefficient resource allocation and reduced QoE [127].

A key advantage of adaptive streaming is reducing data traffic to the origin server through edge caching, which enhances QoE for mobile clients by decreasing latency and buffering times, while also easing central server loads [128]. The integration of edge computing supports adaptive streaming by providing localized processing power for quicker quality adjustments based on real-time assessments.

Adaptive streaming represents a significant advancement in video delivery technologies, offering robust solutions to variable network conditions. By dynamically adjusting video quality and optimizing resource allocation through in-network computing and joint optimization frameworks, adaptive streaming significantly enhances QoE. It prioritizes essential video packets to reduce transmission loss and employs efficient cache management and transcoding strategies at multi-access edge computing (MEC) servers, making it a vital component in modern digital communication systems [33, 85].

9.2 Bitrate Selection and Caching Strategies

Bitrate selection and caching strategies are crucial in adaptive streaming, optimizing video delivery under varying network conditions while maintaining high QoE. Mehrabi et al.'s joint optimization model exemplifies a sophisticated approach that balances QoE with data traffic management, enabling dynamic bitrate adjustments in response to real-time network assessments [128].

Caching strategies enhance adaptive streaming efficiency. The integration of edge computing in caching mechanisms, as demonstrated by ECAS-ML, offers a robust framework for managing video streaming parameters via machine learning techniques [126]. ECAS-ML leverages localized processing power to dynamically adjust video quality based on network conditions, reducing latency and buffering.

The Multi-User Cooperative Video Streaming Framework (MUCVSF) by Gao et al. illustrates the potential of cooperative strategies in adaptive streaming. By enabling users to download video data for others, MUCVSF optimizes playback and enhances network efficiency, particularly in multi-user environments where coordination is crucial for high QoE [**?**].

Advanced classification techniques, such as the LSTM classifier used by Madanapalli et al., facilitate real-time analysis of request packet patterns, allowing effective classification and QoE assessment [16].

Integrating advanced bitrate selection and sophisticated caching strategies within adaptive streaming systems is essential for optimizing video delivery and enhancing user experience. In mobile edge computing environments, strategic caching at the network edge can significantly reduce backhaul traffic and operational costs for mobile network operators. By minimizing cache misses and balancing video quality with traffic efficiency, these systems achieve refined video segment management, ultimately leading to improved QoE and resource utilization [38, 128, 13, 16].

9.3 Quality of Experience (QoE) Optimization

QoE optimization in adaptive streaming involves strategic bitrate allocation and cooperative frameworks to enhance user satisfaction. The CSBAA model exemplifies an approach that balances QoE with traffic reduction, optimizing bitrate allocation to ensure smooth playback and minimal buffering for video streaming clients [128]. By dynamically adjusting bitrates based on network conditions, CSBAA effectively enhances viewing experiences while managing network resources.

The ECAS-ML framework further emphasizes real-time data utilization in QoE optimization. By integrating machine learning with edge computing, ECAS-ML makes informed decisions that improve streaming quality and minimize stalls, enhancing overall user satisfaction [126]. This approach illustrates the potential of edge computing in providing localized processing power for quick adaptations to changing network conditions.

Additionally, Gao et al.'s cooperative approach leverages idle resources from users not actively streaming, improving QoE through enhanced resource utilization [**?**]. This strategy ensures efficient distribution of network resources, allowing all users to experience high-quality streaming even in multi-user environments.

The integration of innovative frameworks and real-time data analysis techniques, such as machine learning models for live video detection and QoE measurement, showcases their potential in enhancing QoE for adaptive streaming services. These advancements enable precise monitoring and optimization of streaming performance, particularly in congested environments, thereby facilitating improved user satisfaction across various multimedia applications, including live sports and social media broadcasts [38, 16]. By balancing bitrate allocation, leveraging edge computing, and utilizing cooperative strategies, these approaches ensure that adaptive streaming systems can meet the dynamic demands of modern digital communication environments, providing seamless and high-quality viewing experiences.

9.4 Edge and Fog Computing in Adaptive Streaming

Edge and fog computing are pivotal in enhancing adaptive streaming by bringing computational resources closer to end-users, thereby reducing latency and improving QoE. The integration of these paradigms allows for efficient management of video streaming parameters, enabling real-time quality adjustments based on current network conditions. The proximity of edge computing resources facilitates faster data processing and alleviates central server burdens, essential for uninterrupted video playback in fluctuating network environments. This approach minimizes latency and enhances video streaming resilience against congestion and packet loss through techniques like in-network computing and edge-cloud fusion algorithms [33, 36, 69, 34].

The effectiveness of edge computing in adaptive streaming is underscored by the principle that even slight reductions in video quality can significantly decrease backhaul traffic, improving overall system efficiency [128]. Real-time video quality optimization helps manage network resources effectively, ensuring minimal buffering and high-quality playback.

In multi-MEC environments, refined cache update algorithms significantly enhance client QoE by optimizing caching strategies [13]. These algorithms ensure frequently accessed content is stored closer to users, reducing retrieval times and improving the overall streaming experience. Strategic placement of caching resources at the edge minimizes latency and supports the dynamic nature of adaptive streaming, allowing seamless transitions between different video quality levels as network conditions fluctuate.

Fog computing enhances edge computing by acting as an intermediary layer, offering greater distributed processing capabilities than edge devices while remaining less resource-intensive than cloud data centers. This positioning reduces latency and network congestion, improving scalability and flexibility in adaptive streaming systems, which benefit from a balanced approach to resource management and performance optimization [29, 129, 91]. By leveraging both edge and fog resources, streaming platforms can dynamically allocate computational tasks and optimize video delivery, ensuring a consistent and high-quality viewing experience across diverse network conditions.

9.5 Emerging Trends and Hybrid Solutions

Emerging trends in adaptive streaming increasingly focus on integrating advanced machine learning techniques and cooperative strategies to enhance video delivery system efficiency and quality. A significant area of exploration is enhancing parameter prediction through sophisticated machine learning algorithms, which provide accurate assessments of network conditions and user demands, optimizing video quality in real-time [126]. These advancements are crucial for maintaining high QoE in dynamic network environments.

In addition to machine learning, cooperative strategies are gaining traction for optimizing resource utilization in adaptive streaming. Developing incentive mechanisms for user participation is a promising avenue for future research, encouraging users to share resources and improve overall efficiency [**?**]. By effectively managing user cooperation, streaming platforms can leverage idle resources to enhance playback quality for all users.

Hybrid solutions combining edge computing with satellite edge computing are also being explored to address unique challenges in resource prediction and management in distributed networks. Integrating machine learning techniques in this context can provide better resource prediction capabilities, enabling more efficient allocation of computational tasks and improving service delivery in satellite edge environments [22].

The evolving landscape of adaptive streaming highlights the significant potential of integrating advanced technologies—such as in-network computing and multi-access edge computing—with collaborative frameworks to enhance video delivery systems. This integration aims to address challenges like network congestion and packet loss, optimize resource allocation, and improve overall QoE for users, ultimately leading to more efficient and resilient video streaming solutions [38, 69, 33, 126, 85]. By leveraging machine learning, incentive mechanisms, and hybrid computing paradigms, adaptive streaming can effectively meet the evolving demands of modern digital communication, ensuring seamless and high-quality viewing experiences across diverse network conditions.

10 Conclusion

10.1 Future Research Directions

Advancing digital communication and data processing requires a focus on adaptive resource allocation strategies, leveraging machine learning for real-time performance optimization. Priorities include developing adaptive consistency and replication techniques within edge computing to bolster scalability and minimize latency. Flexible scheduling schemes for Quality of Service in diverse environments are crucial for addressing scalability while safeguarding data confidentiality. Integrating machine learning into resource management across cloud and edge computing offers substantial potential for enhancing adaptability and resource utilization. Investigating adaptive pricing mechanisms that reflect demand elasticity and user behavior could lead to more efficient resource allocation and service delivery.

In network congestion control, expanding benchmarks to cover a wider range of scenarios and evaluating additional algorithms will enrich understanding and performance in congestion management. Optimizing resource allocation strategies and deploying advanced algorithms across various domains can enhance their effectiveness. Dynamic bandwidth allocation and adapting algorithms for client mobility are essential for improving service capacity and task management in mobile networks. Developing adaptive scheduling algorithms using machine learning can further enhance resource management, ensuring efficient service delivery and improved user experiences.

Enhancing in-router identification schemes for varying network conditions and broadening their application beyond video can improve their effectiveness in diverse contexts. Incorporating advanced traffic monitoring and machine learning into edge computing frameworks will strengthen smart traffic management systems, providing more accurate and timely data for decision-making. Exploring distributed virtual network function placement algorithms to reduce bandwidth costs and user delays will contribute to more efficient network management.

Emerging trends indicate a shift towards hybrid solutions that combine edge and cloud computing capabilities, enhancing real-time processing while reducing network traffic. Future research should focus on optimizing memory management strategies and improving interoperability between hardware-accelerated transport protocols. Enhancing capabilities in scenarios where data compression is infeasible and exploring alternative task-offloading strategies remain critical areas for further investigation.

In multimedia services, refining edge computing strategies, integrating machine learning for real-time adaptation, and addressing privacy concerns are pivotal for future advancements. Further optimization of resource allocation algorithms and exploring their applicability across various machine learning domains should be pursued. Moreover, integrating machine learning techniques to enhance active queue management algorithms and their adaptability to changing network conditions is a promising research avenue.

Developing efficient resource discovery and allocation algorithms is a key challenge in opportunistic edge computing. Creating heuristics to apply existing methods to larger infrastructures and optimizing additional Quality of Service features beyond response time are essential for advancing software-defined networking-edge computing frameworks. Research on 5G networks highlights crucial technologies for achieving ambitious latency goals, alongside recommendations for future research directions.

Future inquiries should address kernel implementations, scheduling issues in device clouds, and conduct comprehensive analyses of network parameters affecting performance. Developing generalized versions of existing algorithms adaptable to various network topologies and conditions, alongside exploring additional performance-enhancing parameters, should be prioritized. Optimizing migration methods, including integrating machine learning techniques and other approaches, will enhance accuracy.

Lastly, research should continue exploring collaborative cloud-edge frameworks for heterogeneous systems, incorporating robust designs with imperfect channel state information and investigating energy-efficient designs. The exploration of adaptive caching strategies and further optimizations in task scheduling to enhance the robustness and efficiency of frameworks is also essential. Future efforts should focus on improving resource management, enhancing security protocols, and developing intelligent orchestration mechanisms for urgent edge computing.

10.2 Emerging Trends and Future Directions

Emerging trends in digital communication optimization emphasize energy efficiency, particularly in 5G applications. The substantial energy demands of these technologies necessitate innovative solutions to reduce power consumption while maintaining high performance. This focus on energy efficiency is critical as 5G networks expand, driving the need for sustainable and cost-effective solutions.

In edge computing, there is a growing emphasis on developing a general model for edge video processing, aimed at standardizing the deployment of cooperative processing functions across heterogeneous edge nodes. Containerization technologies like Docker play a pivotal role in enabling such deployments, offering flexibility and scalability essential for diverse edge computing scenarios.

Moreover, the reliability of edge computing systems is a crucial area of focus, with new evaluation methods being developed to assess and enhance their dependability across various scenarios. These methods are essential for ensuring that edge computing meets the stringent reliability requirements of modern digital applications, providing robust and uninterrupted service delivery.

Looking ahead, future research directions in digital communication optimization are likely to explore the integration of advanced machine learning techniques for real-time resource management and dynamic task allocation. These techniques can significantly enhance the adaptability and efficiency of digital communication systems, enabling swift responses to changing network conditions and user demands.

Additionally, developing hybrid computing solutions that combine the strengths of edge, fog, and cloud computing will be pivotal in addressing the complex challenges of modern digital ecosystems. By leveraging the complementary capabilities of these computing paradigms, future systems can achieve greater scalability, flexibility, and efficiency, ultimately optimizing the performance and reliability of digital communication networks.

References

[1] Venkatraman Balasubramanian, Kees Kroep, Kishor Chandra Joshi, and R. Venkatesha Prasad. Reinforcing edge computing with multipath tcp enabled mobile device clouds, 2019.

[2] Lei Liu, Chen Chen, Qingqi Pei, Sabita Maharjan, and Yan Zhang. Vehicular edge computing and networking: A survey. *Mobile networks and applications*, 26:1145–1168, 2021.

[3] Sam Kayyali. Resource management and quality of service provisioning in 5g cellular networks, 2020.

[4] Seok-Hwan Park, Seongah Jeong, Jinyeop Na, Osvaldo Simeone, and Shlomo Shamai. Collaborative cloud and edge mobile computing in c-ran systems with minimal end-to-end latency, 2021.

[5] Arshin Rezazadeh and Hanan Lutfiyya. Multi-microservice migration modelling, comparison, and potential in 5g/6g mobile edge computing: A non-average parameter values approach, 2023.

[6] Akshitha Sriraman, Sihang Liu, Sinan Gunbay, Shan Su, and Thomas F. Wenisch. Deconstructing the tail at scale effect across network protocols, 2017.

[7] Patrizio Dazzi, Luca Ferrucci, Marco Danelutto, Konstantinos Tserpes, Antonis Makris, Theodoros Theodoropoulos, Jacopo Massa, Emanuele Carlini, and Matteo Mordacchini. Urgent edge computing, 2024.

[8] Dilip Bethanabhotla, Giuseppe Caire, and Michael J. Neely. Utility optimal scheduling and admission control for adaptive video streaming in small cell networks, 2013.

[9] Xiao Yang, Zhiyong Chen, Kuikui Li, Yaping Sun, and Hongming Zheng. Optimal task scheduling in communication-constrained mobile edge computing systems for wireless virtual reality, 2017.

[10] Intisar Al-Mejibli and Martin Colley Salah Al-Majeed. Networks utilization improvements for service discovery performance, 2011.

[11] Eugen Šlapak, Matúš Dopiriak, Mohammad Abdullah Al Faruque, Juraj Gazda, and Marco Levorato. Distributed radiance fields for edge video compression and metaverse integration in autonomous driving, 2024.

[12] Peixuan Song, JunKyu Lee, and Lev Mukhanov. A case study on latency, bandwidth and energy efficiency of mobile 5g and youtube edge service in london. why the 5g ecosystem and energy efficiency matter?, 2023.

[13] Xinyu Huang, Lijun He, Xin Chen, Fan Li, and Guizhong Liu. A more refined mobile edge cache replacement scheme for adaptive video streaming with mutual cooperation in multi-mec servers, 2019.

[14] Yang Cai, Jaime Llorca, Antonia M. Tulino, and Andreas F. Molisch. Optimal multicast service chain control: Packet processing, routing, and duplication, 2022.

[15] Sapna Chaudhary, Prince Sachdeva, Abhijit Mondal, Sandip Chakraborty, and Mukulika Maity. Youtube over google's quic vs internet middleboxes: A tug of war between protocol sustainability and application qoe, 2022.

[16] Sharat Chandra Madanapalli, Alex Mathai, Hassan Habibi Gharakheili, and Vijay Sivaraman. Modeling live video streaming: Real-time classification, qoe inference, and field evaluation, 2021.

[17] C. Briat, E. A. Yavuz, H. Hjalmarsson, K. H. Johansson, U. T. Jönsson, G. Karlsson, and H. Sandberg. The conservation of information, towards an axiomatized modular modeling approach to congestion control, 2013.

[18] Alfredo Petrella, Marco Miozzo, and Paolo Dini. Mobile traffic prediction at the edge through distributed and transfer learning, 2023.

[19] Nguyen Cong Luong, Ping Wang, Dusit Niyato, Wen Yonggang, and Zhu Han. Resource management in cloud networking using economic analysis and pricing models: A survey, 2017.

[20] Juan Luis Herrera, Jaime Galán-Jiménez, Javier Berrocal, and Juan Manuel Murillo. Optimizing response time in sdn-edge environments for time-strict iot applications, 2021.

[21] Maxim A. Kolosovskiy and Elena N. Kryuchkova. Network congestion control using netflow, 2009.

[22] Xiangqiang Gao, Rongke Liu, and Aryan Kaushik. A distributed virtual network function placement approach in satellite edge and cloud computing, 2021.

[23] Narges Mehran, Dragi Kimovski, and Radu Prodan. A two-sided matching model for data stream processing in the cloud-fog continuum, 2021.

[24] Imtiaz Parvez, Ali Rahmati, Ismail Guvenc, Arif I. Sarwat, and Huaiyu Dai. A survey on low latency towards 5g: Ran, core network and caching solutions, 2018.

[25] Lorenzo Testa, Giorgio Audrito, Ferruccio Damiani, and Gianluca Torta. Aggregate processes as distributed adaptive services for the industrial internet of things, 2022.

[26] Haijian Sun, Zekun Zhang, Rose Qingyang Hu, and Yi Qian. Wearable communications in 5g: Challenges and enabling technologies, 2018.

[27] Andrea Hamm, Alexander Willner, and Ina Schieferdecker. Edge computing: A comprehensive survey of current initiatives and a roadmap for a sustainable edge computing development, 2019.

[28] Mahadev Satyanarayanan. The emergence of edge computing. *Computer*, 50(1):30–39, 2017.

[29] Yogesh Simmhan. Big data and fog computing, 2017.

[30] Koustabh Dolui and Soumya Kanti Datta. Comparison of edge computing implementations: Fog computing, cloudlet and mobile edge computing. In *2017 Global Internet of Things Summit (GIoTS)*, pages 1–6. IEEE, 2017.

[31] Amal A. Alahmadi, T. E. H. El-Gorashi, and Jaafar M. H. Elmirghani. Energy efficient and delay aware vehicular edge cloud, 2020.

[32] Contents lists available at scie.

[33] Leonardo Gobatto, Mateus Saquetti, Claudio Diniz, Bruno Zatt, Weverton Cordeiro, and Jose Rodrigo Azambuja. Improving content-aware video streaming in congested networks with in-network computing, 2022.

[34] Anurag Ghosh, Srinivasan Iyengar, Stephen Lee, Anuj Rathore, and Venkat N Padmanabhan. Streaming video analytics on the edge with asynchronous cloud support, 2022.

[35] Devi Klein, Josef Spjut, Ben Boudaoud, and Joohwan Kim. The influence of variable frame timing on first-person gaming, 2023.

[36] Miao Hu, Zhenxiao Luo, Amirmohammad Pasdar, Young Choon Lee, Yipeng Zhou, and Di Wu. Edge-based video analytics: A survey, 2023.

[37] Sohaib Ahmed, Hassan Khalid, Muhammad Hamza, and Danyal Farhat. Mobile edge computing, 2024.

[38] Alcardo Alex Barakabitze, Nabajeet Barman, Arslan Ahmad, Saman Zadtootaghaj, Lingfen Sun, Maria G. Martini, and Luigi Atzori. Qoe management of multimedia streaming services in future networks: A tutorial and survey, 2019.

[39] Babak Badnava, Jacob Chakareski, and Morteza Hashemi. Multi-task decision-making for multi-user 360 video processing over wireless networks, 2024.

[40] Quoc-Viet Pham, Fang Fang, Vu Nguyen Ha, Md. Jalil Piran, Mai Le, Long Bao Le, Won-Joo Hwang, and Zhiguo Ding. A survey of multi-access edge computing in 5g and beyond: Fundamentals, technology integration, and state-of-the-art, 2020.

[41] Shah Zeb, Muhammad Ahmad Rathore, Aamir Mahmood, Syed Ali Hassan, JongWon Kim, and Mikael Gidlund. Edge intelligence in softwarized 6g: Deep learning-enabled network traffic predictions, 2021.

[42] Francesco De Pellegrini, Antonio Massaro, Leonardo Goratti, and Rachid El-Azouzi. Competitive caching of contents in 5g edge cloud networks, 2016.

[43] Yang Cai, Jaime Llorca, Antonia M. Tulino, and Andreas F. Molisch. Decentralized control of distributed cloud networks with generalized network flows, 2022.

[44] Ilias Benkacem, M. Bagaa, T. Taleb, Q. N. Nguyen, T. Tsuda, and T. Sato. Integrated icn and cdn slice as a service, 2022.

[45] Raz Segal, Chen Avin, and Gabriel Scalosub. Constrained in-network computing with low congestion in datacenter networks, 2022.

[46] Anishi Mehta. Time series analysis and forecasting of internet congestion data, 2018.

[47] Tongyu Dai, Xinggong Zhang, Yihang Zhang, and Zongming Guo. Statistical learning based congestion control for real-time video communication, 2019.

[48] Sanaullah Manzoor, Ghulam Abbas, and Masroor Hussain. Choked: A fair active queue management system, 2017.

[49] Meysam Nasimi, Mohammad Asif Habibi, Bin Han, and Hans D Schotten. Edge-assisted congestion control mechanism for 5g network using software-defined networking. In *2018 15th International symposium on wireless communication systems (ISWCS)*, pages 1–5. IEEE, 2018.

[50] Emmanuel Lochin and Bruno Talavera. Managing network congestion with a kohonen-based red queue, 2007.

[51] Chen Tessler, Yuval Shpigelman, Gal Dalal, Amit Mandelbaum, Doron Haritan Kazakov, Benjamin Fuhrer, Gal Chechik, and Shie Mannor. Reinforcement learning for datacenter congestion control, 2022.

[52] Wei Li, Fan Zhou, Kaushik Roy Chowdhury, and Waleed Meleis. Qtcp: Adaptive congestion control with reinforcement learning. *IEEE Transactions on Network Science and Engineering*, 6(3):445–458, 2018.

[53] Tommaso Bonato, Abdul Kabbani, Daniele De Sensi, Rong Pan, Yanfang Le, Costin Raiciu, Mark Handley, Timo Schneider, Nils Blach, Ahmad Ghalayini, Daniel Alves, Michael Papamichael, Adrian Caulfield, and Torsten Hoefler. Fastflow: Flexible adaptive congestion control for high-performance datacenters, 2024.

[54] Nehal Baganal-Krishna, Tuan-Dat Tran, Ralf Kundel, and Amr Rizk. Accelerating end-host congestion response using p4 programmable switches, 2023.

[55] Qian Liu, Robert D. Russell, Fabrice Mizero, Malathi Veeraraghavan, John Dennis, and Benjamin Jamroz. Implementation of pfc and rcm for rocev2 simulation in omnet++, 2015.

[56] Viswanath Sivakumar, Olivier Delalleau, Tim Rocktäschel, Alexander H. Miller, Heinrich Küttler, Nantas Nardelli, Mike Rabbat, Joelle Pineau, and Sebastian Riedel. Mvfst-rl: An asynchronous rl framework for congestion control with delayed actions, 2021.

[57] Walid A. Hanafy, Limin Wang, Hyunseok Chang, Sarit Mukherjee, T. V. Lakshman, and Prashant Shenoy. Understanding the benefits of hardware-accelerated communication in model-serving applications, 2023.

[58] Kanika Sharma, Bernard Butler, and Brendan Jennings. Graph-based heuristic solution for placing distributed video processing applications on moving vehicle clusters, 2021.

[59] Yang Liu, Xiaoping Liu, Yuanwei Jing, and Shaowei Zhou. Adaptive backstepping h tracking control with prescribed performance for internet congestion. *ISA transactions*, 72:92–99, 2018.

[60] Ashkan Aghdai, Michael I. C. Wang, Yang Xu, Charles H. P. Wen, and H. Jonathan Chao. In-network congestion-aware load balancing at transport layer, 2019.

[61] Ying Liu, Hongying Liu, Ke Xu, Meng Shen, and Yifeng Zhong. A large family of multi-path dual congestion control algorithms, 2011.

[62] Ashkan Moharrami, Mohammad Ghasempour, and Mohammad Ghanbari. An in-router identification scheme for selective discard of video packets, 2021.

[63] Jiachen Xue, Muhammad Usama Chaudhry, Balajee Vamanan, T. N. Vijaykumar, and Mithuna Thottethodi. Dart: Divide and specialize for fast response to congestion in rdma-based datacenter networks, 2019.

[64] Quantifying the impact of networ.

[65] Abuthahir Abuthahir, Nizar Malangadan, and Gaurav Raina. Effect of two forms of feedback on the performance of the rate control protocol (rcp), 2019.

[66] Marwah Almasri, Khaled Elleithy, and Abdul Razaque. Analytical study of pre-congestion notification (pcn) techniques, 2012.

[67] Huiling Jiang, Qing Li, Yong Jiang, Gengbiao Shen, Richard Sinnott, Chen Tian, and Mingwei Xu. When machine learning meets congestion control: A survey and comparison. *Computer Networks*, 192:108033, 2021.

[68] Inayat Ali, Seungwoo Hong, and Taesik Cheung. Congestion or no congestion: Packet loss identification and prediction using machine learning, 2024.

[69] Hoang-Loc La, Anh-Tu Ngoc Tran, Quang-Trai Le, Masato Yoshimi, Takuma Nakajima, and Nam Thoai. A use case of content delivery network raw logfile analysis, 2021.

[70] Giovanna Carofiglio and Luca Muscariello. On the impact of tcp and per-flow scheduling on internet performance (extended version), 2009.

[71] Sminesh C. N., Grace Mary Kanaga E., and Ranjitha K. A proactive flow admission and re-routing scheme for load balancing and mitigation of congestion propagation in sdn data plane, 2018.

[72] G. F. Ali Ahammed and Reshma Banu. Analyzing the performance of active queue management algorithms, 2010.

[73] Soheil Abbasloo. Internet congestion control benchmarking, 2023.

[74] Yang Cao, Zeyu Xu, Peng Qin, and Tao Jiang. Video processing on the edge for multimedia iot systems, 2018.

[75] Xiang Chen, Wenjie Zhu, Jiayuan Chen, Tong Zhang, Changyan Yi, and Jun Cai. Edge computing enabled real-time video analysis via adaptive spatial-temporal semantic filtering, 2024.

[76] Karim Sonbol, Öznur Özkasap, Ibrahim Al-Oqily, and Moayad Aloqaily. Edgekv: Decentralized, scalable, and consistent storage for the edge, 2020.

[77] Jennifer Andreoli-Fang and John T Chapman. Mobile-aware scheduling for low latency backhaul over docsis, 2022.

[78] Suvadip Batabyal. How svc enables distributed caching in mec?, 2022.

[79] Viswa Narayanan Sankaranarayanan, Gerasimos Damigos, Achilleas Santi Seisa, Sumeet Gajanan Satpute, Tore Lindgren, and George Nikolakopoulos. Paced-5g: Predictive autonomous control using edge for drones over 5g, 2023.

[80] Shengmei Liu, Xiaokun Xu, and Mark Claypool. A survey and taxonomy of latency compensation techniques for network computer games. *ACM Computing Surveys (CSUR)*, 54(11s):1–34, 2022.

[81] Zhuo Li, Xu Zhou, Taixin Li, and Yang Liu. An optimal-transport-based reinforcement learning approach for computation offloading, 2021.

[82] Jinkun Zhang, Yuezhou Liu, and Edmund Yeh. Result and congestion aware optimal routing and partial offloading in collaborative edge computing, 2022.

[83] Noah Martin and Fahad Dogar. Measuring latency reduction and the digital divide of cloud edge datacenters, 2023.

[84] Xiaolin Jiang, Hossein Shokri-Ghadikolaei, Gabor Fodor, Eytan Modiano, Zhibo Pang, Michele Zorzi, and Carlo Fischione. Low-latency networking: Where latency lurks and how to tame it. *Proceedings of the IEEE*, 107(2):280–306, 2018.

[85] Xinyu Huang, Lijun He, Liejun Wang, and Fan Li. Towards 5g: Joint optimization of video segment cache, transcoding and resource allocation for adaptive video streaming in a muti-access edge computing network, 2020.

[86] Mathieu Simon, Alessandro Spallina, Loic Dubocquet, and Andrea Araldo. Parsimonious edge computing to reduce microservice resource usage, 2021.

[87] Xiangqiang Gao, Rongke Liu, and Aryan Kaushik. Virtual network function placement in satellite edge computing with a potential game approach, 2021.

[88] Jihong Park, Petar Popovski, and Osvaldo Simeone. Minimizing latency to support vr social interactions over wireless cellular systems via bandwidth allocation, 2018.

[89] Cheng Zhang, Yinuo Deng, Hailiang Zhao, Tianlv Chen, and Shuiguang Deng. Tail-learning: Adaptive learning method for mitigating tail latency in autonomous edge systems, 2023.

[90] Yijun Hao, Shusen Yang, Fang Li, Yifan Zhang, Shibo Wang, and Xuebin Ren. Edgetimer: Adaptive multi-timescale scheduling in mobile edge computing with deep reinforcement learning, 2024.

[91] Cheol-Ho Hong and Blesson Varghese. Resource management in fog/edge computing: A survey, 2018.

[92] Toke Høiland-Jørgensen, Michał Kazior, Dave Täht, Per Hurtig, and Anna Brunstrom. Ending the anomaly: Achieving low latency and airtime fairness in wifi, 2017.

[93] Jialei Liu, Ao Zhou, Chunhong Liu, Tongguang Zhang, Lianyong Qi, Shangguang Wang, and Rajkumar Buyya. Reliability-enhanced task offloading in mobile edge computing environments. *IEEE Internet of Things Journal*, 9(13):10382–10396, 2021.

[94] Prabhu Kaliyammal Thiruvasagam, Abhishek Chakraborty, and C. Siva Ram Murthy. Latency-aware and survivable mapping of vnfs in 5g network edge cloud, 2021.

[95] Monitoring and proactive management of qos levels in pervasive applications.

[96] Md Washik Al Azad and Spyridon Mastorakis. The promise and challenges of computation deduplication and reuse at the network edge, 2022.

[97] Panagiotis Oikonomou, Anna Karanika, Christos Anagnostopoulos, and Kostas Kolomvatsos. On the road from edge computing to the edge mesh, 2020.

[98] Saurabh Jha, Archit Patke, Jim Brandt, Ann Gentile, Mike Showerman, Eric Roman, Zbigniew T Kalbarczyk, Bill Kramer, and Ravishankar K Iyer. A study of network congestion in two supercomputing high-speed interconnects. In *2019 IEEE Symposium on High-Performance Interconnects (HOTI)*, pages 45–48. IEEE, 2019.

[99] Anastasia-Dimitra Lipitakis, George Kousiouris, Mara Nikolaidou, Cleopatra Bardaki, and Dimosthenis Anagnostopoulos. Empirical investigation of factors influencing function as a service performance in different cloud/edge system setups, 2023.

[100] Francisco Carpio, Marta Delgado, and Admela Jukan. Engineering and experimentally benchmarking a container-based edge computing system, 2020.

[101] G. F. Ali Ahammed and Reshma Banu. Anakyzing the performance of active queue management algorithms, 2010.

[102] David Candal-Ventureira, Francisco Javier González-Castaño, Felipe Gil-Castiñeira, and Pablo Fondo-Ferreiro. Is the edge really necessary for drone computing offloading? an experimental assessment in carrier-grade 5g operator networks, 2024.

[103] Chiara Caiazza, Silvia Giordano, Valerio Luconi, and Alessio Vecchio. Edge computing vs centralized cloud: Impact of communication latency on the energy consumption of lte terminal nodes, 2023.

[104] Ali Al-Shuwaili and Osvaldo Simeone. Energy-efficient resource allocation for mobile edge computing-based augmented reality applications, 2017.

[105] Jiadong Yu, Ahmad Alhilal, Tailin Zhou, Pan Hui, and Danny H. K. Tsang. Attention-based qoe-aware digital twin empowered edge computing for immersive virtual reality, 2023.

[106] Jin Heo, Ketan Bhardwaj, and Ada Gavrilovska. Poster: Enabling flexible edge-assisted xr, 2023.

[107] Sam Aleyadeh, Abdallah Moubayed, and Abdallah Shami. Mobility aware edge computing segmentation towards localized orchestration, 2021.

[108] Fei Hu, Kunal Mehta, Shivakant Mishra, and Mohammad AlMutawa. A dynamic distributed scheduler for computing on the edge, 2023.

[109] Boubakr Nour and Soumaya Cherkaoui. Matching-based service offloading for compute-less driven iot networks, 2022.

[110] Yang Cai, Jaime Llorca, Antonia M. Tulino, and Andreas F. Molisch. Mobile edge computing network control: Tradeoff between delay and cost, 2022.

[111] Xiaofeng Cao, Guoming Tang, Deke Guo, Yan Li, and Weiming Zhang. Edge federation: Towards an integrated service provisioning model, 2020.

[112] Arkadiusz Madej, Nan Wang, Nikolaos Athanasopoulos, Rajiv Ranjan, and Blesson Varghese. Priority-based fair scheduling in edge computing, 2020.

[113] Li Lin, Peng Li, Jinbo Xiong, and Mingwei Lin. Distributed and application-aware task scheduling in edge-clouds, 2019.

[114] Md Washik Al Azad and Spyridon Mastorakis. Deduplicator: When computation reuse meets load balancing at the network edge, 2024.

[115] Eugenio Moro and Ilario Filippini. Joint management of compute and radio resources in mobile edge computing: a market equilibrium approach, 2020.

[116] Faria Kalim and Shadi A. Noghabi. Bené: On demand cost-effective scaling at the edge, 2018.

[117] Weibin Ma and Lena Mashayekhy. Truthful computation offloading mechanisms for edge computing, 2020.

[118] Mingjin Zhang, Jiannong Cao, Lei Yang, Liang Zhang, Yuvraj Sahni, and Shan Jiang. Ents: An edge-native task scheduling system for collaborative edge computing, 2022.

[119] Richard Olaniyan, Olamilekan Fadahunsi, Muthucumaru Maheswaran, and Mohamed Faten Zhani. Opportunistic edge computing: Concepts, opportunities and research challenges, 2018.

[120] Ahmed Ali-Eldin, Bin Wang, and Prashant Shenoy. The hidden cost of the edge: A performance comparison of edge and cloud latencies, 2021.

[121] Chao Yao, Xiaoyang Wang, Zijie Zheng, Guangyu Sun, and Lingyang Song. Edgeflow: Open-source multi-layer data flow processing in edge computing for 5g and beyond, 2018.

[122] Sahraoui Dhelim, Tahar Kechadi, Liming Chen, Nyothiri Aung, Huansheng Ning, and Luigi Atzori. Edge-enabled metaverse: The convergence of metaverse and mobile edge computing, 2022.

[123] Tero Lähderanta, Teemu Leppänen, Leena Ruha, Lauri Lovén, Erkki Harjula, Mika Ylianttila, Jukka Riekki, and Mikko J. Sillanpää. Edge computing server placement with capacitated location allocation, 2021.

[124] TianZhang He, Adel N Toosi, and Rajkumar Buyya. Efficient large-scale multiple migration planning and scheduling in sdn-enabled edge computing, 2021.

[125] Colton Powell, Christopher Desiniotis, and Behnam Dezfouli. The fog development kit: A development platform for sdn-based edge-fog systems, 2020.

[126] Jesús Aguilar-Armijo, Ekrem Çetinkaya, Christian Timmerer, and Hermann Hellwagner. Ecas-ml: Edge computing assisted adaptation scheme with machine learning for http adaptive streaming, 2022.

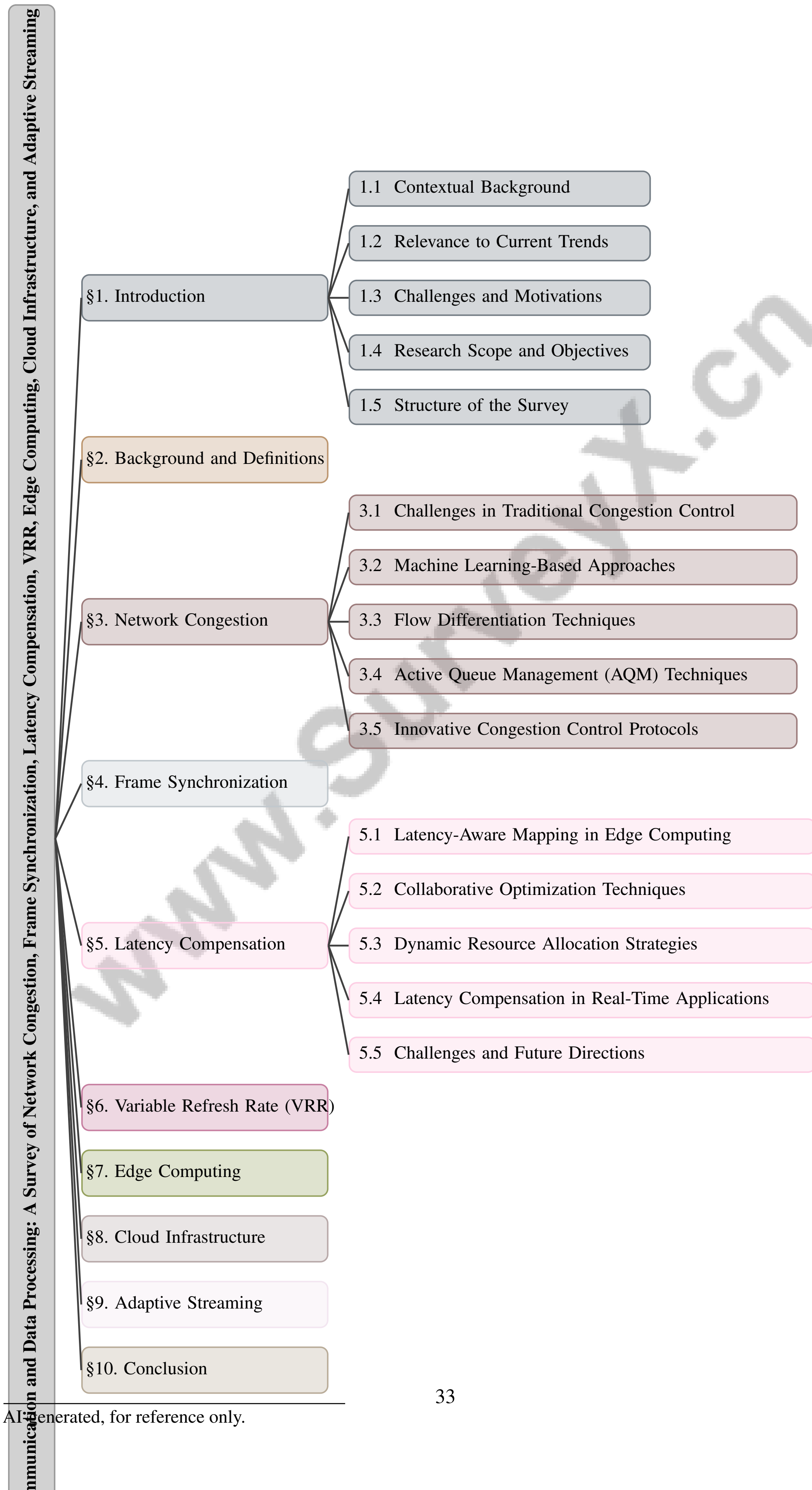
[127] Lin Gao, Ming Tang, Haitian Pang, Jianwei Huang, and Lifeng Sun. Multi-user cooperative mobile video streaming: Performance analysis and online mechanism design, 2018.

[128] Abbas Mehrabi, Matti Siekkinen, and Antti Ylä-Jääski. Cache-aware qoe-traffic optimization in mobile edge assisted adaptive video streaming, 2018.

[129] Harshit Gupta and Ajay Kumar Bharti. Fog computing iot: Overview, architecture and applications, 2023.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.


mmunication and Data Processing: A Survey of Network Congestion, Frame Synchronization, Latency Compensation, VRR, Edge Computing, Cloud Infrastructure, and Adaptive Streaming
§1. Introduction
1.1 Contextual Background
1.2 Relevance to Current Trends
1.3 Challenges and Motivations
1.4 Research Scope and Objectives
1.5 Structure of the Survey
§2. Background and Definitions
§3. Network Congestion
3.1 Challenges in Traditional Congestion Control
3.2 Machine Learning-Based Approaches
3.3 Flow Differentiation Techniques
3.4 Active Queue Management (AQM) Techniques
3.5 Innovative Congestion Control Protocols
§4. Frame Synchronization
§5. Latency Compensation
5.1 Latency-Aware Mapping in Edge Computing
5.2 Collaborative Optimization Techniques
5.3 Dynamic Resource Allocation Strategies
5.4 Latency Compensation in Real-Time Applications
5.5 Challenges and Future Directions
§6. Variable Refresh Rate (VRR)
§7. Edge Computing
§8. Cloud Infrastructure
§9. Adaptive Streaming
§10. Conclusion

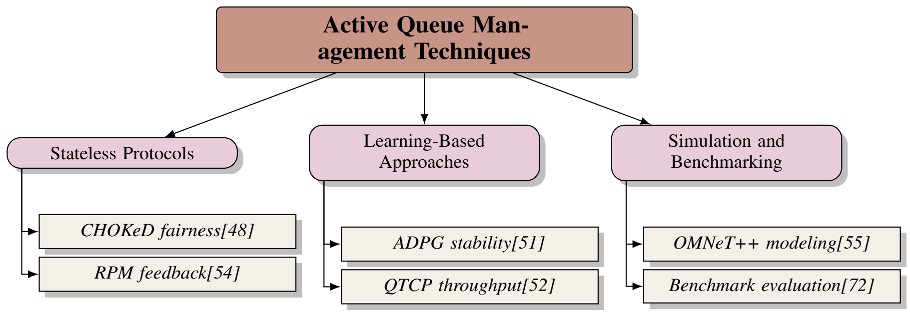


Figure 2: This figure illustrates the categorization of Active Queue Management (AQM) techniques, highlighting stateless protocols, learning-based approaches, and simulation/benchmarking efforts to manage network congestion effectively.

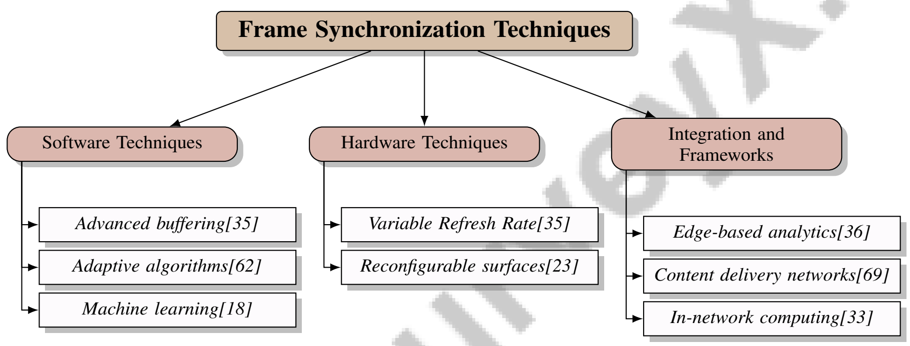


Figure 3: This figure illustrates the key techniques for achieving frame synchronization, categorized into software and hardware methods, as well as integrated frameworks. Software techniques include advanced buffering, adaptive algorithms, and machine learning. Hardware techniques involve variable refresh rate and reconfigurable intelligent surfaces. Integration and frameworks highlight edge-based video analytics, content delivery networks, and in-network computing, showcasing a comprehensive approach to optimizing video frame synchronization.

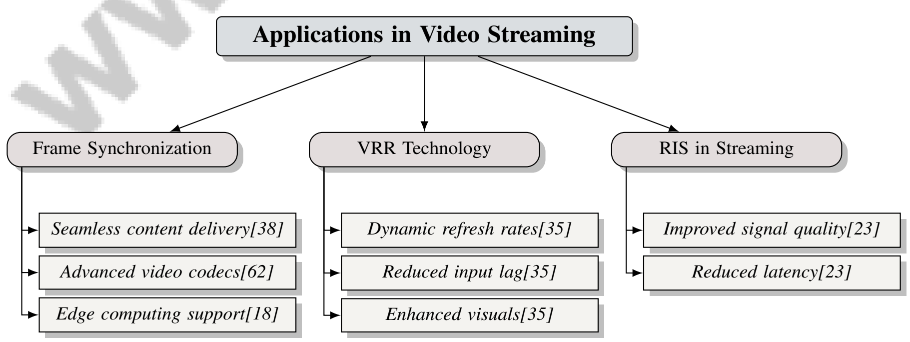


Figure 4: This figure illustrates the key advancements in video streaming applications, focusing on frame synchronization, VRR technology, and RIS in streaming. Frame synchronization ensures seamless content delivery through advanced codecs and edge computing. VRR technology enhances video playback by dynamically adjusting refresh rates, reducing input lag, and improving visuals. RIS technology contributes to better signal quality and reduced latency in streaming environments.

AI-generated, for reference only.

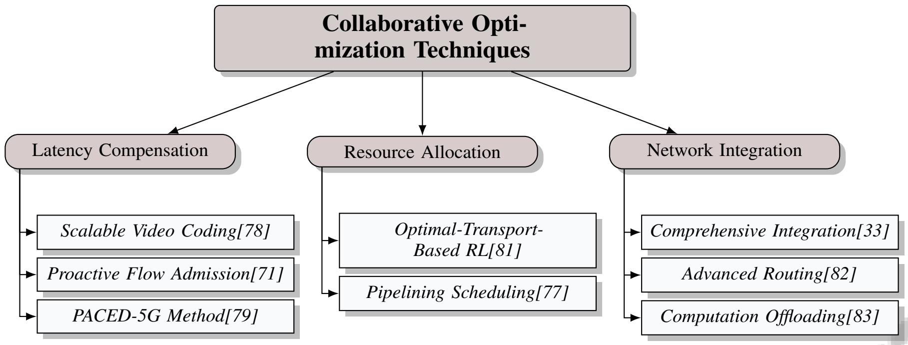
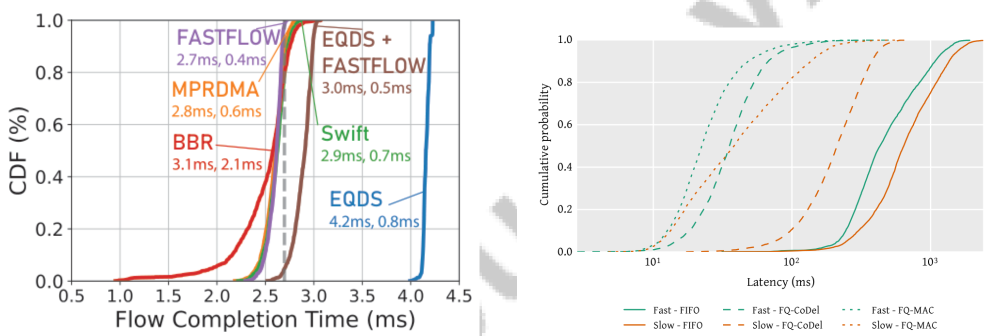


Figure 5: This figure illustrates the hierarchical classification of collaborative optimization techniques in network latency management, emphasizing latency compensation, resource allocation, and network integration strategies.



(a) Comparison of Flow Completion Times for Different Flowing Algorithms[53]

(b) Comparison of Latency Distribution for Different Protocols[92]

Figure 6: Examples of Latency Compensation in Real-Time Applications

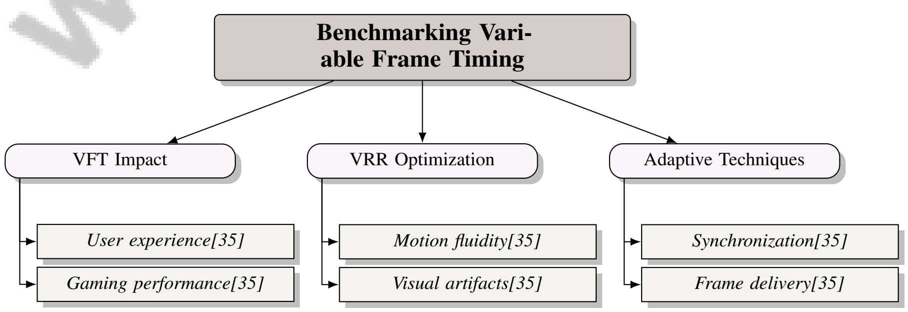


Figure 7: This figure illustrates the hierarchical structure of benchmarking Variable Frame Timing (VFT) in gaming, focusing on its impact on user experience, VRR optimization, and the development of adaptive techniques for enhanced gaming performance.

AI-generated, for reference only.