A Survey on Content Delivery Networks: Technologies and Strategies for Optimization

www.surveyx.cn

Abstract

Content Delivery Networks (CDNs) are pivotal in addressing the escalating demand for high-quality internet content, particularly video streaming. This survey paper provides a comprehensive analysis of the technologies and strategies employed in CDNs to optimize content distribution and enhance user experience. Key areas explored include cache replacement policies, Dynamic Adaptive Streaming over HTTP (DASH), Quality of Experience (QoE) optimization, edge server placement, bandwidth cost reduction, and network optimization for latency reduction. Traditional and advanced cache replacement policies, including machine learning-based approaches, are examined for their impact on performance. The integration of DASH with CDNs, leveraging Software Defined Networking (SDN) and Network Functions Virtualization (NFV), is highlighted for its role in improving video streaming quality. QoE optimization strategies focus on adaptive streaming and error correction to enhance user satisfaction. Edge server placement strategies, employing heuristic models and emerging technologies, are critical for reducing latency and improving content accessibility. The survey also addresses bandwidth cost reduction techniques, emphasizing pricing models and the utilization of underutilized resources. The role of emerging technologies in network optimization, particularly SDN and NFV, is discussed for their contribution to latency reduction. The paper concludes by emphasizing the necessity for continuous innovation and research in CDN technologies to meet the evolving demands of modern internet infrastructure, suggesting future research directions in resource allocation, security, and scalability. By leveraging advanced technologies and strategic methodologies, CDNs can enhance content delivery efficiency, scalability, and user satisfaction, paving the way for future advancements in content distribution networks.

1 Introduction

1.1 Overview of Content Delivery Networks

Content Delivery Networks (CDNs) are essential to modern internet infrastructure, addressing the significant increase in internet traffic, particularly from video content, which was projected to account for 81% of all internet traffic by 2021 [1]. This surge presents challenges for Internet Service Providers (ISPs) and negatively impacts the Quality of Experience (QoE) for users. CDNs alleviate network congestion and enhance QoE by distributing content closer to users, thereby reducing latency and improving load times [2].

The rapid growth of video streaming and mobile data traffic, fueled by high-resolution content and the rise of Internet of Things (IoT) applications, underscores the critical role of CDNs in managing these demands [3]. In 2015, mobile video traffic constituted 55% of total mobile data traffic, with expectations to rise to 75% by 2020, making CDNs indispensable [4].

To meet evolving demands, CDNs have integrated advanced technologies such as Network Functions Virtualization (NFV) and Software Defined Networking (SDN), which enable dynamic resource

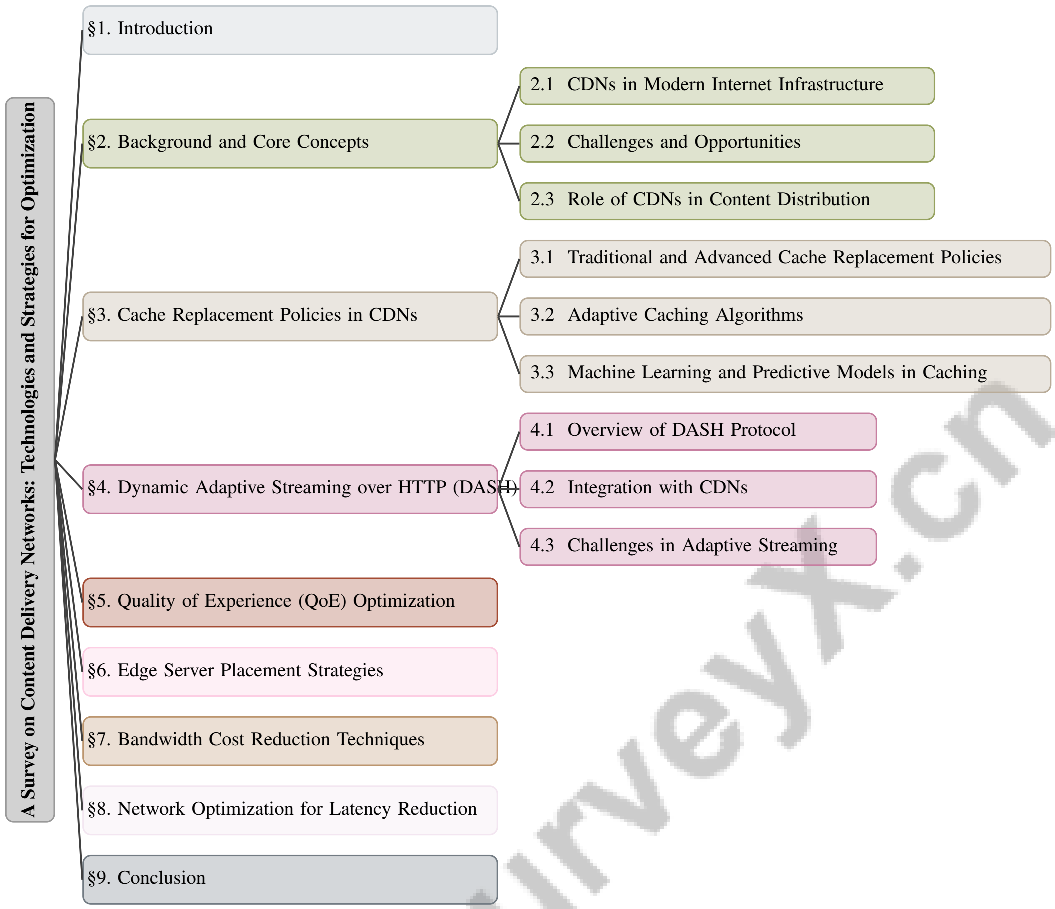

Figure 1: chapter structure

allocation and improved network management [5]. These technologies enhance the flexibility and cost-effectiveness of CDN services, addressing traditional content delivery inefficiencies. Furthermore, the incorporation of edge-cloud computing, which shifts computations from centralized cloud servers to edge nodes, significantly improves content delivery efficiency and responsiveness [6]. Platforms like WebAssembly are also employed to optimize service distribution and deployment.

Energy efficiency has emerged as a crucial design consideration for CDNs, ensuring sustainability for future internet infrastructure [2]. Innovations like InviCloak enhance CDN performance while preserving user privacy by preventing the sharing of sensitive keys with CDN providers [7].

The economic and technical challenges within the internet content delivery market necessitate strategic methodologies for CDN deployment [8]. As the web evolves, the increase in dynamic content makes web page loading times critical for user experience, further emphasizing the importance of efficient CDNs [9].

1.2 Structure of the Survey

This survey provides a comprehensive analysis of the technologies and strategies employed in Content Delivery Networks (CDNs) to optimize content distribution and enhance user experience. Initial sections introduce CDNs and their integration into modern internet infrastructure, followed by an examination of the challenges and opportunities they face, elucidating their pivotal role in content distribution.

Subsequent sections focus on specific technological strategies crucial for CDN optimization, starting with cache replacement policies. Both traditional methods, such as Least Recently Used (LRU) and Least Frequently Used (LFU), and advanced machine learning-based approaches are analyzed for their impact on performance and efficiency. The survey then transitions to Dynamic Adaptive

AI-generated, for reference only.

Streaming over HTTP (DASH), highlighting its significance in adaptive video streaming, particularly for long-form two-dimensional videos streamed using HTTP adaptive streaming (HAS) with client-side adaptive bitrate (ABR) algorithms [10].

The discussion extends to Quality of Experience (QoE) optimization strategies, emphasizing adaptive streaming techniques and their role in enhancing user satisfaction. Additionally, edge server placement strategies are examined, critical for reducing latency and improving content accessibility. The survey analyzes techniques for reducing bandwidth costs through various pricing models and innovative methods for leveraging underutilized resources, alongside the evolution of Internet resource pricing strategies that facilitate effective resource allocation and utility optimization. Moreover, it highlights the importance of adaptive streaming technologies and caching strategies in CDNs, aiming to enhance user experience while minimizing operational expenses [11, 12, 13].

In the final sections, the focus shifts to network optimization for latency reduction, exploring the role of emerging technologies such as Software Defined Networking (SDN) and Network Functions Virtualization (NFV) in enhancing network performance. The survey concludes by summarizing key findings and insights, underscoring the necessity to increase CDN efficiency to support future streaming scenarios involving millions of users [14], and discussing potential future research and development directions in CDN technologies.The following sections are organized as shown in Figure 1.

2 Background and Core Concepts

2.1 CDNs in Modern Internet Infrastructure

Content Delivery Networks (CDNs) are essential to modern internet infrastructure, addressing the rising demands of high-resolution video-on-demand (VoD) and live streaming services. These services require effective delay and bandwidth management, prompting the deployment of CDNs to alleviate network congestion and optimize resource distribution by positioning cache servers closer to end-users, thereby enhancing scalability and reducing latency [15]. Proprietary CDN architectures can lead to inefficiencies in meeting user demands, compounded by the cache replacement problem that involves selecting data for eviction to optimize future cache hits [16, 17]. The performance and user satisfaction of CDNs, particularly in video streaming, are significantly influenced by infrastructure and user access patterns [18].

Economic considerations are crucial in CDN integration, particularly in adopting new protocols by Internet Service Providers (ISPs) [19]. Platforms like Facebook and Netflix exemplify the necessity for robust CDN solutions to minimize access delays and maintain high user experience standards. Facebook's infrastructure supports over a billion users, while Netflix's Open Connect CDN utilizes global Internet Exchange Points (IXPs) and ISPs, underscoring efficient content distribution strategies. Emerging technologies such as Fog Radio Access Networks (F-RANs) leverage fog access points (FAPs) for caching and resource management, enhancing content delivery with reduced latency in the 5G era [20]. Additionally, energy consumption and the carbon footprint of video streaming are critical considerations, prompting methodologies to assess network infrastructure usage impact [21].

The evolution of CDNs is supported by structured approaches to application-level caching, integrating static and adaptive strategies to improve performance [22]. As CDNs advance, they address challenges related to dynamic content, content freshness, and aging costs, providing essential infrastructure for efficient content distribution [23]. The consolidation of DNS and web hosting services highlights the need to understand vulnerabilities and dependencies within internet services, offering a framework for assessing CDN architecture robustness [24].

2.2 Challenges and Opportunities

CDNs face numerous challenges affecting performance, scalability, and security while offering opportunities for optimization and innovation. A primary challenge is current mobile network designs' inability to dynamically scale and adapt to fluctuating traffic demands, degrading video delivery quality [25]. Hierarchical telco architectures requiring multiple electronic conversions inefficiently manage increasing data traffic and service demands [26]. Effective application-level caching implementation and management remain critical, requiring informed decisions on content caching [22]. Traditional caching methods often fail to adapt to varying user demands and distribution

AI-generated, for reference only.

requirements, resulting in inefficiencies [27]. Existing CDN architectures struggle to meet growing video content delivery demand, increasing latency and resource wastage [28]. Overlay multicast networks also present challenges, as server and network bottlenecks, including limited server capacity and unpredictable packet loss, hinder performance [29]. Additionally, ISPs face difficulties in monitoring and understanding changes within evolving CDNs like YouTube, affecting user Quality of Experience (QoE) [30].

Security concerns are prominent in CDNs, particularly regarding Peer-assisted Delivery Networks (PDNs) used in video streaming, which introduce unexamined security risks threatening content delivery [31]. Despite these challenges, significant opportunities for innovation in CDN technologies exist. Advances in caching strategies, such as size-aware policies, can enhance performance in modern databases and data stores [32]. Integrating blockchain-based solutions offers promising avenues to address inefficiencies and privacy concerns, potentially reducing multiple subscriptions and lowering operational costs [33]. Additionally, optimizing edge resource utilization to prevent node overloading while ensuring balanced distribution remains a critical opportunity for enhancing CDN efficiency [34].

2.3 Role of CDNs in Content Distribution

CDNs are vital for efficiently distributing internet content, particularly amid increasing demand for high-quality video and evolving mobile network dynamics. By strategically positioning servers near end-users, CDNs minimize latency and significantly enhance Quality of Experience (QoE), essential for accommodating rising mobile video traffic driven by 5G and IoT advancements. CDNs replicate content across geographically dispersed servers, allowing rapid request fulfillment from the nearest server, improving service reliability and speed. They employ sophisticated routing techniques and overlay structures to identify optimal data paths, ensuring efficient video content delivery while adhering to critical Quality of Service (QoS) parameters such as jitter and packet loss [35, 36, 37].

Integrating advanced technologies like Information-Centric Networking (ICN) with CDN architectures addresses scalability, reliability, and quality of service challenges amid increasing mobile video traffic [27]. This integration optimizes resource management and service delivery by leveraging virtualization capabilities [28]. Anycast addressing in CDNs facilitates distributed load management across multiple proxy nodes, though it presents challenges in avoiding overload situations [38]. In video streaming, Peer-assisted Delivery Networks (PDNs) utilize peer-to-peer connections for content delivery but introduce potential security vulnerabilities requiring careful management to protect users [31]. The complexity of modeling energy consumption across CDN elements, including user devices and network components, complicates content delivery network optimization [2].

In the modern digital ecosystem, CDNs are crucial for efficiently distributing content, alleviating network congestion from the exponential growth of video streaming and bandwidth-intensive applications, managing network resources effectively, and ensuring a consistent and enhanced user experience while addressing challenges related to performance, privacy, and legal compliance [36, 39, 40, 18]. By adapting to challenges posed by new network architectures and traffic patterns, CDNs continue to support efficient content distribution in an increasingly connected world.

3 Cache Replacement Policies in CDNs

The effectiveness of cache replacement policies in Content Delivery Networks (CDNs) is crucial for optimizing content delivery and resource utilization, directly influencing the speed and efficiency of content retrieval. These policies strategically manage the storage of frequently accessed files closer to users, which is essential for reducing latency and enhancing download rates in response to the growing demand for high-quality media experiences [15, 13]. A comprehensive exploration of both traditional and advanced cache management strategies is necessary to address evolving challenges in this domain.

As illustrated in Figure 2, the hierarchical classification of cache replacement policies in CDNs distinguishes between traditional and advanced methods, including adaptive caching algorithms and machine learning predictive models. This figure outlines the evolution of strategies from basic approaches such as Least Recently Used (LRU) and Least Frequently Used (LFU) to more sophisticated machine learning techniques. It highlights key innovations, such as Startrail and LRU-

BaSE, while emphasizing the importance of adaptive and predictive techniques in optimizing content delivery networks.

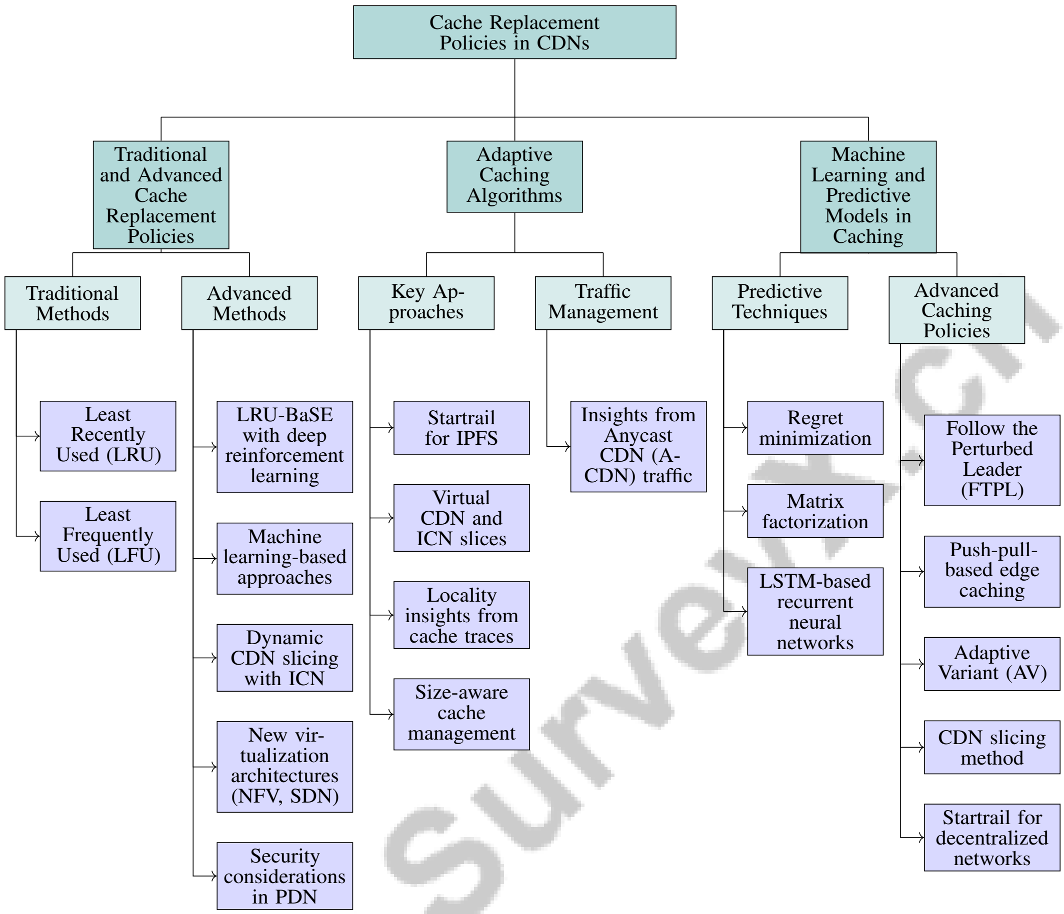

Figure 2: This figure shows a hierarchical classification of cache replacement policies in CDNs, distinguishing between traditional and advanced methods, adaptive caching algorithms, and machine learning and predictive models. It outlines the evolution of strategies from basic LRU and LFU to sophisticated machine learning approaches, highlighting key innovations like Startrail and LRU-BaSE, and emphasizing the importance of adaptive and predictive techniques in optimizing content delivery networks.

3.1 Traditional and Advanced Cache Replacement Policies

Cache replacement policies are vital for CDN efficiency, dictating which content to retain or evict from cache storage to optimize user experience by improving content availability and download speeds while reducing latency. Advanced algorithms like LRU-BaSE employ deep reinforcement learning to dynamically manage eviction decisions, balancing object and byte miss ratios and minimizing access latency and backing traffic [15, 41]. While traditional methods such as Least Recently Used (LRU) and Least Frequently Used (LFU) are popular for their simplicity, they often falter in environments with variable content request sizes and frequencies.

To address these limitations, advanced methods have emerged, including machine learning-based approaches that utilize historical data and real-time analytics to predict content demand. Integrating dynamic CDN slicing with Information-Centric Networking (ICN) slicing optimizes delivery by caching content nearer to users, reducing response times [27]. Additionally, new virtualization architectures for CDNs employ optimization algorithms to reduce migration costs and enhance performance, improving cache replacement strategies [28]. These architectures support dynamic

AI-generated, for reference only.

resource allocation and strategic content placement, crucial for maintaining high cache hit ratios and minimizing latency.

As illustrated in Figure 3, the hierarchical structure of cache replacement policies categorizes them into traditional methods, advanced techniques, and security considerations. Traditional methods include LRU and LFU, while advanced techniques encompass LRU-BaSE, dynamic CDN slicing, and virtualization architectures. Security considerations are addressed by the PDN-AF framework, which assesses the security implications of PDN services and their effect on caching efficiency [31]. Understanding the interplay between security protocols and caching strategies enables CDNs to enhance resilience against threats while maintaining optimal performance.

The evolution of cache replacement policies reflects the increasing complexity of CDN environments, where traditional methods are augmented by advanced techniques to meet modern internet demands. By leveraging machine learning, innovative virtualization architectures like Network Function Virtualization (NFV) and Software Defined Networking (SDN), and robust security frameworks, CDNs can significantly enhance content delivery capabilities. This integration improves the Quality of Experience (QoE) for users and optimizes resource management, addressing challenges like rising video traffic and efficient server capacity utilization [28, 39, 36].

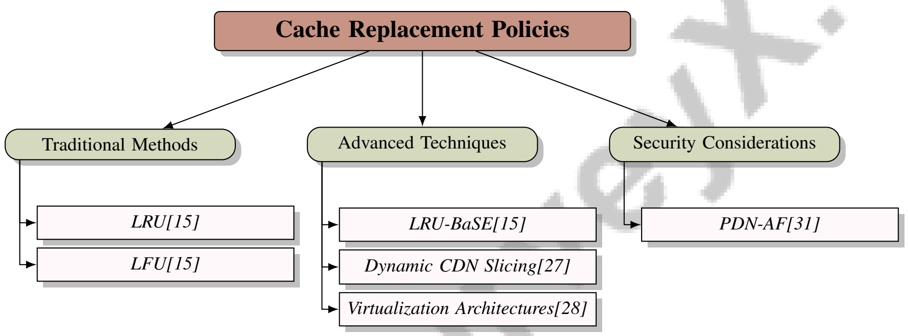

Figure 3: This figure illustrates the hierarchical structure of cache replacement policies, categorizing them into traditional methods, advanced techniques, and security considerations. Traditional methods include LRU and LFU, while advanced techniques encompass LRU-BaSE, dynamic CDN slicing, and virtualization architectures. Security considerations are addressed by the PDN-AF framework.

3.2 Adaptive Caching Algorithms

Adaptive caching algorithms are essential for optimizing content delivery in CDNs, dynamically adjusting cache strategies to accommodate changing network conditions and user demands. By optimizing content replication and delivery, these algorithms ensure users access content from geographically closer servers, reducing latency and network congestion and significantly enhancing the overall Quality of Experience (QoE), especially in bandwidth-intensive applications like video streaming [42, 43, 18, 36].

An innovative adaptive caching approach, Startrail, enhances the InterPlanetary File System (IPFS) by progressively distributing and caching content closer to users, effectively reducing latency and improving throughput [44]. The integration of virtual CDN and ICN slices exemplifies dynamic adaptation to user demands, optimizing caching and content delivery across multiple cloud domains [27].

Recent studies underscore the importance of locality insights from cache traces, improving prediction accuracy and adaptability beyond traditional policies like LRU and LFU by leveraging reuse distance [45]. Additionally, size-aware cache management strategies gather potential victims until their total size is sufficient, then compare frequencies to make more efficient caching decisions [32].

A dataset of 2.0 billion TCP flows monitored over a month provides insights into Anycast CDN (A-CDN) traffic characteristics, emphasizing the importance of adaptive caching in managing diverse traffic patterns [46].

3.3 Machine Learning and Predictive Models in Caching

Integrating machine learning and predictive models into caching strategies marks a significant advancement in optimizing CDNs. These technologies facilitate intelligent decision-making regarding content caching, enhancing both performance and user experience. By leveraging historical data and real-time analytics, machine learning models can forecast content demand and improve cache management. Techniques such as regret minimization and advanced algorithms like matrix factorization optimize cache admission and eviction processes, ensuring popular content is stored closer to users, thus improving access speed and reducing latency. Innovative approaches like LSTM-based recurrent neural networks predict future data accesses with high accuracy, facilitating efficient cache replacements and enhancing CDN performance [47, 15, 43, 18, 45].

The Follow the Perturbed Leader (FTPL) method adapts effectively to user requests without prior knowledge of their distribution, minimizing regret over time [48]. This method enhances content demand prediction, facilitating effective cache placement and replacement strategies. Additionally, the effectiveness of push-pull-based edge caching relies on real-time information about content updates and user requests, allowing for timely cache updates that enhance content freshness [23].

Advanced caching policies such as Adaptive Variant (AV) leverage size-aware strategies to achieve competitive hit-ratios and byte hit-ratios, as demonstrated in experiments with real-world traces from enterprise storage systems and CDN deployments [32]. This integration of size-awareness into traditional caching methods showcases the potential of machine learning to optimize cache content management.

The CDN slicing method utilizes empirical models of video Quality-of-Experience (QoE) to inform resource allocation and management decisions, enhancing caching strategies [25]. Furthermore, Startrail intelligently caches popular content based on access trends, optimizing content delivery in decentralized networks [44].

4 Dynamic Adaptive Streaming over HTTP (DASH)

4.1 Overview of DASH Protocol

Dynamic Adaptive Streaming over HTTP (DASH) is pivotal in adaptive video streaming, allowing real-time video quality adjustments based on network conditions and client capabilities to enhance user Quality of Experience (QoE) [11]. DASH divides video content into smaller, time-synchronized segments encoded at multiple bitrates, facilitating smooth quality transitions according to bandwidth and device performance, thus optimizing delivery [1].

Advanced techniques like the Buffer Occupancy-based Lyapunov Algorithm (BOLA) improve DASH's adaptability by focusing on real-time buffer management rather than unreliable bandwidth predictions, enhancing adaptation accuracy and user satisfaction [49]. Additionally, Blockchain-based architectures ensure secure, efficient content delivery while protecting user identities [33]. The Buffer-based Streaming Control (BSC) method further complements DASH by minimizing interruptions through simultaneous transmission of multiple video quality layers [50]. Comprehensive studies emphasize effective packet loss management for high-quality streaming [29].

4.2 Integration with CDNs

Integrating DASH with Content Delivery Networks (CDNs) significantly enhances video streaming by leveraging both technologies' strengths. This integration addresses latency reduction, network adaptability, and resource optimization through in-network caching, HTTP adaptive streaming, and multipath streaming, ensuring high-quality delivery across diverse networks [1, 51, 10, 52].

Software Defined Networking (SDN) enhances this integration by implementing IP multicast for video content, optimizing delivery paths and reducing traffic load [53]. Combining SDN with DASH allows dynamic flow scheduling, optimizing resources and improving QoE [54]. BOLA offers novel bitrate adaptation based on buffer occupancy, smoothing transitions and minimizing playback interruptions [49]. Joint Collaborative Caching and Processing (JCCP) enables real-time transcoding and caching, enhancing adaptability and reducing latency [28]. CDN and Information-Centric Networking (ICN) slices further improve content delivery by caching at optimal locations [27].

Machine learning techniques like the Dynamic Closed-Loop QoE Optimization (DCL-QOE) method provide real-time video quality optimization based on user feedback and predicted network conditions, ensuring CDNs maintain high QoE [55].

4.3 Challenges in Adaptive Streaming

Adaptive streaming technologies like DASH face challenges affecting video delivery efficiency and quality. Integrating advanced quality metrics such as the Structural Similarity Index (SSIM) into Adaptive Bitrate (ABR) algorithms poses complexity, requiring refinement for optimal use [56]. TCP's limitations, particularly in congestion control, hinder DASH performance, necessitating advanced protocols or optimization techniques [57]. Solutions like the CANE cascade control framework model ABR behavior as a black box, using model predictive control to allocate bandwidth effectively [58].

Buffer management is crucial; inadequate handling can cause buffer starvation, impacting user experience. Recursive buffer management offers flexibility but introduces higher computational complexity, challenging real-time implementation [59]. Future research could explore real-time adjustments to bandwidth allocation models to enhance streaming quality [60].

Error management in CDNs is vital for adaptive streaming. Clustering techniques analyzing CDN error logs provide insights into error patterns, optimizing service quality and mitigating disruptions [61]. Leveraging user viewport similarities enhances caching efficiency, saving bandwidth and improving delivery [62].

Adaptive streaming must also address security threats like random query string DoS attacks. Mitigation strategies, including a proposed gossip protocol, promise to safeguard CDNs from vulnerabilities, ensuring reliable, secure content delivery [63].

5 Quality of Experience (QoE) Optimization

5.1 Adaptive Streaming and QoE Enhancement

Adaptive streaming is pivotal for optimizing Quality of Experience (QoE) by dynamically adjusting video quality according to network conditions and user preferences. Advanced metrics like the Structural Similarity Index (SSIM) enhance video quality assessment beyond traditional bitrate metrics, aligning quality adjustments with user perceptions for improved satisfaction [30]. The Dynamic Adaptive Streaming using Internet Protocol (DAS-IP) method exemplifies this by optimizing streaming decisions based on buffer levels and network conditions, thus enhancing QoE [28].

Integrating Information-Centric Networking (ICN) with Content Delivery Networks (CDNs) significantly boosts content delivery efficiency, mitigates network congestion, and shortens response times, directly influencing QoE [27]. Adaptive caching solutions, which minimize manual intervention, are crucial for optimizing content delivery, as highlighted by application-specific caching techniques [22]. Startrail, an adaptive edge content delivery mechanism, demonstrates a 30

Advancements in optical networking technologies have significantly improved bandwidth capacity and reduced electronic processing requirements, essential for supporting adaptive streaming in high-demand scenarios [26]. Leveraging these technologies allows CDNs to deliver high-quality content efficiently, adapting to the evolving demands of modern digital landscapes.

Adaptive streaming techniques are thus crucial for enhancing QoE through advanced metrics, personalized quality adjustments, and efficient network integration. Ongoing research in CDNs is advancing optimization techniques, such as matrix factorization for cache management, enhancing content popularity predictions and improving cache admission and eviction algorithms. Efforts to interconnect distinct CDNs address challenges related to network congestion and quality of service, while innovative security frameworks utilizing unsupervised anomaly detection aim to protect against cyber threats. New dynamic cache management strategies are proposed to optimize content placement based on real-time demand, ultimately improving user experiences through reduced latency and higher download rates [64, 15, 43, 39, 18].

5.2 Buffer Management and Error Correction

Effective buffer management and error correction are vital for optimizing QoE in multimedia streaming, especially in dynamic and heterogeneous network environments. The variability of network resources and the rise of new multimedia services necessitate context-aware QoE management strategies that adapt to fluctuating conditions [65]. One strategy involves dynamically adjusting video source sending rates based on real-time network conditions, coupled with session admission management to ensure consistent QoE [66], facilitating smooth video delivery despite network variability.

Challenges of fluctuating network performance are amplified in mobile scenarios, such as with Unmanned Aerial Vehicles (UAVs), where high mobility can introduce delays and complications in transmitting high-quality video [67]. Addressing these requires sophisticated buffer management techniques that preemptively allocate resources and predict congestion points, minimizing buffer underruns and playback interruptions.

Error correction mechanisms are crucial for maintaining video quality by compensating for packet loss and transmission errors. By modeling QoE from visual renderings, it becomes possible to capture the visual information influencing user experience, providing a comprehensive understanding of how errors affect perceived quality [68]. This approach enables more effective error correction strategies that prioritize the visual elements most critical to user satisfaction.

5.3 Trade-offs and Metrics in QoE Optimization

Benchmark	Size	Domain	Task Format	Metric
BOLA-BASIC[56]	685,022	Video Streaming	Adaptive Bitrate Streaming	SSIM, Stall Ratio
FB-Akamai[69]	2,000,000	Network Infrastructure	Latency Measurement	Average Access Delay
IACK[70]	600,000	Transport Protocols	Connection Setup Performance Evaluation	Time to First Byte, Probe Timeout
PeerCDN[71]	3,000,000	Video Streaming	Content Delivery Performance Evaluation	QoS metrics, latency
CDN-Convert[72]	1,000,000	Cdn Security	Ddos Attack Simulation	Amplification Factor
DNS-Bench[73]	17,573	Content Delivery Networks	Latency Measurement	DNS Latency, Mapping Latency
DNS-WH[24]	10,000	Web Hosting	Quantitative Analysis	Unreachable, Affected

Table 1: This table provides a comprehensive overview of representative benchmarks used in the evaluation of Quality of Experience (QoE) optimization techniques across various domains. It details the benchmark size, domain, task format, and the specific metrics employed in each case, highlighting the diversity of applications and measurement criteria involved in QoE research.

Optimizing QoE within CDNs involves navigating the complex interplay between user satisfaction and various performance metrics in dynamic network environments. A primary challenge is the inability of existing adaptive streaming algorithms to balance trade-offs between video quality and starvation probability, particularly under fluctuating conditions [74]. This highlights the need for sophisticated QoE optimization techniques that can adapt to real-time network performance changes. Table 1 presents a detailed compilation of benchmarks utilized in the study of trade-offs and metrics in QoE optimization, illustrating the range of domains and performance metrics that are critical to understanding and enhancing user experience in content delivery networks.

Understanding user QoE through visual renderings facilitates the development of more effective optimization strategies [68]. These strategies must consider diverse user expectations regarding video quality, which can vary significantly across contexts and applications.

Integrating adaptive frameworks, such as distributed load management strategies, offers substantial benefits in improving load distribution and mitigating overload situations within CDNs [38]. These approaches emphasize managing performance trade-offs, ensuring efficient resource allocation to maintain high QoE without burdening network infrastructure.

However, current studies often overlook application-specific characteristics, resulting in generalized solutions that may not perform optimally across all scenarios [22]. Tailoring QoE optimization techniques to the specific requirements of different applications is vital for achieving optimal performance and user satisfaction.

6 Edge Server Placement Strategies

6.1 Models and Algorithms for Server Placement

Strategic edge server placement is vital for optimizing Content Delivery Networks (CDNs), particularly for latency-sensitive applications such as live video streaming. Various models and algorithms have been devised to position servers effectively, ensuring optimal content delivery and user satisfaction. The CPVNF method employs a PageRank-based heuristic to select optimal surrogate servers for Virtual Network Function (VNF) deployment, balancing server capacity with outgoing link quality, thereby enhancing CDN adaptability and performance [75]. InviCloak introduces a unilateral deployment strategy that requires no changes to existing CDN infrastructure, facilitating easier adoption [7]. Insights from cloud consolidation studies indicate that strategic proxy deployment can enhance performance without requiring service provider alterations, offering an innovative perspective on server placement [76]. Startrail exemplifies the potential of decentralized models by allowing nodes to cache and deliver popular content, improving access times and reducing load on original content providers [44]. The OPAC model provides a mathematical framework for determining optimal locations for migrating or placing virtual CDN components to meet user quality requirements, aiding strategic decision-making in server placement [28].

6.2 Integration with Emerging Technologies

Incorporating emerging technologies into server placement strategies is revolutionizing the efficiency and adaptability of CDNs, allowing for better accommodation of dynamic network conditions and user demands. Network-Assisted Video Streaming (NAVS) frameworks utilize Software Defined Networking (SDN), Network Functions Virtualization (NFV), and Multi-access Edge Computing (MEC) paradigms to enhance video streaming services by dynamically allocating resources and optimizing server placement to minimize latency and ensure reliability [25]. Future research is expected to advance these frameworks through adaptive learning techniques to refine workload predictions, enhancing server placement and scheduling strategies [77]. Focus areas include developing robust Quality of Experience (QoE) models, exploring edge computing solutions, and fostering collaboration between application-layer algorithms and network infrastructure [10]. Adaptive scheduling algorithms leveraging machine learning techniques show promise in accurately predicting QoE, enabling real-time adjustments in server placement and resource allocation [78]. The effectiveness of these integrated technologies is further demonstrated by methods optimizing server locations based on customer profiles, improving cache hit ratios and reducing access latency through balanced server load and strategic placement [79]. Merger-aware routing techniques maximize traffic merging opportunities, enhancing overall network efficiency [80]. Additionally, integrating Peer-to-Peer (P2P) streaming with CDN capabilities significantly advances scalability and quality adaptation for video delivery, leveraging the decentralized nature of P2P networks to complement CDN infrastructures and manage increased traffic loads efficiently [51].

7 Bandwidth Cost Reduction Techniques

7.1 Pricing Models and Economic Viability

Effective bandwidth cost reduction in Content Delivery Networks (CDNs) relies on strategic pricing models that enhance resource utilization while maintaining high service quality. Key models include flat pricing, usage pricing, and congestion pricing, each offering distinct benefits and challenges [12]. Flat pricing provides simplicity but may lead to inefficiencies by not reflecting actual usage. Usage pricing aligns costs with data consumption, ensuring accurate billing but introducing cost variability. Congestion pricing encourages users to shift demand during peak times, effectively mitigating network congestion.

The Network Bandwidth Allocation (NBA) problem offers a mathematical framework for optimizing resource allocation in cloud computing environments, crucial for effective data transmission scheduling and bandwidth cost reduction [81]. Advanced methods like PCDN+ utilize predictive models to manage bandwidth by anticipating demand fluctuations and adjusting resource allocation, although their efficacy can be challenged by unpredictable workloads [34, 41].

Energy-aware load balancing strategies contribute to economic viability by achieving over 55

7.2 Utilization of Underutilized Resources

Maximizing the use of underutilized resources is essential for reducing expenses in CDNs amid growing demand for high-quality content delivery. Cloud-based solutions enable dynamic resource allocation based on real-time demand, optimizing server utilization and minimizing idle times [41]. This approach enhances resource efficiency and aligns with CDNs' economic goals by reducing operational costs.

Integrating Peer-to-Peer (P2P) streaming within CDN infrastructures is another strategy for optimizing underutilized resources. By distributing content delivery across a network of peers, this method alleviates the load on centralized servers, improving scalability and resource utilization [51]. It also supports effective load balancing, preventing bottlenecks and enhancing network performance.

Energy-aware load balancing further enhances resource utilization by dynamically adjusting server workloads according to energy consumption patterns, achieving significant energy savings while ensuring service availability [82]. This not only reduces operational costs but also supports CDN sustainability and environmental goals.

Advanced caching techniques, particularly those employing machine learning algorithms, optimize underutilized resources by predicting content demand and refining cache placement [34]. These techniques enable informed decisions about content caching, ensuring efficient resource use and optimized content delivery for both cost-effectiveness and performance.

8 Network Optimization for Latency Reduction

8.1 Emerging Technologies in Network Optimization

Emerging technologies such as Software Defined Networking (SDN) and Network Functions Virtualization (NFV) are pivotal in enhancing network performance in Content Delivery Networks (CDNs) by minimizing latency and boosting data transmission efficiency. SDN improves network agility by separating control and data planes, allowing dynamic management of resources and supporting innovative prefetching techniques that enhance data transmission, particularly in scalable video streaming [65, 83]. SDN-based infrastructures offer superior network utilization and reduced flow setup times compared to traditional methods, optimizing overall performance [84].

NFV complements SDN by virtualizing network services on commodity hardware, reducing costs and improving scalability. This virtualization facilitates dynamic resource allocation, effectively meeting varying CDN service demands. The potential of NFV is further enhanced by real-time data utilization, which allows efficient resource allocation in response to client requests without over-provisioning [85]. Future research should focus on refining resource allocation algorithms and exploring additional CDN as a Service (CDNaaS) use cases, integrating machine learning for predictive resource management to enhance service delivery [86].

Advanced algorithms like the Honeybee algorithm balance communication cost and load distribution, reducing bottlenecks and latency in CDNs [87]. The integration of IoT devices and Event-driven Content Optimization (ECoV) for merging redundant video streams illustrates the potential of event-driven processing in optimizing network traffic [80]. Empirical studies in cloud computing scenarios, including content and live video delivery networks, underscore the importance of these technologies in optimizing network performance [81]. Tools such as YTrace aid in performance data capture and analysis across the service and network stack, providing insights into user experience issues and enabling precise optimizations [88].

8.2 Latency Reduction through Strategic Server Placement

Strategic server placement is critical for reducing latency in CDNs by positioning content closer to end-users. Challenges arise from inefficient utilization of existing fiber networks, often affected by ISP routing policies that prioritize cost over latency, complicating latency measurement and validation across diverse networks [89]. Innovative placement strategies are required to enhance network efficiency and minimize delays.

The PACK algorithm provides a scalable solution for edge server placement, accommodating both high and low-capacity servers, addressing deployment complexities, and ensuring optimal resource allocation to satisfy user demands [90]. By strategically positioning servers at the network edge, CDNs can significantly reduce data travel distance, thereby decreasing latency and improving user experience.

A framework for analyzing CDN performance using raw log files allows detailed categorization of data by user access patterns and service types, offering insights for system optimization and guiding informed decisions regarding server placement and resource allocation [18]. Understanding user interactions and service usage patterns enables effective server positioning, further reducing latency.

Distributing the load across multiple peers and paths can alleviate congestion and improve streaming quality. Recent studies highlight the importance of decentralized architectures to balance network load, minimizing bottlenecks and enhancing content delivery [51]. Implementing such distributed strategies can strengthen CDNs' resilience against traffic spikes, ensuring consistent performance across varied network conditions.

9 Conclusion

Content Delivery Networks (CDNs) play a crucial role in enhancing content distribution within today's internet infrastructure. By employing advanced frameworks, such as closed-loop Quality of Experience (QoE) optimization, CDNs have significantly improved video delivery quality by dynamically adjusting to user feedback and fluctuating network conditions. The strategic implementation of IPv6 in densely populated regions offers valuable insights for further refinement and optimization. Nonetheless, the growing prevalence of encrypted traffic and the imperative for enhanced user privacy pose substantial challenges to the sustainability of CDNs.

Innovations like InviCloak present promising solutions for bolstering privacy while maintaining CDN performance. However, future research should focus on minimizing computational overhead and effectively deploying integrity verifiers. Additionally, recent findings underscore the security vulnerabilities in Peer-assisted Delivery Networks (PDNs), highlighting the need for comprehensive mitigation strategies and increased user awareness. Expanding the geographical range of web performance metrics and integrating machine learning techniques could greatly enhance predictive capabilities for loading times and performance analyses.

The centralization of DNS and web hosting services has significant implications for internet resilience and security, necessitating strategic resource management and improved system robustness. Moreover, integrating Information-Centric Networking (ICN) with CDN architectures has demonstrated potential in enhancing content delivery efficiency, alleviating network congestion, and reducing response times. Future research endeavors could focus on augmenting system scalability and optimizing resource management strategies to achieve superior performance.

Ongoing innovation and research in CDN technologies are vital to meet the evolving demands of modern internet infrastructure. Future directions should include refining collaborative frameworks, exploring automated integration mechanisms, enhancing resource allocation algorithms with predictive models, and developing self-adaptive caching techniques that leverage application-specific insights. By leveraging emerging technologies and strategic approaches, CDNs can substantially improve content delivery efficiency, scalability, and user satisfaction, thus setting the stage for future advancements in content distribution networks.

References

[1] Vitalii Poliakov. *Combining in-network caching, HTTP adaptive streaming and multipath to improve video quality of experience*. PhD thesis, Université Côte d'Azur, 2018.

[2] Mohammadhassan Safavi and Saeed Bastani. Invited abstract: A simulation package for energy consumption of content delivery networks (cdns), 2015.

[3] Mojgan Ghasemi, Partha Kanuparthy, Ahmed Mansy, Theophilus Benson, and Jennifer Rexford. Performance characterization of a commercial video streaming service, 2016.

[4] Rahul Singh and P. R. Kumar. Optimal decentralized dynamic policies for video streaming over wireless channels, 2019.

[5] Louiza Yala. *Content delivery networks as a service (CDNaaS)*. PhD thesis, Université de Rennes, 2018.

[6] Javier Cabrera-Arteaga, Pierre Laperdrix, Martin Monperrus, and Benoit Baudry. Multi-variant execution at the edge, 2022.

[7] Shihan Lin, Rui Xin, Aayush Goel, and Xiaowei Yang. Invicloak: An end-to-end approach to privacy and performance in web content distribution, 2022.

[8] Stefano Iellamo, Guenter Klas, and Kevin Smith. New generation value networks for content delivery, 2016.

[9] Antoine Saverimoutou, Bertrand Mathieu, and Sandrine Vaton. Web view: A measurement platform for depicting web browsing performance and delivery, 2020.

[10] Leonardo Peroni and Sergey Gorinsky. An end-to-end pipeline perspective on video streaming in best-effort networks: A survey and tutorial, 2025.

[11] Miran Taha Abdullah Abdullah, Jaime Lloret, Alejandro Cánovas Solbes, and Laura García-García. Survey of transportation of adaptive multimedia streaming service in internet. *Network Protocols and Algorithms*, 9(1-2):85–125, 2017.

[12] Huan He, Ke Xu, and Ying Liu. Internet resource pricing models, mechanisms, and methods, 2011.

[13] Jing Zhang. A literature survey of cooperative caching in content distribution networks, 2012.

[14] Reza Farahani. Network-assisted delivery of adaptive video streaming services through cdn, sdn, and mec, 2024.

[15] Srujan Teja Thomdapu, Palash Katiyar, and Ketan Rajawat. Dynamic cache management in content delivery networks, 2020.

[16] Al-Mukaddim Khan Pathan, Rajkumar Buyya, James Broberg, and Kris Bubendorfer. Economy-based content replication for peering content delivery networks, 2006.

[17] Mario E. Consuegra, Wendy A. Martinez, Giri Narasimhan, Raju Rangaswami, Leo Shao, and Giuseppe Vietri. Analyzing adaptive cache replacement strategies, 2017.

[18] 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.

[19] Li Ye, Hong Xie, John C. S. Lui, and Kenneth L. Calvert. Quantifying deployability evolvability of future internet architectures via economic models, 2020.

[20] Yaohua Sun, Mugen Peng, and Shiwen Mao. A game-theoretic approach to cache and radio resource management in fog radio access networks. *IEEE Transactions on Vehicular Technology*, 68(10):10145–10159, 2019.

[21] Gaël Guennebaud, Aurélie Bugeau, and Antoine Dudouit. Assessing vod pressure on network power consumption, 2023.

[22] Jhonny Mertz and Ingrid Nunes. Understanding application-level caching in web applications: a comprehensive introduction and survey of state-of-the-art approaches. *ACM Computing Surveys (CSUR)*, 50(6):1–34, 2017.

[23] Bahman Abolhassani, John Tadrous, Atilla Eryilmaz, and Serdar Yüksel. Optimal push and pull-based edge caching for dynamic content, 2024.

[24] Synthia Wang, Kyle MacMillan, Brennan Schaffner, Nick Feamster, and Marshini Chetty. Measuring the consolidation of dns and web hosting providers, 2024.

[25] T. Taleb, P. A. Frangoudis, I. Benkacem, and A. Ksentini. Cdn slicing over a multi-domain edge cloud, 2022.

[26] Oscar Gonzalez de Dios, Ramon Casellas, Filippo Cugini, and Jose Alberto Hernandez. Beyond 5g domainless network operation enabled by multiband: Toward optical continuum architectures, 2023.

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

[28] Hatem Ibn Khedher. *Optimization and virtualization techniques adapted to networking*. PhD thesis, Institut National des Télécommunications, 2018.

[29] Konstantin Andreev, Bruce M. Maggs, Adam Meyerson, Jevan Saks, and Ramesh K. Sitaraman. Algorithms for constructing overlay networks for live streaming, 2011.

[30] Danilo Giordano, Stefano Traverso, Luigi Grimaudo, Marco Mellia, Elena Baralis, Alok Tongaonkar, and Sabyasachi Saha. Youlighter: An unsupervised methodology to unveil youtube cdn changes, 2015.

[31] Siyuan Tang, Eihal Alowaisheq, Xianghang Mi, Yi Chen, XiaoFeng Wang, and Yanzhi Dou. Stealthy peers: Understanding security risks of webrtc-based peer-assisted video streaming, 2022.

[32] Gil Einziger, Ohad Eytan, Roy Friedman, and Benjamin Manes. Lightweight robust size aware cache management, 2021.

[33] Thang X. Vu, Symeon Chatzinotas, and Bjorn Ottersten. Blockchain-based content delivery networks: Content transparency meets user privacy, 2019.

[34] Rui-Xiao Zhang, Haiping Wang, Shu Shi, Xiaofei Pang, Yajie Peng, Zhichen Xue, and Jiangchuan Liu. Enhancing resource management of the world's largest {PCDN} system for {On-Demand} video streaming. In *2024 USENIX Annual Technical Conference (USENIX ATC 24)*, pages 951–965, 2024.

[35] Gonca Gursun. Routing-aware partitioning of the internet address space for server ranking in cdns, 2018.

[36] Gang Peng. Cdn: Content distribution network, 2004.

[37] Paolo Medagliani, Stefano Paris, Jérémie Leguay, Lorenzo Maggi, Xue Chuangsong, and Haojun Zhou. Overlay routing for fast video transfers in cdn, 2017.

[38] Abhishek Sinha, Pradeepkumar Mani, Jie Liu, Ashley Flavel, and David A. Maltz. Distributed load management in anycast-based cdns, 2015.

[39] Rajkumar Buyya, Al-Mukaddim Khan Pathan, James Broberg, and Zahir Tari. A case for peering of content delivery networks, 2006.

[40] Anne Edmundson, Paul Schmitt, Nick Feamster, and Jennifer Rexford. Ocdn: Oblivious content distribution networks, 2017.

[41] Peng Wang and Yu Liu. Optimizing replacement policies for content delivery network caching: Beyond belady to attain a seemingly unattainable byte miss ratio, 2022.

[42] Peter Hillmann, Tobias Uhlig, Gabi Dreo Rodosek, and Oliver Rose. Simulation and optimization of content delivery networks considering user profiles and preferences of internet service providers, 2020.

[43] Adolf Kamuzora, Wadie Skaf, Ermiyas Birihanu, Jiyan Mahmud, Péter Kiss, Tamás Jursonovics, Peter Pogrzeba, Imre Lendák, and Tomáš Horváth. Matrix factorization for cache optimization in content delivery networks (cdn), 2022.

[44] João Tiago, David Dias, and Luís Veiga. Adaptive edge content delivery networks for web-scale file systems, 2022.

[45] Pengcheng Li and Yongbin Gu. Learning forward reuse distance, 2020.

[46] Danilo Cicalese, Danilo Giordano, Alessandro Finamore, Marco Mellia, Maurizio Munafò, Dario Rossi, and Diana Joumblatt. A first look at anycast cdn traffic, 2021.

[47] Giuseppe Vietri, Liana V Rodriguez, Wendy A Martinez, Steven Lyons, Jason Liu, Raju Rangaswami, Ming Zhao, and Giri Narasimhan. Driving cache replacement with {ML-based}{LeCaR}. In *10th USENIX Workshop on Hot Topics in Storage and File Systems (HotStorage 18)*, 2018.

[48] Rajarshi Bhattacharjee, Subhankar Banerjee, and Abhishek Sinha. Fundamental limits of online network-caching, 2020.

[49] Kevin Spiteri, Rahul Urgaonkar, and Ramesh K. Sitaraman. Bola: Near-optimal bitrate adaptation for online videos, 2020.

[50] Zakaria Ye. *Performance analysis of http adaptive video streaming services in mobile networks*. PhD thesis, Université d'Avignon, 2017.

[51] Mathias Lacaud. *Towards pragmatic solutions to improve the quality of video streaming in current and future networks*. PhD thesis, Université de Bordeaux, 2020.

[52] Ge Ma, Zhi Wang, Miao Zhang, Jiahui Ye, Minghua Chen, and Wenwu Zhu. Understanding performance of edge content caching for mobile video streaming, 2017.

[53] Kyoomars Alizadeh Noghani and M. Oguz Sunay. Streaming multicast video over software-defined networks, 2015.

[54] Ming Ma, Zhi Wang, Yankai Zhang, and Lifeng Sun. Dynamic flow scheduling strategy in multihoming video cdns, 2016.

[55] Imen Triki, Quanyan Zhu, Rachid Elazouzi, Majed Haddad, and Zhiheng Xu. Learning from experience: A dynamic closed-loop qoe optimization for video adaptation and delivery, 2017.

[56] Emily Marx, Francis Y. Yan, and Keith Winstein. Implementing bola-basic on puffer: Lessons for the use of ssim in abr logic, 2020.

[57] Wen Hu, Zhi Wang, and Lifeng Sun. A measurement study of tcp performance for chunk delivery in dash, 2016.

[58] Mehdi Hosseinzadeh, Karthick Shankar, Maria Apostolaki, Jay Ramachandran, Steven Adams, Vyas Sekar, and Bruno Sinopoli. Cane: A cascade-control approach for network-assisted video qoe management, 2023.

[59] Yuedong Xu, Eitan Altman, Rachid El-Azouzi, Majed Haddad, Salaheddine Elayoubi, and Tania Jimenez. Analysis of buffer starvation with application to objective qoe optimization of streaming services, 2013.

[60] Iskandar Aripov. The statistical analysis of the live tv bit rate, 2019.

[61] Ermiyas Birihanu, Jiyan Mahmud, Péter Kiss, Adolf Kamuzora, Wadie Skaf, Tomáš Horváth, Tamás Jursonovics, Peter Pogrzeba, and Imre Lendák. Client error clustering approaches in content delivery networks (cdn), 2022.

[62] Niklas Carlsson and Derek Eager. Cross-user similarities in viewing behavior for 360° video and caching implications, 2021.

[63] Stefano Ferretti and Vittorio Ghini. Mitigation of random query string dos via gossip, 2012.

[64] Li Yang, Abdallah Moubayed, Abdallah Shami, Parisa Heidari, Amine Boukhtouta, Adel Larabi, Richard Brunner, Stere Preda, and Daniel Migault. Multi-perspective content delivery networks security framework using optimized unsupervised anomaly detection, 2021.

[65] 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.

[66] Qahhar Muhammad Qadir, Alexander A. Kist, and Zhongwei Zhang. A quality of experience-aware cross-layer architecture for optimizing video streaming services, 2020.

[67] Liyana Adilla binti Burhanuddin, Xiaonan Liu, Yansha Deng, Ursula Challita, and Andras Zahemszky. Qoe optimization for live video streaming in uav-to-uav communications via deep reinforcement learning, 2021.

[68] Junchen Jiang and Siddhartha Sen. A new abstraction for internet qoe optimization, 2020.

[69] Reza Farahbakhsh, Angel Cuevas, Antonio M. Ortiz, Xiao Han, and Noel Crespi. How far is facebook from me? facebook network infrastructure analysis, 2017.

[70] Jonas Mücke, Marcin Nawrocki, Raphael Hiesgen, Thomas C. Schmidt, and Matthias Wählisch. Reacked quicer: Measuring the performance of instant acknowledgments in quic handshakes, 2024.

[71] Ming Ma, Zhi Wang, Ke Su, and Lifeng Sun. Understanding the smartrouter-based peer cdn for video streaming, 2016.

[72] Ziyu Lin, Zhiwei Lin, Ximeng Liu, Zuobing Ying, and Cheng Chen. Unveiling the bandwidth nightmare: Cdn compression format conversion attacks, 2024.

[73] Rami Al-Dalky and Michael Rabinovich. Revisiting comparative performance of dns resolvers in the ipv6 and ecs era, 2020.

[74] Rahul Singh and P. R. Kumar. Dynamic adaptive streaming using index-based learning algorithms, 2016.

[75] Mouhamad Dieye, Shohreh Ahvar, Jagruti Sahoo, Ehsan Ahvar, Roch Glitho, Halima Elbiaze, and Noel Crespi. Cpvnf:cost-efficient proactive vnf placement and chaining for value-added services in content delivery networks, 2018.

[76] Debopam Bhattacherjee, Muhammad Tirmazi, and Ankit Singla. Measuring and exploiting the cloud consolidation of the web, 2019.

[77] Shizhen Zhao, Xiao Zhang, Peirui Cao, and Xinbing Wang. Design of robust and efficient edge server placement and server scheduling policies: Extended version, 2021.

[78] Ivo Sousa, Maria Paula Queluz, and António Rodrigues. A survey on qoe-oriented wireless resources scheduling, 2020.

[79] Peter Hillmann, Tobias Uhlig, Gabi Dreo Rodosek, and Oliver Rose. Modeling the location selection of mirror servers in content delivery networks, 2020.

[80] Kutalmis Akpinar. Utilizing edge in iot and video streaming applications to reduce bottlenecks in internet traffic. 2019.

[81] Changpeng Yang, Jintao You, Xiaoming Yuan, and Pengxiang Zhao. Network bandwidth allocation problem for cloud computing, 2022.

[82] Vimal Mathew, Ramesh K. Sitaraman, and Prashant Shenoy. Energy-aware load balancing in content delivery networks, 2011.

[83] Jordi Ortiz and Antonio Skarmeta. Sdn enabled information centric networking (icn) as a service prefetching mechanism for hypertext transfer protocol (http) based services. the scalable video streaming case, 2020.

[84] Dirk Trossen, Sebastian Robitzsch, Scott Hergenhan, Janne Riihijarvi, Martin Reed, and Mays Al-Naday. Service-based routing at the edge, 2019.

[85] Elif Ak, Taner Ozdas, Serkan Sevim, and Berk Canberk. Wae: Workload automation engine for cdn-specialized container orchestration, 2020.

[86] Abdelhak Bentaleb, May Lim, Mehmet N Akcay, Ali C Begen, Sarra Hammoudi, and Roger Zimmermann. Toward one-second latency: Evolution of live media streaming. *arXiv preprint arXiv:2310.03256*, 2023.

[87] Hamid Ghasemi, Mahdi Jafari Siavoshani, and Saeed Hadadan. A novel communication cost aware load balancing in content delivery networks using honeybee algorithm, 2019.

[88] Partha Kanuparthy, Yuchen Dai, Sudhir Pathak, Sambit Samal, Theophilus Benson, Mojgan Ghasemi, and P. P. S. Narayan. Ytrace: End-to-end performance diagnosis in large cloud and content providers, 2016.

[89] Ilker Nadi Bozkurt, Waqar Aqeel, Debopam Bhattacherjee, Balakrishnan Chandrasekaran, Philip Brighten Godfrey, Gregory Laughlin, Bruce M. Maggs, and Ankit Singla. Dissecting latency in the internet's fiber infrastructure, 2018.

[90] 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.

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.

AI-generated, for reference only.