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 integral to modern internet infrastructure, addressing the substantial rise in internet traffic, particularly from video content, which was projected to account for 81% of all internet traffic by 2021 [1]. This increase presents challenges for Internet Service Providers (ISPs) and adversely affects the Quality of Experience (QoE) for users. CDNs mitigate network congestion and enhance QoE by distributing content closer to end-users, thereby reducing latency and improving load times [2].

The rapid growth of video streaming and mobile data, fueled by high-resolution content and the expansion of Internet of Things (IoT) applications, underscores the critical role of CDNs [3]. Mobile video traffic constituted 55% of total mobile data traffic in 2015, projected to rise to 75% by 2020, making CDNs essential in this landscape [4].

To adapt to evolving demands, CDNs have adopted advanced technologies such as Network Functions Virtualization (NFV) and Software Defined Networking (SDN), facilitating dynamic resource allo-

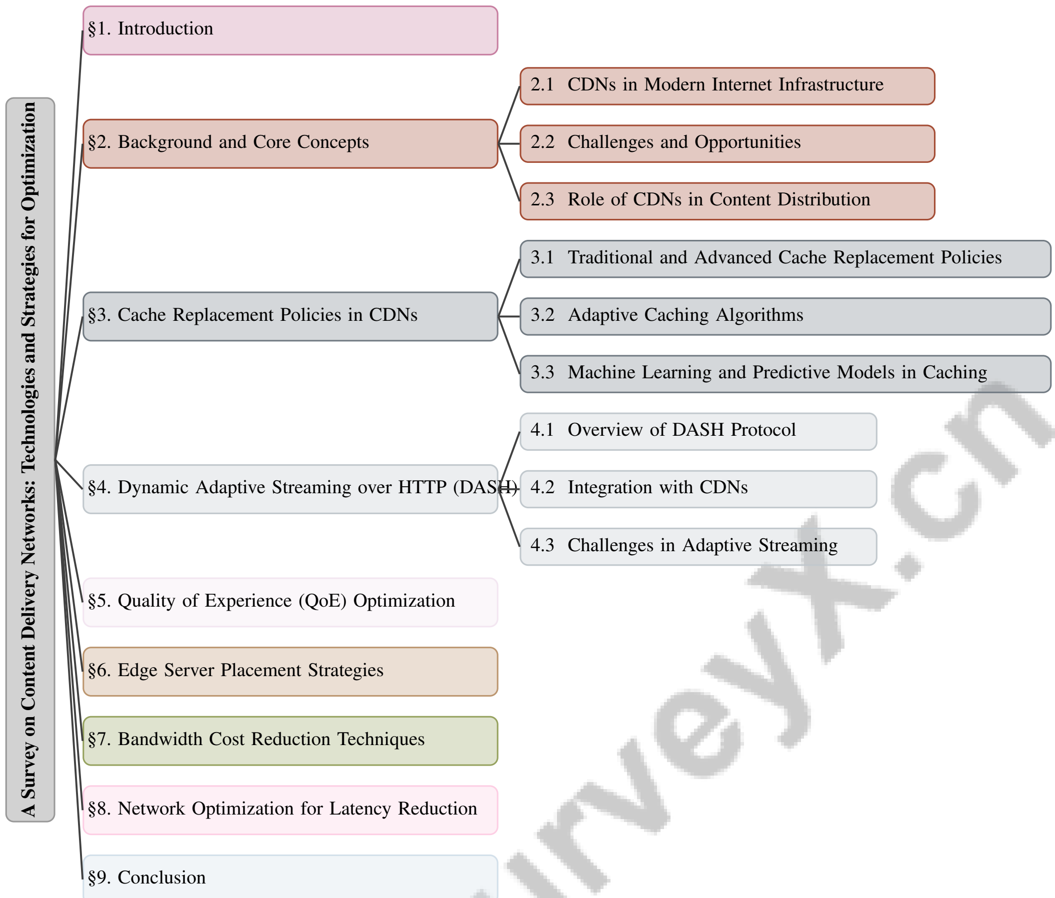

Figure 1: chapter structure

cation and enhanced network management [5]. These innovations enable CDNs to provide flexible, cost-effective services that address inefficiencies in traditional content delivery methods. Additionally, edge-cloud computing, which shifts computations from centralized servers to edge nodes, improves content delivery efficiency and responsiveness [6]. Platforms like WebAssembly further optimize service distribution and deployment.

Energy efficiency has emerged as a vital design consideration for CDNs, ensuring sustainability in future internet infrastructure [2]. Solutions such as InviCloak enhance CDN performance while protecting user privacy by preventing sensitive key sharing with providers [7].

The economic and technical challenges in the content delivery market necessitate strategic methodologies for CDN deployment [8]. As web content becomes increasingly dynamic, optimizing web page loading times is crucial for user experience, emphasizing the need for efficient CDNs [9].

1.2 Structure of the Survey

This survey provides a comprehensive analysis of the technologies and strategies employed in CDNs to optimize content distribution and enhance user experience. The initial sections introduce CDNs and their integration into modern internet infrastructure, followed by an exploration of core concepts, challenges, and opportunities faced by CDNs in content distribution.

Subsequent sections focus on specific technological strategies essential for CDN optimization. The discussion begins with cache replacement policies, analyzing traditional methods like Least Recently Used (LRU) and Least Frequently Used (LFU), alongside advanced machine learning-based approaches for their performance impacts. The survey then transitions to Dynamic Adaptive Streaming over HTTP (DASH), emphasizing its significance in adaptive video streaming, particularly

AI-generated, for reference only.

for long-form two-dimensional videos utilizing HTTP adaptive streaming (HAS) with client-side adaptive bitrate (ABR) algorithms [10].

Quality of Experience (QoE) optimization strategies are examined, highlighting adaptive streaming techniques that enhance user satisfaction. The analysis of edge server placement strategies is crucial for optimizing application performance by minimizing latency and improving content accessibility in edge computing contexts, considering factors such as server capacity, deployment budget, and network topology. Effective placement can significantly reduce data transfer loads, particularly in mobile video streaming, where caching strategies at Wi-Fi access points and cellular base stations are vital for meeting user demands. Innovative algorithms like PACK facilitate strategic edge server positioning to balance workloads and improve reliability, while dynamic cache management techniques ensure high content availability and efficient resource utilization across CDNs [11, 12, 13, 14]. Additionally, bandwidth cost reduction techniques are evaluated, focusing on pricing models and methods to utilize underutilized resources effectively.

The final sections address network optimization for latency reduction, exploring the roles of emerging technologies such as SDN and NFV in enhancing network performance. The survey concludes by summarizing key findings and insights, emphasizing the need to improve CDN efficiency to support future streaming scenarios involving millions of users [15], 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 the modern internet, facilitating the delivery of high-resolution video-on-demand (VoD) and live streaming services. These applications demand low latency and high bandwidth, necessitating robust CDN infrastructures to manage congestion and improve scalability by deploying cache servers near end-users [12]. The proprietary nature of many CDNs can lead to inefficiencies, particularly in cache management, where optimization of data eviction is crucial for enhancing cache hits [16]. The effectiveness of CDNs, especially in video streaming, is influenced by their infrastructure and user access patterns [17].

Economic considerations are vital for CDN integration, notably in adopting new protocols by Internet Service Providers (ISPs) [18]. Large platforms like Facebook and Netflix demonstrate the necessity of efficient CDN strategies to minimize delays and maintain high user experience standards. Facebook's infrastructure caters to over a billion users, while Netflix's Open Connect CDN leverages global Internet Exchange Points (IXPs) and ISPs for optimal content distribution. Emerging technologies, such as Fog Radio Access Networks (F-RANs), offer promising solutions for the 5G era by utilizing fog access points (FAPs) to complement CDNs in delivering high-quality content with reduced latency [19]. Additionally, video streaming services' energy consumption and carbon footprint are critical considerations, with methodologies assessing network infrastructure impact at a territorial scale [20].

CDNs continue to evolve, integrating static and adaptive caching strategies to enhance performance [21]. They address challenges related to dynamic content, content freshness, and aging costs [22]. The consolidation of DNS and web hosting services underscores the importance of understanding vulnerabilities and dependencies within internet services, providing a framework for evaluating CDN architecture robustness [23].

2.2 Challenges and Opportunities

CDNs face significant challenges impacting their performance, scalability, and security, while also presenting opportunities for optimization and innovation. A key challenge is the inability of current mobile networks to dynamically scale with fluctuating traffic, degrading video delivery quality [24]. Hierarchical telco architectures exacerbate this issue by inefficiently supporting increasing data traffic [25]. Effective application-level caching remains crucial, requiring informed decisions about content caching and timing [21]. Traditional caching methods often fail to adapt to varying user demands, resulting in inefficiencies [26]. Additionally, existing CDN architectures struggle with rising video content demand, leading to increased latency and resource wastage [27].

AI-generated, for reference only.

Overlay multicast networks present challenges, with server and network bottlenecks, including limited capacity and unpredictable packet loss, posing significant obstacles [28]. ISPs also face difficulties monitoring evolving CDNs like YouTube, affecting the Quality of Experience (QoE) for users [29]. Security concerns are paramount, particularly in Peer-assisted Delivery Networks (PDNs) used in video streaming, which introduce unexamined risks [30].

Despite these challenges, significant opportunities exist for CDN innovation. Advances in caching strategies, such as size-aware policies, can enhance performance in modern databases [31]. Blockchain-based solutions offer promising avenues for addressing inefficiencies and privacy concerns, potentially reducing subscriptions and operational costs [32]. Optimizing edge resource utilization to prevent node overloading while ensuring balanced distribution remains a critical opportunity for enhancing efficiency [33].

2.3 Role of CDNs in Content Distribution

CDNs are crucial for efficient internet content distribution, particularly with the rising demand for high-quality video and evolving mobile networks. By positioning servers closer to end-users, CDNs minimize latency and improve the Quality of Experience (QoE), crucial for accommodating increased mobile video traffic driven by 5G and IoT advancements. CDNs replicate content across multiple servers and use sophisticated algorithms to select the optimal server based on real-time latency measurements, ensuring swift video content delivery amid growing internet utilization [34, 35, 36].

Integrating advanced technologies like Information-Centric Networking (ICN) with CDNs addresses scalability, reliability, and quality of service challenges amid increasing mobile video traffic [26]. This integration optimizes resource management and service delivery through virtualization [27]. Anycast addressing in CDNs facilitates distributed load management across multiple proxy nodes, though it presents challenges in avoiding overload situations [37].

In video streaming, PDNs utilize peer-to-peer connections for content delivery, introducing potential security vulnerabilities that must be managed [30]. Modeling energy consumption across various CDN elements, including user devices and network components, further complicates content delivery optimization [2].

CDNs are vital in the contemporary digital landscape, using sophisticated technologies and collaborative frameworks to enhance content delivery efficiency, alleviate network congestion, and improve user experiences. By strategically positioning servers and data centers globally, CDNs replicate and distribute content closer to users, addressing challenges posed by bandwidth-intensive applications like video streaming. The evolution of CDNs has led to multi-CDN strategies, enabling content providers to optimize performance and reduce costs by dynamically switching between networks based on geographic needs and service quality [34, 38, 11, 17, 39]. Adapting to 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 optimization of cache replacement policies within Content Delivery Networks (CDNs) is crucial for enhancing content delivery efficiency and resource utilization, directly impacting user experience by reducing latency and improving download speeds. Recent advancements, such as the LRU-BaSE algorithm, have significantly improved cache performance by dynamically optimizing object eviction strategies based on real-time monitoring of object miss ratios (OMR) and byte miss ratios (BMR). This approach decreases access latency and backing-to-OS traffic by over 30

To illustrate the complexities of these strategies, Figure 2 presents a hierarchical structure of cache replacement policies in CDNs, categorizing traditional and advanced strategies, adaptive caching algorithms, and the integration of machine learning and predictive models. This figure highlights the evolution from basic policies like Least Recently Used (LRU) and Least Frequently Used (LFU) to sophisticated methods such as LRU-BaSE and machine learning-based approaches, emphasizing their applications in optimizing content delivery and enhancing user experience. A thorough examination of cache management strategies, including both traditional and advanced policies, is necessary to comprehend the innovations addressing the evolving challenges in this domain.

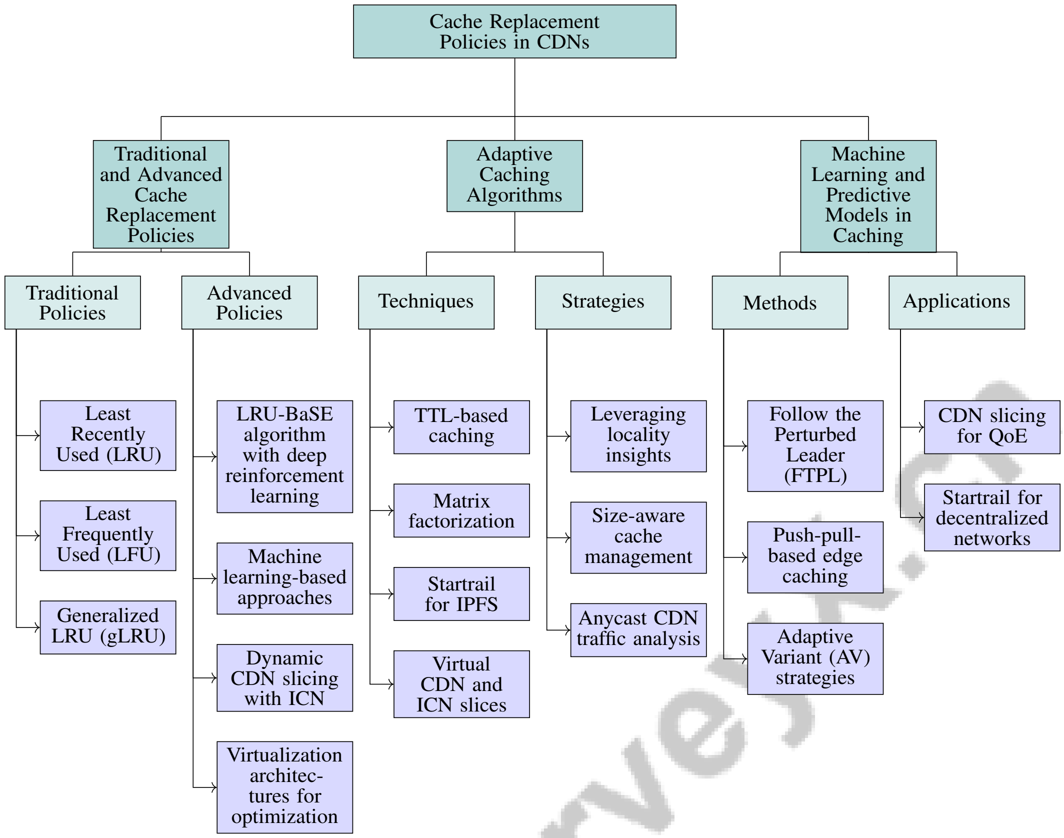

Figure 2: This figure illustrates the hierarchical structure of cache replacement policies in CDNs, categorizing traditional and advanced strategies, adaptive caching algorithms, and the integration of machine learning and predictive models. It highlights the evolution from basic policies like LRU and LFU to sophisticated methods such as LRU-BaSE and machine learning-based approaches, emphasizing their applications in optimizing content delivery and enhancing user experience.

3.1 Traditional and Advanced Cache Replacement Policies

Cache replacement policies are pivotal in optimizing CDN efficiency and performance by determining which content should be retained or evicted from the cache. Traditional policies like Least Recently Used (LRU) and Least Frequently Used (LFU) are favored for their straightforward implementation and effectiveness in predictable access patterns. However, LRU's performance declines with asymmetric file sizes, leading to adaptations like generalized LRU (gLRU), which optimizes performance by caching file chunks instead of entire files. Traditional policies may falter in scenarios requiring content freshness, prompting modified policies that prioritize cached content based on user requests. Recent advancements, such as the LRU-BaSE algorithm, employ deep reinforcement learning to enhance cache performance by dynamically managing eviction strategies, improving byte miss ratios while maintaining object miss ratios. These innovations reflect the ongoing evolution of caching strategies to meet the diverse demands of modern applications, particularly in video streaming and real-time content delivery [40, 17, 41, 42].

To overcome the limitations of traditional caching strategies, advanced methods have emerged, including machine learning-based approaches that utilize historical data and real-time analytics to predict content demand. For instance, integrating dynamic CDN slicing with Information-Centric Networking (ICN) slicing optimizes content delivery by caching content closer to users, thereby reducing response times [26]. Additionally, new virtualization architectures for CDNs incorporate optimization algorithms to minimize migration costs and enhance performance, supporting effective cache replacement strategies [27]. Security considerations are integral to cache replacement policies, as evidenced by the Peer-assisted Delivery Network (PDN) analysis framework (PDN-AF), which evaluates the security implications of PDN services and their impact on caching efficiency [30].

The evolution of cache replacement policies in CDNs underscores the increasing complexity of modern internet infrastructure, with traditional methods enhanced by innovative techniques such as dynamic optimization algorithms and deep reinforcement learning models. As illustrated in Figure 3, this figure highlights the hierarchical categorization of cache replacement policies in CDNs, distinguishing between traditional and advanced methods, and emphasizing virtualization approaches to optimize performance and security. These advancements effectively manage time-varying content popularity, improve byte miss ratios, and ensure higher content availability and reduced latency, ultimately enhancing user experience in bandwidth-intensive services like video streaming [34, 43, 40, 17, 12]. By integrating machine learning strategies, virtualization architectures, and security frameworks, CDNs can improve content delivery efficiency, ensuring high Quality of Experience (QoE) and optimal resource management across diverse network conditions.

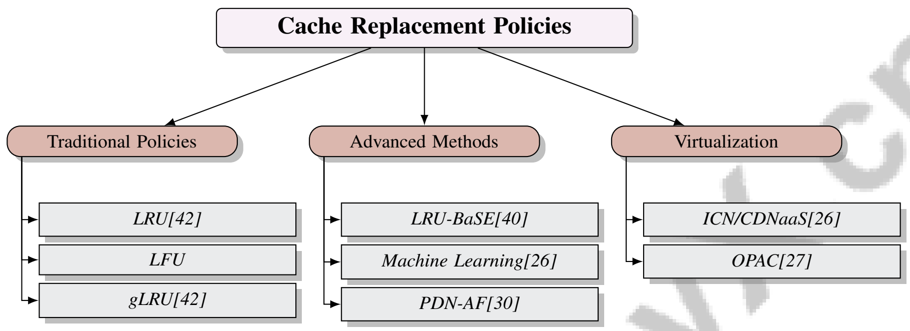

Figure 3: This figure illustrates the hierarchical categorization of cache replacement policies in CDNs, distinguishing between traditional and advanced methods, and highlighting virtualization approaches to optimize performance and security.

3.2 Adaptive Caching Algorithms

Adaptive caching algorithms are crucial for enhancing content delivery in CDNs by dynamically modifying cache strategies to align with fluctuating network conditions and varying user demands. These algorithms address challenges posed by the non-stationary and bursty nature of real-world content requests, ensuring high-quality user experiences. Techniques such as TTL-based caching and matrix factorization optimize cache content selection and management, enabling CDNs to efficiently serve popular content while minimizing latency and maximizing hit rates. By leveraging real-time data and predictive models, adaptive algorithms significantly improve content availability and download speeds, leading to a more responsive content delivery system [44, 12, 45, 43].

An innovative approach like Startrail enhances the InterPlanetary File System (IPFS) by progressively distributing and caching content closer to users, significantly reducing latency and improving throughput [46]. The integration of virtual CDN and ICN slices exemplifies dynamic adaptation to user demands, optimizing caching and content delivery across multiple cloud domains [26].

Recent studies highlight the method of leveraging locality insights from cache traces to improve prediction accuracy and adaptability compared to traditional policies like LRU and LFU, using reuse distance for enhanced cache management [47]. Additionally, size-aware cache management strategies gather potential victims until their total size is sufficient, then compare candidate frequencies to the total frequency of the gathered victims, leading to more efficient caching decisions [31].

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 [48]. Understanding these characteristics enables CDNs to implement effective caching strategies aligned with observed patterns.

3.3 Machine Learning and Predictive Models in Caching

Integrating machine learning and predictive models into caching strategies represents a transformative advancement in optimizing CDNs. These technologies enable intelligent decision-making for content placement and retrieval, optimizing mobile video streaming delivery through edge networks like Wi-Fi access points and cellular base stations. This strategic approach boosts performance metrics, such as download speeds and latency, while enhancing user experience by ensuring popular content is readily available closer to users, alleviating network congestion [43, 12, 17, 14]. By analyzing historical data and real-time analytics, machine learning models can predict content demand and optimize cache management.

An innovative method, Follow the Perturbed Leader (FTPL), adapts to user requests without prior knowledge of their distribution, minimizing regret over time [49]. This method enhances content demand prediction, facilitating effective cache placement and replacement strategies. Additionally, push-pull-based edge caching utilizes real-time information about content updates and user requests, allowing timely cache updates that enhance content freshness [22].

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 [31]. This integration exemplifies the potential of machine learning to adaptively manage cache content, ensuring optimal performance.

The CDN slicing method employs empirical models of video Quality-of-Experience (QoE) to drive resource allocation and management decisions, enhancing caching strategies [24]. This approach underscores the significance of real-time data in refining caching strategies through predictive modeling. Furthermore, Startrail, an adaptive caching extension for IPFS, intelligently caches popular content based on access trends, optimizing content delivery in decentralized networks [46].

4 Dynamic Adaptive Streaming over HTTP (DASH)

4.1 Overview of DASH Protocol

Dynamic Adaptive Streaming over HTTP (DASH) is pivotal for adaptive video streaming, allowing real-time quality adjustments based on network conditions and client capabilities to optimize user Quality of Experience (QoE) [50]. It segments videos into smaller, time-aligned chunks encoded at various bitrates, enabling seamless quality switches according to bandwidth and device performance [1]. Advanced techniques like the Buffer Occupancy-based Lyapunov Algorithm (BOLA) enhance DASH by focusing on real-time buffer management over bandwidth predictions, addressing common algorithmic pitfalls [51]. Additionally, Blockchain-based architectures bolster secure content delivery and user privacy [32]. The Buffer-based Streaming Control (BSC) method further augments DASH by reducing interruptions through simultaneous multi-quality layer transmissions [52]. Comprehensive surveys highlight DASH's evolution, protocols, and methodologies for enhancing QoE in diverse networks [50], emphasizing the criticality of robust adaptive streaming protocols for managing packet loss [28].

4.2 Integration with CDNs

Integrating DASH with Content Delivery Networks (CDNs) significantly improves video streaming by reducing latency, enhancing network adaptability, and optimizing resources. Techniques like in-network caching, HTTP adaptive streaming, and multipath transmission ensure high-quality delivery across varied environments. Edge caching reduces latency and boosts bandwidth by storing frequently accessed content closer to users, while adaptive bitrate algorithms adjust video quality in real-time [10, 53, 1, 14].

Figure 4 illustrates the integration of DASH with CDNs, highlighting techniques, optimization methods, and collaborative approaches used to enhance video streaming quality and efficiency. Utilizing Software Defined Networking (SDN) for IP multicast optimizes delivery paths and reduces traffic [54], with dynamic flow scheduling strategies enhancing resources and QoE [55]. The BOLA algorithm prioritizes buffer occupancy for smoother transitions and fewer interruptions [51]. Collaborative algorithms like Joint Collaborative Caching and Processing (JCCP) aid real-time transcoding and caching, enhancing adaptability and reducing latency [27]. Integrating CDN and ICN slices further optimizes delivery by strategic caching [26]. Machine learning methods, such as Dynamic Closed-Loop QoE Optimization (DCL-QOE), fine-tune video quality based on user feedback and network conditions, maintaining high QoE in dynamic environments [56].

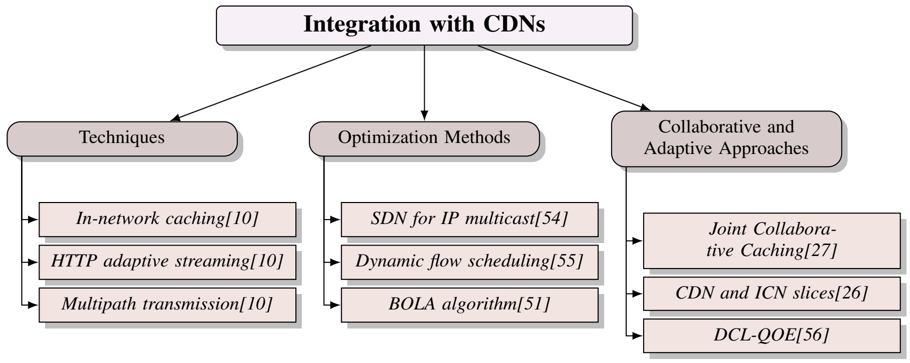

Figure 4: This figure illustrates the integration of DASH with CDNs, highlighting techniques, optimization methods, and collaborative approaches used to enhance video streaming quality and efficiency.

4.3 Challenges in Adaptive Streaming

Adaptive streaming technologies, including DASH, face challenges impacting efficiency and quality. Integrating advanced quality metrics like the Structural Similarity Index (SSIM) into Adaptive Bitrate (ABR) algorithms complicates implementation and requires refinement [57]. Delivering video chunks over TCP is challenging under fluctuating conditions; TCP's congestion control can impede DASH performance, necessitating advanced protocols or optimizations [58]. Solutions like the CANE cascade control framework model ABR behavior for effective bandwidth allocation through predictive control [59]. Buffer management is crucial, as poor handling can lead to starvation and degrade user experience. While recursive buffer management offers flexibility, it often involves high complexity, challenging real-time implementation [60]. Future research should explore real-time bandwidth allocation adjustments for better streaming quality under variable conditions [61]. Error management in CDNs is vital for adaptive streaming; clustering CDN error logs can reveal patterns for optimizing service quality and mitigating disruptions [62]. Leveraging user viewport similarities can enhance caching efficiency, saving bandwidth and improving delivery [63]. Security threats like random query string DoS attacks must be addressed; strategies such as the proposed gossip protocol can protect CDNs from vulnerabilities, ensuring reliable content delivery [64].

5 Quality of Experience (QoE) Optimization

5.1 Adaptive Streaming and QoE Enhancement

Adaptive streaming is crucial for optimizing Quality of Experience (QoE) by dynamically adjusting video quality in response to network conditions and user preferences. Advanced metrics, like the Structural Similarity Index (SSIM), integrated into Adaptive Bitrate (ABR) algorithms, enhance video quality assessment beyond traditional metrics, thereby improving user satisfaction [29]. The Dynamic Adaptive Streaming using Internet Protocol (DAS-IP) exemplifies this by enabling informed streaming decisions based on buffer levels and network conditions, enhancing QoE [27].

Information-Centric Networking (ICN) combined with Content Delivery Networks (CDNs) significantly boosts content delivery efficiency, reduces congestion, and shortens response times, positively impacting QoE [26]. Adaptive caching solutions, requiring minimal manual intervention, are vital for optimizing content delivery, as application-specific caching techniques demonstrate [21]. Startrail, an adaptive edge content delivery mechanism, improves request latency by 30

Advancements in optical networking technologies enhance bandwidth capacity and reduce electronic processing needs, supporting adaptive streaming in high-demand scenarios [25]. Leveraging these technologies ensures that CDNs deliver high-quality content efficiently, addressing evolving digital landscape demands. HTTP Adaptive Streaming (HAS) enhances QoE by dynamically adjusting video quality based on real-time network conditions, employing advanced metrics to monitor bandwidth fluctuations and implement personalized quality adjustments [50, 65, 66]. Integrating efficient network solutions like Software-Defined Networking (SDN) and Network Function Virtualization (NFV) further enhances resource management, ensuring seamless streaming across diverse multimedia applications.

Figure 5 illustrates the hierarchical categorization of adaptive streaming strategies and QoE enhancement techniques, highlighting key areas such as adaptive streaming, content delivery mechanisms, and networking technologies. This figure underscores the integration of advanced metrics, content delivery innovations, and networking advancements in enhancing user experience.

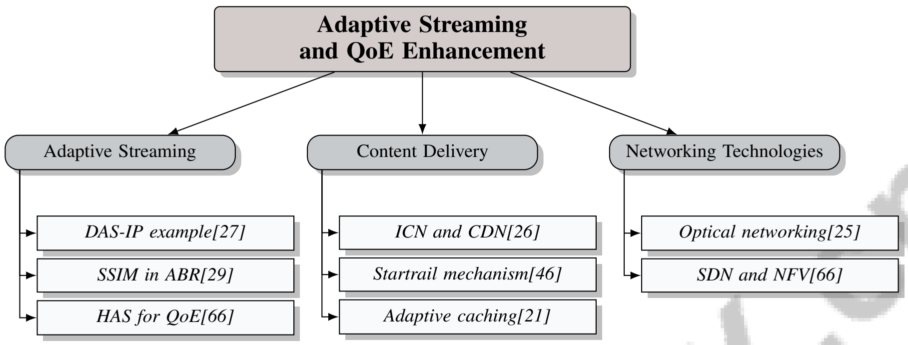

Figure 5: This figure illustrates the hierarchical categorization of adaptive streaming strategies and QoE enhancement techniques, highlighting key areas such as adaptive streaming, content delivery mechanisms, and networking technologies. The figure underscores the integration of advanced metrics, content delivery innovations, and networking advancements in enhancing user experience.

5.2 Buffer Management and Error Correction

Effective buffer management and error correction are pivotal for optimizing QoE in multimedia streaming within dynamic and heterogeneous network environments. The variability of network resources and the emergence of new multimedia services necessitate context-aware QoE management strategies capable of adapting to fluctuating conditions [66]. Dynamically adjusting video source sending rates based on real-time network conditions and managing session admissions ensures consistent QoE [67], ensuring smooth video delivery even amidst varying network conditions.

Mobile scenarios, such as those involving Unmanned Aerial Vehicles (UAVs), pose additional challenges due to high mobility, which can introduce delays and complications in transmitting high-quality video [68]. Sophisticated buffer management techniques that proactively allocate resources and predict potential congestion points are essential to minimize buffer underruns and playback interruptions. Error correction mechanisms are crucial for maintaining video quality by compensating for packet loss and transmission errors. Modeling QoE from visual renderings captures the visual information influencing user experience, allowing for effective error correction strategies that prioritize elements critical to user satisfaction [69].

5.3 Trade-offs and Metrics in QoE Optimization

Optimizing QoE within CDNs involves navigating the complex interplay between user satisfaction and various performance metrics in dynamic network environments. Existing adaptive streaming algorithms often struggle to balance trade-offs between video quality and starvation probability under fluctuating conditions [70], highlighting the need for sophisticated QoE optimization techniques that can adapt to real-time network performance changes.

A comprehensive understanding of user QoE, informed by visual renderings, provides a nuanced perspective on user satisfaction and informs the development of effective optimization strategies [69]. These strategies must consider diverse user expectations regarding video quality, which can vary significantly across different contexts and applications. Integrating adaptive frameworks, such as distributed load management strategies, improves load distribution and reduces overload situations within CDNs [37]. These approaches underscore the importance of managing performance trade-offs to allocate resources efficiently, ensuring high QoE without overburdening network infrastructure.

However, current studies often overlook application-specific characteristics, leading to generalized solutions that may not perform optimally across all scenarios [21]. Tailoring QoE optimization

AI-generated, for reference only.

techniques to the specific requirements of various applications is crucial for achieving optimal performance and user satisfaction.

6 Edge Server Placement Strategies

6.1 Models and Algorithms for Server Placement

Optimizing edge server placement is essential for enhancing Content Delivery Networks (CDNs), particularly for latency-sensitive applications like live video streaming. Diverse models and algorithms have been devised to strategically position servers, ensuring efficient content delivery and high user satisfaction. The CPVNF method employs a PageRank-based heuristic for selecting surrogate servers for Virtual Network Function (VNF) deployment, balancing server capacity and link quality to enhance CDN adaptability and performance under changing user demands and network conditions [71].

InviCloak offers a strategy for unilateral deployment by websites without altering CDN infrastructure, simplifying adoption compared to previous solutions [7]. This underscores the importance of flexible deployment strategies in server placement. Insights from cloud consolidation research suggest that strategic proxy deployment can enhance performance without requiring changes from service providers, providing a novel perspective on server placement [72].

Startrail improves access times and reduces the load on original content providers by enabling decentralized nodes to cache and serve popular content [46]. This decentralized model highlights edge computing's potential in optimizing server placement strategies. Additionally, the OPAC model provides a mathematical framework for determining optimal locations for migrating or placing virtual CDN components, ensuring alignment with user quality requirements [27].

6.2 Integration with Emerging Technologies

Emerging technologies are revolutionizing server placement strategies, enhancing CDN efficiency and adaptability to dynamic network conditions and user demands. Network-Assisted Video Streaming (NAVS) frameworks leverage Software Defined Networking (SDN), Network Functions Virtualization (NFV), and Multi-access Edge Computing (MEC) paradigms to improve video streaming services [24]. These technologies dynamically allocate resources and optimize server placement, minimizing latency and ensuring reliable content delivery.

As illustrated in Figure 6, the integration of these emerging technologies in server placement strategies highlights the roles of NAVS, adaptive learning techniques, and the synergy between Peer-to-Peer (P2P) streaming and CDN capabilities. These advancements collectively enhance CDN efficiency, Quality of Experience (QoE), scalability, and network traffic management.

Future advancements are expected to incorporate adaptive learning techniques for refining workload predictions, strengthening server placement and scheduling strategies [73]. Key areas include developing robust QoE models, exploring edge computing solutions, and fostering collaboration between application-layer algorithms and network infrastructure [10]. Adaptive scheduling algorithms using machine learning show promise in accurately predicting QoE, enabling real-time adjustments in server placement and resource allocation [74].

The effectiveness of these integrated technologies is further demonstrated by methods optimizing server locations based on customer profiles, enhancing cache hit ratios and reducing access latency through balanced server loads and strategic placements [75]. Additionally, merger-aware routing techniques maximize traffic merging opportunities, improving overall network efficiency [76].

Integrating P2P streaming with CDN capabilities represents a significant innovation for enhancing scalability and quality adaptation in video delivery [53]. This approach leverages the decentralized nature of P2P networks to complement CDN infrastructures, increasing their capacity to manage traffic loads and deliver content efficiently.

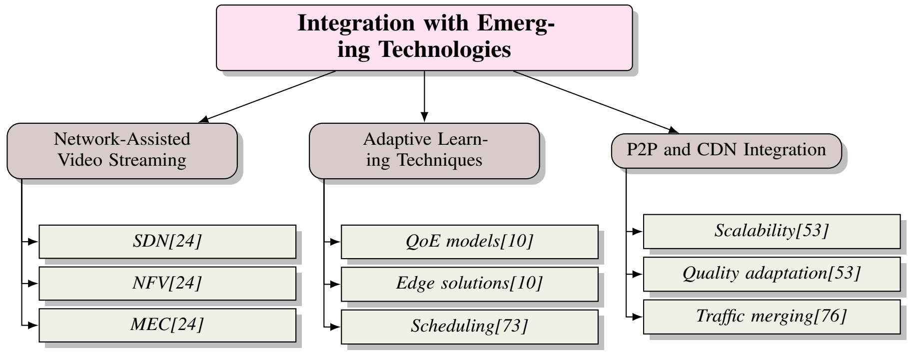

Figure 6: This figure illustrates the integration of emerging technologies in server placement strategies, highlighting the roles of Network-Assisted Video Streaming, Adaptive Learning Techniques, and the integration of P2P with CDN capabilities. These advancements enhance CDN efficiency, QoE, scalability, and network traffic management.

Benchmark	Size	Domain	Task Format	Metric
BOLA-BASIC[57]	685,022	Video Streaming	Adaptive Bitrate Streaming	SSIM, Stall Ratio
FB-Akamai[77]	2,000,000	Network Infrastructure	Latency Measurement	Average Access Delay
IACK[78]	600,000	Transport Protocols	Connection Setup Performance Evaluation	Time to First Byte, Probe Timeout
PeerCDN[79]	3,000,000	Video Streaming	Content Delivery Performance Evaluation	QoS metrics, latency
CDN-Convert[80]	1,000,000	Cdn Security	Ddos Attack Simulation	Amplification Factor
DNS-Bench[81]	17,573	Content Delivery Networks	Latency Measurement	DNS Latency, Mapping Latency
DNS-WH[23]	10,000	Web Hosting	Quantitative Analysis	Unreachable, Affected

Table 1: This table presents a comprehensive overview of various benchmarks utilized in the evaluation of network performance across different domains. It details the benchmark names, their respective sizes, domains of application, task formats, and the metrics used for performance evaluation. Such benchmarks are crucial for assessing the effectiveness of pricing models and resource allocation strategies in Content Delivery Networks (CDNs).

7 Bandwidth Cost Reduction Techniques

7.1 Pricing Models and Economic Viability

The economic feasibility of bandwidth cost reduction in CDNs is largely dependent on the implementation of effective pricing models that enhance resource utilization while maintaining service quality. Notable models include flat pricing, usage pricing, and congestion pricing [82]. Flat pricing offers simplicity but may lead to inefficiencies as it does not account for actual usage. Usage pricing aligns costs with data consumption, providing precise billing but introducing cost variability. Congestion pricing incentivizes users to modify demand during peak periods, effectively managing network congestion.

Figure 7 illustrates the economic viability of CDNs through key pricing models, resource allocation strategies, and security and sustainability measures. This figure highlights the aforementioned pricing models, alongside resource management techniques such as the Network Bandwidth Allocation (NBA) problem, which provides a mathematical framework crucial for optimizing resource allocation and controlling bandwidth costs under dynamic network conditions [83]. Techniques like PCDN+ enhance bandwidth cost management by utilizing predictive models to anticipate demand shifts and adjust resources accordingly [33]. However, these methods face challenges in environments with unpredictable workloads, necessitating adaptive strategies for sustained efficiency [40].

Energy-aware load balancing strategies contribute to economic viability by achieving over 55% energy savings in CDNs while ensuring high service availability and minimizing server transitions [84]. These strategies reduce operational costs and support sustainability, enhancing the economic appeal of CDN solutions. Furthermore, robust security frameworks in Peer-assisted Delivery Networks (PDNs) mitigate security threats, preventing costly breaches and maintaining user trust, thus supporting long-term economic sustainability [30].

Table 1 provides a detailed enumeration of benchmarks that are instrumental in evaluating the performance and economic viability of pricing models in Content Delivery Networks (CDNs).

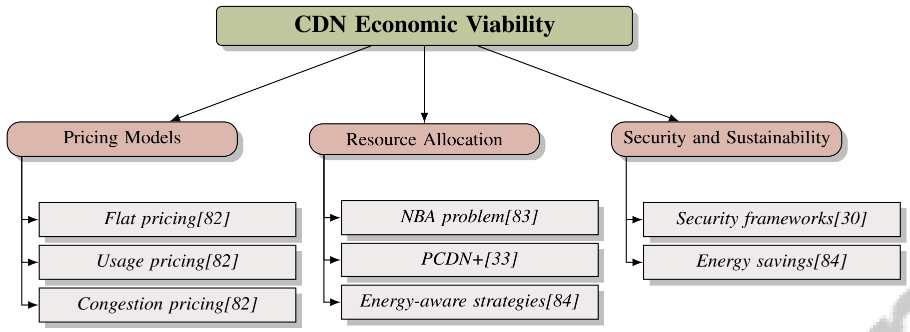

Figure 7: This figure illustrates the economic viability of CDNs through key pricing models, resource allocation strategies, and security and sustainability measures. It highlights flat, usage, and congestion pricing, resource management techniques like NBA and PCDN+, and security frameworks for long-term sustainability.

7.2 Utilization of Underutilized Resources

Maximizing the use of underutilized resources is crucial for reducing expenses in CDNs, particularly as demand for high-quality content delivery grows. Cloud-based solutions for dynamic resource allocation based on real-time demand optimize server usage and minimize idle times, thereby enhancing resource efficiency and reducing operational costs [40].

Integrating P2P streaming within CDN infrastructures provides an innovative approach to exploiting underutilized resources by distributing content delivery across a peer network, alleviating the load on centralized servers, and improving scalability [53]. This method enhances load balancing, ensuring effective resource allocation to prevent bottlenecks and improve network performance.

Energy-aware load balancing further optimizes underutilized resources by adjusting server workloads based on energy consumption patterns, achieving significant energy savings while maintaining service availability [84]. This approach reduces operational costs and supports the sustainability of CDN operations, aligning with broader environmental goals.

Advanced caching techniques, particularly those employing machine learning algorithms, enhance the utilization of underutilized resources by predicting content demand and optimizing cache placement [33]. These techniques enable CDNs to make informed caching decisions, ensuring efficient resource use and optimizing content delivery for cost 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) significantly enhance network performance in Content Delivery Networks (CDNs) by reducing latency and improving data transmission efficiency. SDN improves network flexibility by separating control and data planes, allowing for dynamic resource management [66]. This adaptability supports innovative prefetching methods that boost data transmission rates, particularly in scalable video streaming [85]. SDN-based infrastructures outperform traditional methods in network utilization and flow setup times, proving their efficacy in performance optimization [86].

NFV complements SDN by virtualizing network services, enabling function deployment on standard hardware, which reduces costs and enhances scalability. This virtualization facilitates dynamic resource allocation, crucial for adjusting to CDN service demands. NFV's efficiency is heightened by real-time data utilization, preventing over-provisioning by efficiently meeting client requests [87]. Future studies should refine resource allocation algorithms and explore more CDN as a Service (CDNaaS) use cases, integrating machine learning for predictive resource management [88].

Advanced algorithms like the Honeybee algorithm achieve a balance between communication cost and load balancing, crucial for optimizing CDN performance [89]. These algorithms ensure efficient load distribution across network nodes, minimizing bottlenecks and latency.

Integrating IoT devices and merging redundant video streams through ECoV optimizes network traffic, showcasing event-driven processing's potential in enhancing CDN efficiency [76]. Experiments in cloud computing, including content and live video delivery networks, highlight these technologies' importance in performance optimization [83]. Additionally, tools like YTrace offer comprehensive performance analysis across service and network stacks, providing insights into user experience issues and enabling targeted optimizations [90].

8.2 Latency Reduction through Strategic Server Placement

Strategic server placement is crucial for minimizing latency in Content Delivery Networks (CDNs) by positioning content closer to end-users. A significant challenge in latency reduction is the sub-optimal utilization of existing fiber networks, often dictated by Internet Service Provider (ISP) routing policies that prioritize cost over latency, complicating latency measurement across various networks [91]. Innovative placement strategies are necessary 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, effectively addressing deployment complexities and ensuring optimal resource allocation [13]. By strategically placing servers at the network edge, CDNs can significantly reduce data travel distance, thus minimizing latency and enhancing user experience.

Figure 8 illustrates the hierarchical structure of latency reduction strategies in Content Delivery Networks (CDNs), focusing on server placement, performance analysis, and load distribution. Analyzing CDN performance through raw log files allows detailed categorization by user access patterns and service types. This analysis offers insights for system optimization, facilitating informed decisions on server placement and resource allocation [17]. Understanding user interactions and identifying service usage patterns enable CDNs to strategically position servers to match specific demand profiles, further reducing latency.

Moreover, distributing the load across multiple peers and paths can alleviate congestion and enhance streaming quality. Studies emphasize leveraging decentralized architectures to balance network load, reducing bottlenecks and improving content delivery [53]. Adopting such distributed strategies strengthens CDNs' resilience against traffic spikes, ensuring consistent performance across diverse network conditions.

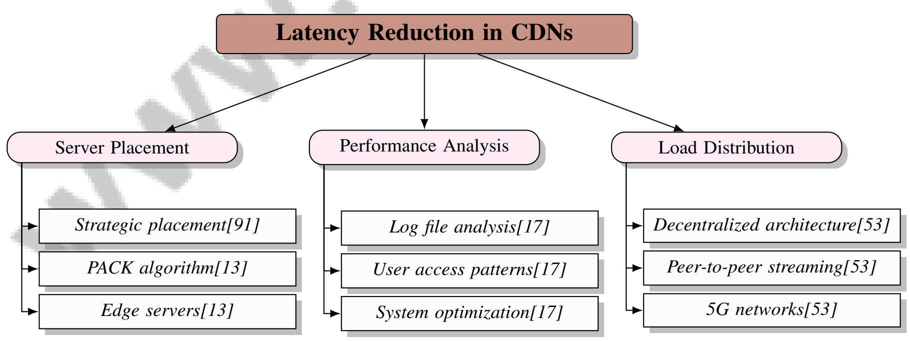

Figure 8: This figure illustrates the hierarchical structure of latency reduction strategies in Content Delivery Networks (CDNs), focusing on server placement, performance analysis, and load distribution.

9 Conclusion

Content Delivery Networks (CDNs) are crucial in enhancing content distribution efficiency within modern internet infrastructures. By integrating advanced techniques such as closed-loop Quality of Experience (QoE) optimization, CDNs have significantly improved video quality delivery, effectively responding to user feedback and network conditions, thereby demonstrating their real-world applicability. The deployment of IPv6 in high-density areas provides valuable insights for further optimization. Nonetheless, CDNs face challenges due to the rise of encrypted traffic and the growing need for enhanced user privacy, which could impact their long-term viability.

Innovative solutions like InviCloak offer promising opportunities to bolster privacy while maintaining CDN performance, though future research must address challenges related to computational overhead and the deployment of integrity verifiers. Additionally, recent findings on the security vulnerabilities of Peer-assisted Delivery Networks (PDNs) highlight the necessity for further exploration of mitigation strategies and increased user awareness. Expanding the geographic range of web performance measurements and utilizing machine learning techniques could enhance predictive capabilities for loading times and performance analysis.

The integration of DNS and web hosting services has significant implications for internet resilience and security, underscoring the need for strategic resource management and enhanced system robustness. Moreover, combining Information-Centric Networking (ICN) with CDN architectures has been demonstrated to improve content delivery efficiency, alleviate network congestion, and reduce response times. Future research should focus on enhancing system scalability and optimizing resource management strategies to achieve superior performance.

Ongoing innovation and research in CDN technologies are critical to addressing the evolving needs of modern internet infrastructure. Future efforts should prioritize refining collaboration frameworks, exploring automated mechanisms for improved technology integration, enhancing resource allocation algorithms with predictive models, and developing self-adaptive caching techniques that utilize application-specific knowledge. By leveraging emerging technologies and strategic methodologies, CDNs can enhance content delivery efficiency, scalability, and user satisfaction, paving the way for future advancements in content distribution networks.

AI-generated, for reference only.

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

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

[13] 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.

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

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

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

[17] 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.

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

[19] 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.

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

[21] 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.

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

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

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

[25] 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.

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

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

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

[29] 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.

[30] 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.

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

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

[33] 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.

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

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

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

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

[38] Oliver Hohlfeld, Jan Rüth, Konrad Wolsing, and Torsten Zimmermann. Characterizing a meta-cdn, 2019.

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

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

[41] Pawan Poojary, Sharayu Moharir, and Krishna Jagannathan. Caching under content freshness constraints, 2017.

[42] Eric Friedlander and Vaneet Aggarwal. Generalization of lru cache replacement policy with applications to video streaming, 2019.

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

[44] 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.

[45] Soumya Basu, Aditya Sundarrajan, Javad Ghaderi, Sanjay Shakkottai, and Ramesh Sitaraman. Adaptive ttl-based caching for content delivery, 2017.

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

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

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

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

[50] 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.

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

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

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

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

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

[56] 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.

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

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

[59] 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.

[60] 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.

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

[62] 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.

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

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

[65] Hiba Yousef. *Adaptive streaming using Peer-to-Peer and HTTP*. PhD thesis, Institut Polytechnique de Paris, 2021.

[66] 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.

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

[68] 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.

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

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

[71] 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

[85] 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.

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

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

[88] 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.

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

[90] 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.

[91] 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.

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.