International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 01 | January - 2024 SJIF Rating: 8.176 ISSN: 2582-3930

A Novel Approach to Improve Quality of Service in Cloud-Based Content Delivery Network

Priyanka D[1], Vedanth R Kumar[2], Gowthami H S[3], Manikant V Hooli[4], Shivaraja pateel K S[5],
[1]Priyanka D, ISE, Vidya Vikas Institute of Engineering & Technology, Mysore
[2]Vedanth R Kumar, ISE, Vidya Vikas Institute of Engineering & Technology, Mysore [3]Gowthami H S, ISE, Vidya Vikas Institute of Engineering & Technology, Mysore [4]Manikant V Hooli, ISE, Vidya Vikas Institute of Engineering & Technology, Mysore
[5]Shivaraja pateel K S, ISE, Vidya Vikas Institute of Engineering & Technology, Mysore

Abstract: A content delivery network (CDN) improves normal website functions and increases customer satisfaction. In addition, CDNs foster a seamless online experience, ensuring reliable performance and bolstering cybersecurity measures for enhanced user trust and data protection. It improves High-speed content delivery, Real-time streaming, Multi-user scaling etc. One of the most notable achievements is the remarkable reduction in latency, thanks to advanced CDN technologies that position content closer to end-users, enhancing responsiveness and making real-time applications, online gaming, and interactive content more enjoyable. Furthermore, cost optimization has been a focal point, with CDNs efficiently managing resources and data transfer to deliver content while keeping operational expenses in check. In essence, the ongoing enhancements in CDNs underscore their pivotal role in shaping the digital content landscape. With a focus on speed, scalability, latency reduction, cost efficiency, security, global reach, and data-driven decision-making, CDNs have become indispensable for organizations seeking to excel in serving an increasingly diverse and demanding global audience. In a world where digital content reigns supreme, these improvements empower organizations to meet and exceed user expectations, contributing to their overall success in the digital age.

Keywords: Content Delivery Network, Cloud-based Content Delivery Network, Quality metrics and monitoring, Resource management, Performance optimisation.

1. INTRODUCTION

A content delivery network (CDN) is a network of interconnected servers that speeds up webpage loading for data-heavy applications. If the user is located far from that server, it will take a long time to load a large file, such as a video or website image. Instead, the website content is stored on CDN servers geographically closer to the users and reaches their computers much faster. The primary purpose of a content delivery network (CDN) is to reduce latency, or reduce the delay in communication created by a network's design. Because of the global and complex nature of the internet, communication traffic between websites (servers) and their users (clients) has to move over large physical distances. The communication is also two-way, with requests going from the client to the server and responses coming back. A Cloud-based Content Delivery Network (CDN) is designed to efficiently distribute and deliver web content, such as web pages, images, videos, and other assets, to users around the world. The architecture of a CDN involves several key components and technologies that work

together to optimize content delivery, reduce latency, improve scalability, and enhance security. The typical architecture of a cloud-based CDN are content origin servers, edge servers (CDN pops - points of presence), content caching, content routing and load balancing, content purge and invalidation, dynamic content acceleration, security and DDOS protection, analytics and monitoring, content compression and optimization, https and content encryption, global network backbone, scalability, billing and usage reporting, content delivery optimization.

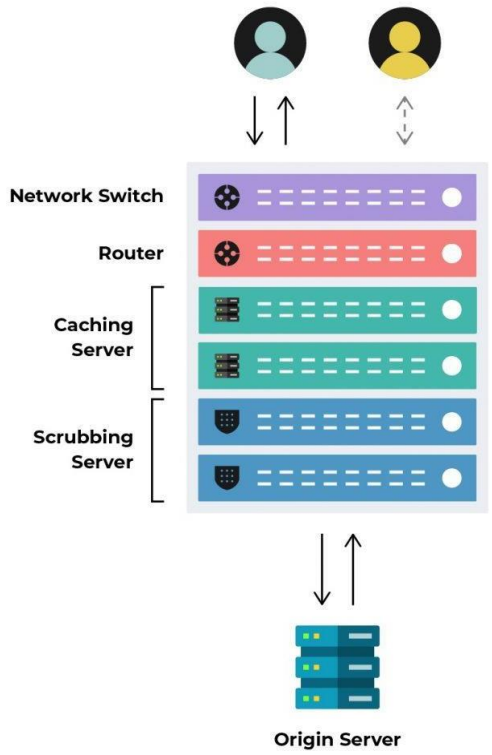
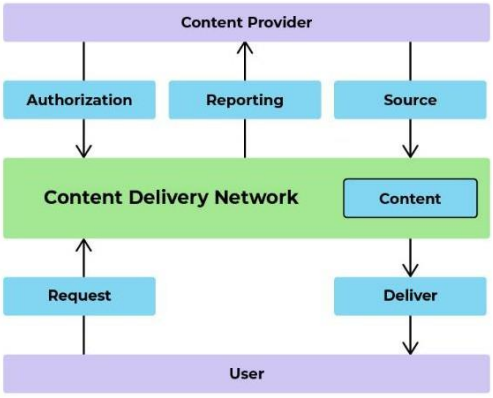


Fig 1:A Simple architecture of CDN

The content origin servers are the original source of the web content, such as web applications, databases, and storage servers. These servers store the master copies of the content. Edge servers are strategically distributed in multiple locations around the world. These edge locations are often hosted in data centers or cloud infrastructure. They serve as caching servers that store copies of the content closer to end-users. This reduces the distance and latency between users and the content they request. Content is cached at edge servers based on various caching policies. Popular and frequently accessed content is cached for faster retrieval. Cache eviction policies and TTL (Time-to-Live) settings determine how long content remains in the cache. CDNs use intelligent routing algorithms to direct user requests to the nearest or most appropriate edge server. Load balancing ensures that edge servers are efficiently utilized. CDNs provide mechanisms to purge or invalidate cached content when changes are made to the original content on the origin server. This ensures that users always receive the latest content. In addition to static content, CDNs can accelerate the delivery of dynamic content, such as personalized web pages and real-time data, through techniques like edge-side caching, serverless computing, and content optimization. CDNs offer security features, including DDoS mitigation, web application firewalls (WAFs), and SSL/TLS termination to protect content and applications from cyber threats. CDNs provide analytics tools to monitor traffic, performance, and user behavior. This data helps optimize content delivery and troubleshoot issues. CDNs often employ techniques like content compression and image optimization to reduce the size of assets, improving load times. CDNs support secure connections through HTTPS, encrypting data in transit between edge servers and users. CDNs maintain a high-speed, global network backbone with redundant connections to ensure reliability and low-latency content delivery. CDNs are designed to handle traffic spikes and accommodate growing demand. Additional edge servers can be deployed as needed. CDNs provide billing and reporting tools that allow customers to monitor usage and manage costs effectively. Advanced CDNs may use machine learning and AI algorithms to optimize content delivery based on user behavior and network conditions.

Let's see a simpler model of a CDN:



: A Simpler model of content delivery network

Content delivery networks (CDNs) provide many benefits that improve website performance and support core network infrastructure. For example, a CDN can do the following tasks like Reduce page load time, Website traffic can decrease if your page load times are too slow. A CDN can reduce bounce rates and increase the time users spend on your site. Reduce bandwidth costs, bandwidth costs are a significant expense because every incoming website request consumes network bandwidth. Through caching and other optimizations, CDNs can reduce the amount of data an origin server must provide, reducing the costs of hosting for website owners. Increase content availability, too many visitors at one time or network hardware failures can cause a website to crash. CDN services can handle more web traffic and reduce the load on web servers. Also, if one or more CDN servers go offline, other operational servers can replace them to ensure uninterrupted service. Improve website security, Distributed denial-of-service (DDoS) attacks attempt to take down applications by sending large amounts of fake traffic to the website. CDNs can handle such traffic spikes by distributing the load between several intermediary servers, reducing the impact on the origin server.

2. PROBLEM STATEMENT

Despite improvements in cloud-based Content Delivery Networks (CDNs), a fresh strategy that can significantly improve end-users' quality of service (QoS) is urgently needed. Optimizing content delivery speed, lowering latency, and guaranteeing consistent performance across various geographic regions and under variable network circumstances are issues faced by current CDNs. Therefore, there is an urgent need to find solution to the following problem: "How can a novel approach be developed to significantly improve the QoS in cloud-based CDNs, addressing problems with speed, latency, and reliability, ultimately enhancing the end-user experience?

3. METHODOLOGY

Methods in Improving quality of service in cloud-based content delivery network by improving:

Performance Optimization: Work on optimizing the performance of the C-CDN infrastructure. This could involve designing efficient caching mechanisms, load balancing algorithms, and traffic routing strategies to reduce latency and improve response times.

Scalability and Resource Management: Focus on making the C-CDN scalable and capable of handling growing traffic demands. Develop resource management techniques to dynamically allocate cloud resources based on real-time demand.

Quality Metrics and Monitoring: Define appropriate quality metrics for C-CDN performance, such as response time, throughput, and error rates. Implement monitoring systems to continuously track and analyze these metrics, allowing for

quick identification of performance bottlenecks.

Machine Learning and AI: Explore the potential of using machine learning and AI algorithms to optimize content delivery. This could include predictive caching, personalized content recommendations, and intelligent traffic management.

Content Optimization: Collaborate with content providers to optimize their content for efficient delivery in the C-CDN. This may involve image and video compression, content transformation, and adaptive streaming techniques.

4. DESIGN AND IMPLEMENTATION

4.1 Performance Optimization:

CDNs can employ techniques like Anycast routing, which directs user requests to the nearest available server, reducing the number of hops and minimizing latency. Intelligent routing algorithms can dynamically adapt to changing network conditions and select the fastest paths. CDNs can implement content-based cache invalidation, where changes to specific content trigger cache updates. Time-based or event-based strategies can also help maintain a balance between cache freshness and efficiency. Predictive resource scaling anticipates demand and maintains a certain level of activity on servers, reducing the likelihood of prolonged cold starts. "Warm-up" procedures can also proactively load necessary resources during low-demand periods.

4.2 Scalability and Resource Management:

Work closely with your cloud service provider to leverage their scaling features, such as auto-scaling groups and load balancers. Develop strategies for resource allocation based on factors like server load, geographic distribution, and traffic patterns. Use load testing to identify resource bottlenecks and optimize resource allocation.

4.3 Quality metrics and Monitoring:

Reduce latency by optimizing routing algorithms and leveraging Anycast routing for faster content delivery. Increase available bandwidth to accommodate larger traffic spikes by working with your hosting provider or scaling up resources. Fine-tune cache policies to ensure that frequently accessed content is effectively cached and reduce cache misses. Implement better error handling and redirection strategies to minimize instances of 404 or 500 errors.

4.4 Machine Learning and AI:

Implement advanced ML models that take into account contextual information such as user location, device type, and browsing history for more accurate predictions. Develop AI-driven algorithms that consider both historical and real-time data to make resource allocation decisions. Fine-tune auto-scaling thresholds and resource provisioning strategies to minimize overprovisioning during sudden demand spikes.

4.5 Content Optimization:

Reduce the size of your website's assets, such as images, CSS, JavaScript, and HTML files. Use compression techniques like GZIP and Bartoli to further shrink file sizes without sacrificing quality. Smaller files load faster and consume less bandwidth. Configure your CCDN to enable content compression. Most modern CCDNs support compression techniques that reduce the size of the data being transmitted between the CCDN and the end user's device.

5. FUTURE TRENDS:

The future trends of cloud-based Content Delivery Networks (CDNs) are likely to include:

5.1 Edge Computing Integration:

CDNs will increasingly integrate with edge computing, bringing processing closer to the end-users for improved latency and faster content delivery. AI and Machine Learning Optimization: Implementation of AI and machine learning algorithms to dynamically optimize content delivery routes, adapt to changing network conditions, and enhance overall performance.

5.2 Security Enhancements:

Advanced security features to protect against evolving cyber threats, with increased emphasis on DDoS mitigation, web application security, and encryption.5G Integration: Integration with 5G networks for faster and more reliable content delivery, especially for bandwidth-intensive applications like augmented reality (AR) and virtual reality (VR).

5.3 Multi-Cloud and Hybrid CDN Solutions:

Adoption of multi-cloud and hybrid CDN solutions to provide flexibility, redundancy, and optimize costs by leveraging multiple cloud providers.

5.4 Serverless CDN Architectures:

Utilization of serverless architectures to scale resources dynamically based on demand, leading to more cost-effective and efficient CDN solutions.

5.5 IoT Support:

CDN services tailored to support the growing number of Internet of Things (IoT) devices, ensuring efficient content delivery to a wide range of connected devices.

5.6 Global Expansion:

Increased emphasis on expanding CDN infrastructure globally to meet the rising demand for content delivery across diverse geographic locations.

5.7 Green Computing:

Integration of eco-friendly practices and energy-efficient technologies to reduce the environmental impact of CDN infrastructure.

5.8 Enhanced Analytics and Reporting:

Improved analytics tools to provide more detailed insights into content delivery performance, user behavior, and overall CDN efficiency.

These trends reflect the evolving landscape of cloud-based CDNs, driven by the need for faster, more reliable, and secure content delivery in an increasingly interconnected world.

6. CONCLUSION

In conclusion, the continuous evolution and improvements in Content Delivery Networks (CDNs) represent a transformative force in the realm of digital content distribution. These enhancements have not only accelerated content delivery but have also enriched the overall user experience. CDNs now excel in speed and efficiency, significantly reducing loading times and buffering issues, ensuring that users can access web content and multimedia swiftly and seamlessly. The integration of scalability and flexibility into CDNs allows them to adapt in real-time to varying traffic demands, guaranteeing uninterrupted, high-quality content delivery during traffic spikes or changing user requirements.

Moreover, with the implementation of real-time monitoring and analytics tools, organizations can proactively track performance, identify issues, and make data-driven decisions, thereby further optimizing content delivery. Finally, these CDN improvements grant organizations a competitive advantage by delivering a superior user experience compared to their competitors. This can lead to increased customer retention and an enhanced brand reputation.

In essence, the ongoing enhancements in CDNs underscore their pivotal role in shaping the digital content landscape. With a focus on speed, scalability, latency reduction, cost efficiency, security, global reach, and data-driven decision-making, CDNs have become indispensable for organizations seeking to excel in serving an increasingly diverse and demanding global audience. In a world where digital content reigns supreme, these improvements empower organizations to meet and exceed user expectations, contributing to their overall success in the digital age.

7. REFERENCES

[1] I. Leyva-Mayorga et al., "Network- coded cooperation and multi- connectivity for massive content delivery,"*IEEE Access*, vol. 8, pp. 15656–15672, 2020, doi: 10.1109/access.2020.2967278.

[2] Y. Fan, B. Yang, D. Hu, X. Yuan, and X. Xu, "Social- and content-aware prediction for video content delivery,"*IEEE Access*, vol. 8, pp. 29219–29227, 2020, doi: 10.1109/access.2020.2972920.

[3] M. S. Al-Abiad, A. Douik, and M. J. Hossain, "Coalition formation game for cooperative content delivery in network coding assisted D2D communications,"*IEEE Access*, vol. 8, pp. 158152–158168, 2020, doi: 10.1109/access.2020.3020472.

[4] J. Chuan, B. Bai, X. Wu, and H. Zhang, "Optimizing content placement and delivery in wireless distributed cache systems through belief propagation,"*IEEE Access*, vol. 8, pp. 100684–100701, 2020, doi: 10.1109/access.2020.2996222.

[5] H. Yang, H. Pan, and L. Ma, "A review on software defined content delivery network: A novel combination of CDN and SDN,"*IEEE Access*, vol. 11, pp. 43822–43843, 2023, doi: 10.1109/access.2023.3267737.

[6] C. T. E. R. Hewage, A. Ahmad, T. Mallikarachchi, N. Barman, and M. G. Martini, "Measuring, modeling and integrating time-varying video quality in end-to-end multimedia service delivery: A review and open challenges,"*IEEE Access*, vol. 10, pp. 60267–60293, 2022, doi: 10.1109/access.2022.3180491.

[7] M. I. A. Zahed, I. Ahmad, D. Habibi, Q. V. Phung, M. M. Mowla, and M. Waqas, "A review on green caching strategies for next generation communication networks,"*IEEE Access*, vol. 8, pp. 212709–212737, 2020, doi: 10.1109/access.2020.3040958.

[8] Z.-H. Wang and A.-L. Dong, "Research on load balancing and caching strategy for central network,"*J. Electr. Comput. Eng.*, vol. 2022, pp. 1–12, 2022, doi: 10.1155/2022/6601965.

[9] M. Karaata, A. Al-Mutairi, and S. Alsubaihi, "Multipath routing over star overlays for quality of service enhancement in hybrid content distribution peer-to-peer networks,"*IEEE Access*, vol. 10, pp. 7042–7058, 2022, doi: 10.1109/access.2021.3139936.

[10] J. Zhang, S. Li, and C. Wang, "A secure dynamic content delivery scheme in named data networking,"*Secur. Commun. Netw.*, vol. 2022, pp. 1–15, 2022, doi: 10.1155/2022/6304927.

[11] C. Fang et al., "Energy-efficient hierarchical collaborative scheme for content delivery in mobile edge computing," Wirel. Commun. Mob. Comput., vol. 2022, pp. 1–10, 2022, doi: 10.1155/2022/8392511.

[12] C. Natalino, A. de Sousa, L. Wosinska, and M. Furdek, "Content placement in 5G-enabled edge/core data center networks resilient to link cut attacks,"*Networks (N. Y.)*, vol. 75, no. 4, pp. 392–404, 2020, doi: 10.1002/net.21930.