A Survey on Secure Edge Computing for Content Delivery

Anitha H M
Department of Information Science and Engineering
B.M.S College of Engineering, Bengaluru, Karnataka, India

Rachita N
Department of Information Science and Engineering
B.M.S College of Engineering, Bengaluru, Karnataka, India

Sai varnitha Reddy
Department of Information Science and Engineering
B.M.S College of Engineering, Bengaluru, Karnataka, India

V Navya
Department of Information Science and Engineering
B.M.S College of Engineering, Bengaluru, Karnataka, India

Abstract- Digital content consumption has skyrocketed in recent years, and Content Delivery Networks (CDNs) have been critical in delivering this content to end users. CDNs employ edge servers to store digital content close to the end users in order to provide high Quality of Service (QoS). However, the edge servers in this architecture are vulnerable to attacks resulting in the deterioration of QoS and services provided to the end users. It is imperative to protect these edge servers in order to ensure the security and continuous delivery of data. To enhance the privacy of data and safeguard the architecture against security threats, certain effective security measures need to be adopted.

Keywords – Edge computing, Content Delivery Networks, Cloud computing, Machine learning, security, access control.

● I. INTRODUCTION

Edge computing is a technology that brings computation and data storage closer to the end user, rather than in a centralized data center. This is done by placing small, powerful computing devices at the network's edge, such as at the base of a cell tower or in a street cabinet. Due to its significant reduction in latency, edge computing is perfect for real-time applications like virtual reality and autonomous vehicles. Second, it improves security by reducing the volume of data sent over a network and the number of points of entry into a system. Thirdly, by spreading processing and storage across various devices, it enhances reliability and resilience. Fourthly, it lowers bandwidth needs and related expenses. Fifth, it aids businesses in observing privacy laws and other regulations.

Edge computing is a versatile technology with a myriad of applications in various industries. For instance, it can be used in industrial automation to process real-time data from sensors and machines, leading to optimized manufacturing processes, improved productivity, and reduced maintenance costs. In smart cities, edge computing can manage traffic, optimize energy consumption, and enhance the overall quality of life for residents. In healthcare, edge computing can process data from wearable devices and other sensors, enabling real-time monitoring of patients and personalized care. Additionally, edge computing can be used in autonomous vehicles to process sensor data and make critical decisions on navigation and safety. Other promising applications of edge computing include video surveillance, augmented and virtual reality, gaming, and more.

Edge computing presents several use cases, including Content Delivery Networks (CDNs), which are extensively utilized to provide web content to users based on their geographical location. CDNs cache content on servers in close proximity to the end users, thereby reducing latency and enhancing performance.

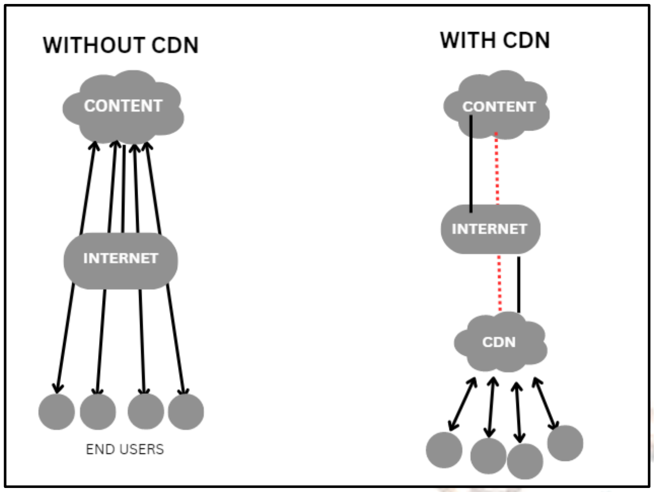

Fig 1. Content Delivery Network

While edge servers can be secure, they are not immune to security threats and it's important to stay vigilant and take necessary steps to protect them. Since they are closer to the network, they are prone to various physical attacks. Edge servers can be secure if they are properly configured and maintained. However, they may also be vulnerable to security threats like any other server, particularly if they are connected to the internet.

● II. BACKGROUND

○

○ *2.1 Edge computing vs Cloud computing*

Edge computing refers to a distributed computing architecture in which data is processed at the edge of the network, near the source of the data, rather than in a centralized data center. This allows for lower latency, faster response times, and the ability to process data even when there is no internet connection. Edge computing is often used in applications such as Internet of Things (IoT) devices, autonomous vehicles, and other situations where real-time data processing is required.

Cloud computing, on the other hand, refers to delivering computing resources and services over the Internet. The data is stored and processed in remote data centers, and users can access it through the Internet. Cloud computing is often used for data storage, software as a service (SaaS), and platform as a service (PaaS). The main advantages of cloud computing are scalability and cost-efficiency, as it allows businesses to access powerful computing resources without having to invest in expensive hardware.

Edge computing and cloud computing have different characteristics and serve different purposes. Edge computing is focused on low-latency and real-time processing while cloud computing is focused on scalability and cost-efficiency. They can be used together to create a hybrid architecture that combines the advantages of both technologies.

As Bilal et al. [1] specify, Content Delivery Networks(CDNs) aid in enhancing delivery system efficiency and minimizing media applications' network footprint. Cache servers spread across various locations are used by CDNs to replicate Internet content and bring it closer to end users. In addition to lessening the load on origin servers, increasing network bandwidth utilization, and improving user experience when accessing the service, this lowers the latency of content delivery and increases content availability. There is a high demand for dynamic CDN deployment practices due to the size and expansion of content providers. The development of edge-cloud computing and the competence of virtualization technologies have made it simpler to deploy virtual cache servers of CDNs over the Internet in order to satisfy these requirements

○ *2.2 Major issues in edge computing and their solutions*

As Parikh et al. [2] suggest the following are some of the main privacy and security issues with edge computing:

- Weak security measures make the system vulnerable to attacks by malicious actors
- Inadequate communication security between devices
- Difficulty in recovering and backing up data during system failures
- Delayed deployment of updates
- Limited visibility into the network
- Lack of user-controlled data collection.

To address the above specified issues, some solutions include

- Ensuring that all edge nodes are as secure as the network
- Delivering constant network monitoring and discoverability to users via an interactive platform
- Implementing encryption and access controls for data

- Utilizing intrusion detection systems to detect unauthorized access and alert users
- Using lightweight cryptography and monitoring hardware performance
- Monitoring user behavior and providing decoy information in case of suspicious activity.

○ *2.3 Content Delivery Network*

The main objective of this project is to focus on how edge computing helps with the security measures taken in Content Delivery Networks. For this, it is important to understand how CDN works using edge computing,

A content delivery network (CDN) is a distributed network of servers that work together to deliver content to end users. In edge computing, CDNs are often used to improve the performance and availability of web applications and other digital content. CDNs function by caching content at multiple edge locations, closer to the end-users, so that when a user requests the content, it can be delivered from the edge server closest to the user, rather than from a centralized origin server. This reduces the latency and improves the speed of content delivery. Edge computing extends the capabilities of CDNs by allowing more complex computations to be performed at the edge locations, such as data processing and analysis, and even running applications. This allows for faster response times and better performance for web applications and other content that requires real-time data processing.

CDNs can also use load balancing and failover mechanisms to ensure high availability and redundancy in case of server failures or network outages. Overall, CDNs play a critical role in edge computing by providing faster, more reliable, and scalable content delivery, while also reducing the load on centralized servers.

○ *2.4 Data Security And Privacy Requirements*

As Zhang et al. [3] specify, the following are the key requirements

- Confidentiality: Ensuring that sensitive information is protected from unauthorized access or disclosure.
- Integrity: Ensuring that the data or information stored in the system is accurate, complete and consistent.
- Availability: Ensuring that the system or service is accessible to authorized users when needed.
- Authentication and access control: Verifying a user, device, or system's identity
- Privacy requirement: Preventing sensitive data's unauthorized acquisition, use, disclosure, and disposal.

○ *2.5 Attacks and their defense mechanisms*

According to Xiao et al. [4], the different attacks and their mechanisms are as follows:

- DDoS attacks and their Defenses - This is the type of attack that disrupts normal services provided by a server based on a cluster of edge devices. It aims to destroy the legitimate use of a server. There are mutliple defense mechanisms for DDoS such as Traffic Filtering, Traffic Shaping, DDoS mitigation service, Botnet tracking, Cloud-based DDoS protection, etc.
- Side Channel Attacks and Defenses - These attacks compromise a user's security and privacy using any publicly accessible information that is not privacy-sensitive in nature, namely side channel information. Some of its defense mechanisms are Timing randomization, Power analysis countermeasures, Masking, Electromagnetic Shielding, etc.
- Malware Injection Attack - Malware is installed stealthily into a computing system. It is indeed one of the most dangerous attacks since it is a significant threat to system security and data integrity. Antivirus/Antimalware software can prevent these attacks including mechanisms like firewalls, Sandboxing, Patch Management, Network Segmentation, etc.
- Cache poisoning: Cache poisoning is a type of attack where attackers inject malicious content into the CDN's cache. When a user requests that content, they receive the malicious version instead of the legitimate one. This can result in malware infections, phishing attacks, and other security issues.
- Cross-site scripting (XSS): Cross-site scripting attacks involve injecting malicious scripts into web pages that are delivered through the CDN. When users visit those pages, the scripts execute in their browsers, allowing attackers to steal sensitive data, take control of user accounts, or launch other attacks.
- Brute force attacks: Brute force attacks can target the CDN's login pages, attempting to guess user credentials by trying different combinations of usernames and passwords. If successful, the attacker can gain unauthorized access to the CDN's management console, allowing them to perform malicious activities.
- DNS hijacking: DNS hijacking is a type of attack where attackers modify the DNS settings of a CDN to redirect users to malicious websites or intercept their traffic. This can lead to data breaches, session hijacking, and other security issues.

To mitigate these attacks, CDNs should implement security best practices, such as using strong encryption, regularly patching software and systems, and monitoring network traffic for suspicious activity. It's also essential to conduct regular security assessments and audits to identify and address vulnerabilities in the CDN's infrastructure.

○ *2.6 Some approaches to CDNs*

- Multi-CDN: A multi-CDN approach involves using multiple CDNs to deliver content to end-users. By leveraging multiple CDNs, organizations can improve the availability, reliability, and performance of their web content. If one CDN experiences a performance issue or outage, traffic can be automatically rerouted to another CDN. Multi-CDN can also help organizations mitigate security threats by spreading the risk across multiple CDNs.

- Hybrid CDN: A hybrid CDN approach combines the advantages of both private and public CDNs. Organizations can use a private CDN to deliver sensitive content, such as financial data or intellectual property, while using a public CDN to deliver non-sensitive content, such as marketing videos or blog posts. This approach can help organizations balance the need for security and performance with cost-effectiveness.
- Serverless computing: Serverless computing involves running applications without the need for dedicated servers. By leveraging serverless computing, CDNs can scale their resources automatically to meet demand, reducing the risk of performance issues or downtime. Serverless computing can also help organizations mitigate security threats by reducing the attack surface and minimizing the risk of vulnerabilities.
- CDN-as-a-Service: CDN-as-a-Service is a cloud-based approach to CDNs that allows organizations to leverage the benefits of CDNs without the need for dedicated infrastructure. CDN-as-a-Service providers offer scalable and cost-effective solutions for delivering web content, reducing the complexity and cost of managing CDNs in-house. This approach can also help organizations mitigate security threats by providing built-in security features and monitoring.

○ *2.7 Mobile Edge Computing for video delivery*

A Mobile Edge Computing (MEC)-based architecture for a better adaptive HTTP video delivery can include the following features:

- Content Caching: MEC can be used to cache video content closer to the user, reducing the latency and improving the quality of experience (QoE) for video streaming.
- Adaptive Bitrate Streaming: MEC can be used to dynamically adjust the video bitrate based on the available bandwidth and the user's device capabilities, providing a more consistent QoE.
- Network Function Virtualization (NFV): MEC can enable the deployment of virtual network functions (VNFs) such as video transcoding and video optimization at the edge, reducing the need for these functions to be performed in a centralized data center.
- Quality of Service (QoS): MEC can be used to provide QoS for video streaming, ensuring that the video traffic is given priority over other types of traffic, improving the overall QoE.
- Security: MEC can be used to secure video delivery by providing secure connections and encrypting the video content.
- Analytics: MEC can be used to collect analytics data on the video delivery, such as the number of users, the video bitrate, and the video resolution, which can be used to optimize the video delivery and improve the QoE.

Overall, a MEC-based architecture can significantly improve the adaptive HTTP video delivery by reducing the latency, improving the QoE and providing a more efficient and secure video delivery.

A Mobile Edge Computing (MEC)-based architecture is a distributed computing paradigm that enables the deployment of computing resources and services closer to the edge of the network, closer to end-users and devices. The architecture can be composed of the following components:

- Edge Cloud: The edge cloud is a cluster of servers and storage devices that are deployed at the edge of the network, typically in a base station or a data center.
- Edge Nodes: Edge nodes are devices that are deployed at the edge of the network, such as routers, gateways, and IoT devices.
- Network Function Virtualization (NFV): MEC allows the deployment of virtual network functions (VNFs) such as firewalls, load balancers, and video transcoding at the edge.
- Management and Orchestration: MEC requires a management and orchestration system to manage the edge cloud, edge nodes, and VNFs.

Overall, a MEC-based architecture enables the deployment of computing resources and services closer to the edge of the network, reducing the latency, improving the quality of experience (QoE) for end-users and providing a more efficient and secure computing environment.

A review of the Multi-Access Edge Computing(MEC) architecture of Ali et al. [5] from a security and privacy perspective can include the following topics:

- Data Encryption: It is important to protect sensitive information from unauthorized access or tampering. MEC can use various encryption methods such as Advanced Encryption Standard (AES) or Transport Layer Security (TLS) to secure data in transit and at rest.
- Firewall: A firewall can be used to control access to the MEC platform and protect it from unauthorized access or attacks.
- Virtualization Security: MEC relies heavily on virtualization to enable multiple tenants to share the same physical resources. Virtualization security is critical to protect against threats and attacks that can compromise the platform.
- Identity and Access Management: It is critical to ensure that only authorized users have access to the MEC platform and its resources.

○ *2.8 Data Encryption*

As Aluvalu [6] iterates the importance of it, the cloud computing environment is distributed and constantly changing, making static policies ineffective for controlling access. Therefore, dynamic policies are necessary to protect the confidentiality of stored data by encrypting it using cryptographic algorithms before uploading it to the cloud.

- One such method is attribute-based encryption, where secret keys and ciphertexts are based on user attributes such as location or subscription type. In this system, decryption is only possible when a certain number of attributes match.
- Another method is key policy attribute-based encryption, where messages are encrypted under a set of attributes and private keys are linked to access structures determining which ciphertexts the key holder can decrypt.

○ *2.9 Identity and Access Management*

Meghanathan et al. [7] briefs the various Access control mechanisms that can be deployed. There are several different access control models that can be used to secure cloud computing environments, including

- Role-based access control (RBAC), which assigns users roles based on the least privilege principle and the minimum permissions required for their job
- Task role-based access control (TRBAC), which dynamically validates access permissions based on the assigned roles and tasks
- Attribute-role-based access control (ARBAC), which assigns attributes and values to data objects and verifies users' access based on their role and the values they provide
- Multi-tenancy model, where a Cloud Service Provider (CSP) manages the addition, removal, and management of tenants and their associated security issues, and tenants manage their own access control lists and capability list

○ *2.10 CDNs deployed in Edge computing environments*

Duc et al. [8]: A CDN consists of 3 components:

- Origin servers
- Cache servers
- Clients

Cloud and edge computing are made possible by virtualization, which abstracts physical resources using virtual machines and containers. As a result, resources like processors, storage, and networks can be combined into software-defined infrastructures (SDIs), which offer automated resource allocation, scaling, and optimization, as well as adaptability to changes in workload requirements. Instead of using expensive fixed physical cache hardware, virtualization in Content Delivery Networks (CDNs) enables the implementation of virtual caches (vCaches) on cloud or edge computing systems.

Typically, CDNs have multiple tiers of directed graphs or asymmetric trees that are organized hierarchically. Nodes connected to users make up the lowest tier of CDNs, and nodes connected to origin servers make up the highest tier. The lowest to the top tiers of nodes have caches deployed on a subset of them. User requests access CDNs at the lowest tier, and traffic-delivering user requests are routed through lower tiers using multi-path routing frameworks from cache memory situated at higher-tier nodes. This enables effective resource management and flexibility in response to shifting workload demands.

Many models for CDN application, workload, and infrastructure are proposed to optimize resource usage in these environments. They show how the models can be used to predict traffic patterns and optimize the placement of virtual caches.

○ *2.11 Dynamic security orchestrations in Edge computing*

As Duc et al. [8] and Jalalpour et al. [9] describe, the papers describe a system for secure virtualized dynamic content delivery networks in edge computing environments. The system consists of three main components:

- The Orchestrator, which is responsible for managing security policies and handling security alerts.
- The Virtual Infrastructure Manager (VIM), which manages resources and security function chains.
- The Security Monitoring Analytics System (SMAS), which collects and analyzes data about security chains and alerts the orchestrator if necessary. The system uses various technologies such as Lactive, Docker, network service header, and open virtual switch to implement the above functionalities.

Rate-limiting is a common technique used in Content Delivery Networks (CDNs) to combat network and application layer attacks. Rate-limiting is a technique that limits the number of requests that a server can handle within a certain time period.

There are different types of rate-limiting mechanisms that can be used, depending on the specific requirements of the CDN. These include

Rate-limiting at different layers of the protocol stack (such as the transport or application layer) and Rate-limiting based on different criteria, such as content, end-user, server, and geography.

For example, rate-limiting based on geography can be useful to mitigate Distributed Denial of Service (DDoS) attacks, as it allows the CDN to limit the number of requests from a specific geographic region. Similarly, rate-limiting per content can be useful to prevent a specific resource from being overloaded, while rate-limiting per end-user or server can help to prevent abuse.

Rate-limiting mechanisms can be implemented at the Edge servers of the CDN which are closer to the end-users and can act as a first line of defense against security attacks. The rate-limiting mechanisms can be dynamically adjusted by the Security Orchestrator based on the type of attack and the resources available.

○ *2.12 Machine learning in detecting network security of Edge computing:*

Traditional security methods are not well-suited for edge computing systems, as they are often unable to keep up with the dynamic and distributed nature of these systems. A combination of supervised and unsupervised machine learning algorithms can be used to detect both known and unknown threats.

A combination of network-based and host-based features can be used to train machine-learning models. Network-based features include information about network traffic, such as packets, flows, and protocols. Host-based features include information about the state of the host, such as system calls, process information, and file system activity.

This methodology (not to mention this paper) discusses the use of machine learning in detecting the network security of an edge computing system. Two types of codes are collected from the hardware system, normal and abnormal, which are then converted into feature vectors. Data is pre-processed and divided into three sets: 70% for training, 15% for testing, and the remaining for cross-validation. A supervised learning algorithm called Support Vector Machines (SVM) is used to classify the training code vectors and identify mutations in the code. The SVM algorithm is implemented using the LIBSVM library, which includes kernel functions to address time prediction problems. The system learns to classify the normal code and attack code after this process, ensuring a balanced situation between regular and mutation datasets.

A technique called "fuzzy rule-based reasoning" can be used to improve the accuracy of the models. This technique involves using a set of fuzzy rules, which are a form of logical reasoning that allows for uncertain or imprecise information, to improve the ability of the model to detect complex and nuanced security threats.

● III. SUMMARY

This paper discusses the significance of safeguarding edge servers in Content Delivery Networks (CDNs) to ensure data security and uninterrupted delivery. It defines edge computing and its benefits, such as reduced latency, increased security, reliability, and lower bandwidth requirements. Industrial automation, smart cities, healthcare, autonomous vehicles, video surveillance, augmented and virtual reality, gaming, and more applications use edge computing. CDNs cache material on servers near end users, lowering latency and enhancing performance. To improve data privacy and protect against security threats, effective security measures must be implemented. Additionally, it covers several attack types and their countermeasures in relation to Content Delivery Networks (CDNs). DDoS, side channel, malware injection, cache poisoning, cross-site scripting, brute force attacks, and DNS hijacking are some of the threats. CDNs should establish security best practices and routinely carry out security audits and assessments in order to lessen the impact of these attacks. Additionally covered in the essay are other CDN strategies, such as multi-CDN, hybrid CDN, serverless computing, and CDN-as-a-Service. The article also covers how Mobile Edge Computing (MEC) can be utilized to enhance adaptive HTTP video distribution by lowering latency, enhancing Quality of Experience (QoE), and delivering videos in a more effective and secure manner. It emphasizes the significance of data encryption and dynamic access control mechanisms in maintaining data confidentiality. Edge computing Content Delivery Networks (CDNs) can be optimized with virtual caches and rate-limiting techniques. In order to handle security rules, resources, and security function chains in edge computing settings, dynamic security orchestrations are also required. Machine learning techniques can be employed in edge computing systems to detect both known and undiscovered network security threats utilizing network-based and host-based data.

● IV.CONCLUSION

Edge server security in Content Delivery Networks (CDNs) is critical for data protection and continuous delivery. Edge computing provides various advantages, such as lower latency, increased security, reliability, and fewer bandwidth requirements. However, in order to ensure data privacy and guard against security threats, appropriate security measures must be applied. To limit the impact of attacks, CDNs should create security best practices and conduct security audits and assessments on a regular basis. Other CDN solutions discussed in the study include multi-CDN, hybrid CDN, serverless computing, and CDN-as-a-Service. Furthermore, Mobile Edge Computing (MEC) can be used to improve adaptive HTTP video delivery, underlining the significance of data encryption and dynamic access control techniques. Finally, edge computing platforms can be used to detect both known and unknown network security risks by utilizing dynamic security orchestrations and machine learning approaches.

● REFERENCES

[1] K. Bilal and A. Erbad, "Edge computing for interactive media and video streaming," 2017 Second International Conference on Fog and Mobile Edge Computing (FMEC), 2017, pp. 68-73, doi: 10.1109/FMEC.2017.7946410.

[2] Shalin Parikh, Dharmin Dave, Reema Patel, Nishant Doshi, "Security and Privacy Issues in Cloud, Fog and Edge Computing", Procedia Computer Science,Volume 160,2019,Pages 734-739,ISSN 1877-0509,https://doi.org/10.1016/j.procs.2019.11.018

[3] J. Zhang, B. Chen, Y. Zhao, X. Cheng and F. Hu, "Data Security and Privacy-Preserving in Edge Computing Paradigm: Survey and Open Issues," in IEEE Access, vol. 6, pp. 18209-18237, 2018, doi: 10.1109/ACCESS.2018.2820162.

[4] Y. Xiao, Y. Jia, C. Liu, X. Cheng, J. Yu and W. Lv, "Edge Computing Security: State of the Art and Challenges," in Proceedings of the IEEE, vol. 107, no. 8, pp. 1608-1631, Aug. 2019, doi: 10.1109/JPROC.2019.2918437.

[5] B. Ali, M. A. Gregory and S. Li, "Multi-Access Edge Computing Architecture, Data Security and Privacy: A Review," in IEEE Access, vol. 9, pp. 18706-18721, 2021, doi: 10.1109/ACCESS.2021.3053233.

[6] Aluvalu, Rajanikanth & Muddana, Lakshmi. (2014). A Survey on Access Control Models in Cloud Computing. Advances in Intelligent Systems and Computing. 337. 10.1007/978-3-319-13728-5_73.

[7] Meghanathan, Natarajan. (2013). Review of Access Control Models for Cloud Computing. Computer Science & Information Technology. 3. 77-85. 10.5121/csit.2013.3508.

[8] T. Le Duc and P. -O. Oestberg, "Application, Workload, and Infrastructure Models for Virtualized Content Delivery Networks Deployed in Edge Computing Environments," 2018 27th International Conference on Computer Communication and Networks (ICCCN), 2018, pp. 1-7, doi: 10.1109/ICCCN.2018.8487450.

[9] E. Jalalpour, M. Ghaznavi, D. Migault, S. Preda, M. Pourzandi and R. Boutaba, "Dynamic Security Orchestration for CDN Edge-Servers," 2018 4th IEEE Conference on Network Softwarization and Workshops (NetSoft), 2018, pp. 329-331, doi: 10.1109/NETSOFT.2018.8459970.