

# An Investigational Study and Analysis of Cloud-based Content Delivery Network: Perspectives

Suman Jayakumar

Research Scholar, VTU,  
Belgaum, India

Prakash .S

Department of CSE, Dr. AIT,  
Bengaluru, India

C.B Akki

Department of CSE, IIT Dharwad,  
India

**Abstract**—The content management includes a major technical strategy in the network paradigm of the internet which is called a Content delivery network. The design and the deployment of the CDN shall ensure optimal Quality of services (QoS). This paper aims to brief the taxonomy of the CDN along with its typical architecture. Much latest advancement in smartphones and smart devices which are content hungry require more efficient and reliable mechanism for the cost-effective delivery of the contents irrespective of bottleneck constraints that leads to redesign the entire architecture of CDN on the cloud as CCDN or a new business model of CCDN as a service. The challenges of design for CCDN along with the evolved architecture are discussed in this paper.

**Keywords**—Content delivery network; cloud computing; distribution network; mobility; scalability; distribution

## I. INTRODUCTION

In the era of highly mobile business and competitive environment, any delay in the site/content latency leads to the loss of the business or failure of an intelligent application. The most rapid approach to improvise the application performance is to adopt the Content Delivery Network (CDN) where without having an additional work overhead of updating code or architecture the data can be placed. The evolution approach of (CDN) aims to deliver the content to the end user with optimal Quality of Services (QoS). It is estimated that by 2020 approximately 67% of multimedia content will be forwarded using CDN [1].

CDN is a physically distributed network with surrogate servers and their data centers. The primary objective is to offer high accessibility and high performance to the end-users in order. CDNs provide a huge segment of the Internet content nowadays, containing web things like a) *Graphics* i) text and scripts, b) *Downloadable things* i) Software, ii) Media files, and iii) Documents, c) *Applications* i) E-commerce, ii) Live streaming media, iii) Portals, iv) Social media.

Many of the web objects including text, scripts along with downloadable objects, applications, live streaming media, social media sites, etc. are served through CDNs, and the economics of the CDNs are fulfilled by the content owner [2]. The CDN is the first layer on the internet, and the typical architecture of the CDN includes proxy servers, datacenters, edge servers, origin servers, Domain Name System (DNS) hosted zone, etc. [3]. One of the leading CDN service providers is Akamai, who have deployed tens of thousands of data centers along with edge serves to distribute the content in

different geographic locations across the globe. With the rapid growth of content especially multimedia contents, the demand of CDN is increasing so in the competition the CDNs need to assure a very cost-effective, consistent with fault-tolerant architecture that ensures lower latency. There is a sharp shift into the content type which needs to deal with CDNs as it includes both dynamic and static contents including video, metadata, etc., so the content management processes require establishing correlation among the content. The correlation computation demands an additional module for the machine learning which even helps to tags the popularity of the content. The typical architecture design of CDN considers the content type, its granularity along with the characteristics and the distribution. Some of the significant benefits of the CDN are as follows viz. i) media and advertising: the need of propagating different numbers of real-time multimedia streams demands a dedicated communication channel that can be provided by CDN only. The availability and response could be further increased by migrating CDN services over the cloud. ii) e-Commerce: The improvement of the application is carried out by e-commerce companies in order to offer a better quality of service to the client. This is possible by using CDN where 100% uptime associated with the applications can be carried out. iii) Education: The revolutionary in the area of online learning system demands real-time transaction of various forms of course contents along with forum management where thousands of internal threads are operated. Such massive set of information is made available for ensuring an effective knowledge delivery system by CDN, and iv) business-based application: Any business-based application demands a highly smooth and responsive interaction between the users and the service providers (or vendors/merchants). The faster responsiveness of the business-application is feasibly maintained using CDN. However, apart from the above advantages, there are many limitations as well as constraints too, for which reasons; it is worth investigating this topic.

This paper discusses the different perspective of the existing models and research work towards improving the performance of the CDN system. The organization of the paper is as follow: Section II discusses significant taxonomies of CDN followed by a brief discussion of the significance of CDN in Section III. The evolution of the CDN network is briefly discussed in Section IV followed by highlights of some of the research-based approaches in Section V. Finally, research gap is briefed in Section VI while the conclusive remarks of this paper are briefed in Section VII.

## II. ESSENTIALS OF CDN SYSTEM

The theoretical concept of CDN is quite big enough and there are various taxonomies of it depending on the networking and application demands. It is essential to understand the taxonomies of the CDN that are frequently considered by researchers. For this reason Figure1 highlights the conventional taxonomies of CDN system that are discussed in the work of Stocker et al. [4] which offers an evidence of various forms of CDN system formed by the existing researchers. The different types of CDN are as follows:

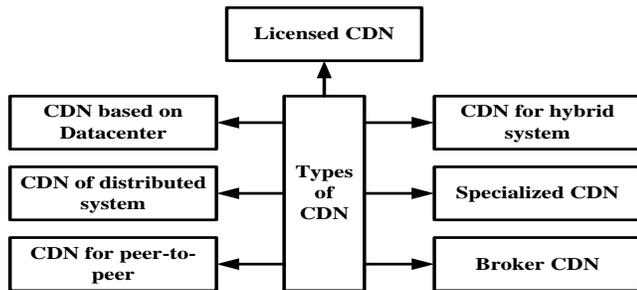


Fig. 1. Types of CDN [4]

By looking into the above Figure1, it is now clear that a different form of networking system demands to utilize different CDN systems. There are CDN systems that are meant for data communication to and from the data center and to the customers using the cloud environment. Peer-to-Peer network also has a dedicated CDN architecture for supporting the exchange of data, which runs completely a different communication protocol. Specialized CDN assists in carrying out a specific on-demand task of communication over better resource availability. Broker CDN assists in explicit communication to achieve a specific performance level using the broker node in the cloud. Finally, hybrid CDN and distributed CDN offers better flexibility to manage an effective data transmission over the distributed network and heterogeneous protocols. The fundamental framework of CDN includes four essential components that can be expressed as an origin server, several surrogate servers, CDN operator and a request redirection mechanism [5]. These components are classified as shown in Figure 2, followed by their discussion:

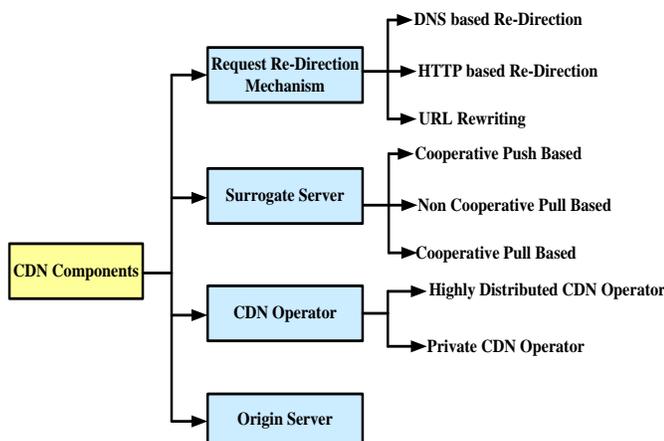


Fig. 2. Classifications of Essential Components in CDN

The essential components in CDN are briefed as follows:

### 1) Origin Server:

The actual contents are reposted over the original server where the data communication is carried out with multimedia providers through a distributed replica server to manage and update a large set of content databases. The replica server plays a role of the media server, cache server or web server. Here, a content provider depends on the CDN operators to globally distribute the contents from the origin server to the user.

### 2) CDN Operator:

The CDN operators are the third party services where the content owner pays to deliver their content to end users and similarly, CDN pays to ISPs and network operator for hosting its servers to their datacenters. Moreover, the CDN operators are run by some branded and commercial organizations. These organizations develop a distributed platform where they provide service for content distribution on rent. The following are the kinds of CDN operators who facilitate services by using different platforms viz. i) highly distributed CDN operator: These CDN operators rent their servers in the data centers of the world's leading Internet service providers, and their networks are highly distributed to most parts of the world. This location is also known as Point of Presence (PoP). The Akamai industry is a good example of a highly distributed CDN operator. ii) Private CDN operators: Private CDN operator established its own data center at a specific region where they present content distribution services over a private network. Also, each location being considered as PoP. A Limelight network is an *example of a private CDN operator*.

### 3) Surrogate Servers:

The CDN operators are the responsible for delivering quality aware content to end user by employing or hosting multiples of edge servers in multi-directional ways, and these servers are popularly known's as surrogate servers. The surrogate servers are proxy servers which are much greater than the origin servers that distribute the content to the user on behalf of origins servers of multiple datacenters. The objective of using surrogate servers is to reduce the network traffic, reduce load and to enhance the availability of the content for both end users and devices. Furthermore, the proxies are categorized into the following which is described as i) *Forward proxy*: These proxies are deployed to capture all the web traffic which is found in the ISP network and ii) *Reverse Proxy*: These proxies are deployed to capture the content request which is addressing to origin servers, and this is implemented as CDN's surrogate servers. Typically, a surrogate server utilizes a caching proxy to store data to provide content to users more frequently. The process involves an HTTP request generated based on a request by the user to access the content, and if the content requested by the user is stored locally in the caching proxy, it is referred to as a HIT, where the user gets the respective content from the caching proxy without accessing from the origin server. In case the caching proxy doesn't retain the requested content, then it is referred to the case of MISS. In this case, the content requested by the user is provided from the origin server, and further it stored locally in the caching proxy for future use. The network interconnection and collaboration between the surrogate servers of CDN's can be implemented using Cooperative push-based [6], Non-Cooperative pull based

[7], Cooperative Pull based [8], Request Redirection Mechanism [9].

### III. EXISTING APPROACHES OF CDN

There are various works being carried out in the area of CDN system in existing research arena. According to reputed research publishers IEEE Xplore, there are 4051 research papers that directly or indirectly deal with discussing the problems and solutions in CDN system. However, a narrowed search shows that there are 480 journals published in last 5 years related to improving the performance of CDN. All these research papers are focused on different types of research problems e.g. placement [10], cost, optimization [11], communication strategy development [12], mobility [13], replica management [14] etc. This section discusses the existing research-based approaches towards improving the performance of CDN. Haghghi et al. [15] have presented a stochastic mechanism to carry out efficient optimization in cloud environment of CDN. They have discussed cloud sites, user sites and cluster graph model for system model and map formulation. In the cluster-graph model formulation, they have mentioned the state space, action space, Rewards and transition probabilities along with feedback based quality of experience. The resource optimization formulation has done by object function dynamic programming, approximate sub-optimal algorithm, unlimited bandwidth & bandwidth algorithm, Limited bandwidth and sort & select algorithm. The performance evaluation of the introduced method by calculating normalized accumulated rewards, time-normalized, and a number of user and improvement gain for software-as-a-Service scheme. Papagianini et al. [16] have presented a modeling of CDN over cloud that performs three steps e.g. service area clustering and inter cluster content distribution graph, inter-cluster content distribution graph over networked cloud environment and replica placement in the network cloud. The performance evaluation was assessed by experimentation set up, different comparative results, relative partitioning cost for 5 and 10 cloud providers and by mapping a number of cloud providers vs. computation time.

Dong et al. [17] have introduced a streaming-based method for constructing a data forwarding architecture to enhance the cloud service provider capacity. The presented study has used game theory using Nash equilibrium with an aid of a unique negotiation-based approach for the purpose of cost modeling of CDN system. Existing approaches has also used the concept of the content-based networks considering the mobility concept of cloud. The study states the importance of content-based management system for data placement considering significant case studies of traffic and resource problems over cloud networks. The study outcome proved its efficiency towards minimizing latency reduction. The problems associated with connectivity of CDN over fog computing was carried out by Mouradian et al. [18]. According to the authors, there are various dependable parameters of fog computing e.g. heterogeneity, QoE management, scalability, mobility, federation, and interoperability that affects the performance of CDN system. Siracusano et al. [19] have presented a re-designing of CDN for dynamic content delivery and effectiveness of the services. The presented architecture was found to support significant level of optimization. The issues

and opportunities like end-user mapping and redirection, network performance discovery and proxy offloading is discussed along with state-of-art, opportunity and envisioned a solution. The implementation and evaluation have done by path selection based on runtime, solution quality and a number of proxies, TCP proxies offloading and internet deployment based on transfer and deployment time.

### IV. EXISTING APPROACHES OF CCDN

With the rise of massive and discrete demands of the data from the user, the service providers of CDN are in consistent threats of competition towards catering up such dynamic needs. The only solution towards it is to depend upon third party a service provider which also offers more dependencies towards its availability over specific geographic location [20]. Therefore, such problems are now overcome using cloud-based CDN (CCDN) system that contributes towards smart placement of data over the distributed cloud storage that is fully capable of offering 24/7 data availability for an user for any part of the world. Apart from this, there are various beneficial points of adopting CCDN that is claimed not only for offering cost effective solution but also certain redefined services towards data delivery system (Figure 3). The benefits of CCDN are to offer value of service as all the models of CCDN offer pay-per-service. Upon hosting CDN services over the cloud, the availability of the services exponentially increases owing to service availability. As cloud supports a better form of interoperability, CCDN offers extensive heterogeneous service deployment even in CDN system too. Figure 3 highlights pictorial depiction of benefits offered by CCDN.

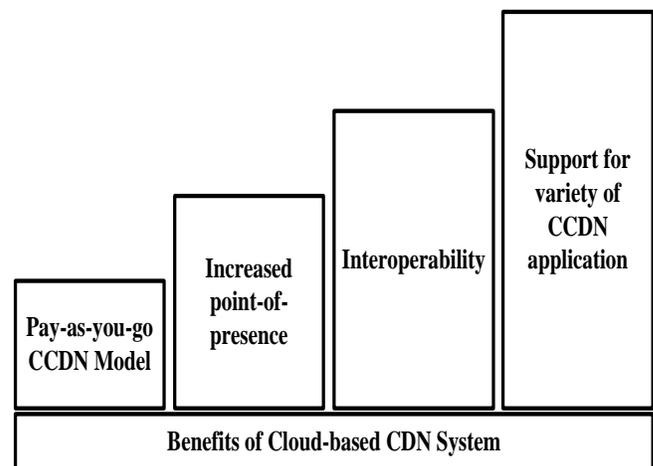


Fig. 3. Benefits of CCDN

Different research approaches towards CCDN can be briefed as following with respect to content placement, fog computing, replica server placement, file management system, mobile edge computing, caching, and content distribution.

#### A. Efficient Content Placement

There have been various studies carried out towards addressing the content placement problem of cloud. The work carried out by Jin et al. [21] has used Topkis-Veinott for addressing the problems associated with inequality in

optimization of location. The study outcome has been found to minimize cost of transport and storage of data. Mathematical approach of Lagrangian relaxation has been adopted by Khalaji and Analoui [22] for the sole purpose of minimizing the complexity issues associated with content placement problems to show the outcome with optimal cost effectiveness. Usage of dynamic programming-based approach for solving similar problem was discussed by Chen and Chen [23] with the study outcome to exhibit low complexity. Adoption of mixed integer linear programming was introduced by Xu et al. [24] as a part of optimization model for solving similar problems. The study outcome has been claimed to offer better capacity of storage. Literature has also witnessed the usage of greedy based approach as well as other search optimization approach e.g. Tabu search as seen on the work carried out by Yang and Huang [25] to prove that it offers an efficient performance of content placement. Apart from this, there are also different set of work being carried out by other researchers e.g. [26] towards similar problems. However, it is still an open end problems in CCDN.

### B. Fog Computing based Approach

Cloud computing has some inherent capability such as scalability, on-demand resource allocation, a flexible pricing model, reduced management, and easy allocation and service provisioning. However, some of the limitations can arrive such as connectivity between the cloud, and the end user or devices, latency due to the overhead by inter-cloud communication and rule may prescribe processing at the location where the cloud provider has no data center. To overcome these limitation fog computing is utilized to provide quality of application. The research work towards the fog system for end-user applications and application specification architectures were found in recent past. Some of the works like Mouradian et al. [18], Stolfo et al. [27] and Wang et al. [28] have discussed end-user application specific architectures for the platform as a service (PaaS), mobile fog and development of IOT applications in Fog respectively. The [18] presents a PaaS architecture to automate applications within a hybrid cloud or fog environment. The mobile fog is a high-level programming model which can be used for future internet needs. The distributed data flow mechanism of [27] gives IoT application which works better in Fog. The application specific architecture for fog system was presented in Stantchev et al. [29], Kapsalis et al. [30] and Datta et al. [31]. The [31] is helpful in healthcare IT professionals to achieve adequate QoS. The three-level architecture of vehicular fog computing is presented [30] for vehicles as infrastructures and yields smart sensor based healthcare system. The consumer specific fog computing architecture that enables roadside units based on IoT services as shown in the Figure 4. The above-mentioned scheme is found more frequently adopted by various researchers in order to introduce the potential of fog computing over CDN. It consists of three different layers or stratum where the top layer is retained for the cloud to host surrogate server. The middle layer holds the fog processing by managing different distributed and synchronized access points. Finally, the end layer connects to edge users.

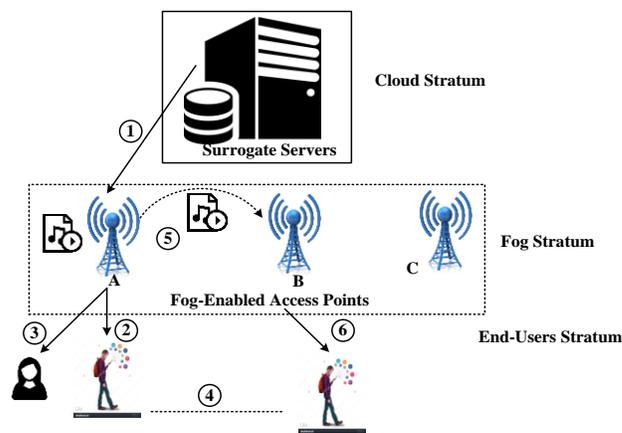


Fig. 4. Schema of Fog Computing using CDN

At present, the approaches of CDN-based research work using fog computing are basically two types, i.e. architecture-based and algorithmic-based. The architecture-based approach has basically focused on developing applications for end-users using agnostic architecture (deals with issues, e.g., Resource management, scheduling application for the end user, etc.) and for exclusive applications (e.g., Applications related to connected vehicular nodes, healthcare, smart appliances, etc.). Similarly, algorithm-based approaches emphasis on storage/distribution, resource consumption, end-user, etc.

### C. Efficient Replica Server Placement

There are issues in the content delivery network while transferring the contents to the end user. To avoid content delivery issues, replica server placement and content placement method in the cloud will achieve the requirements of the user. Hence the replica server placement approach will lead to the fast delivery of content, reduction of data loss and reduce the cost [32]-[35]. Replica server placement algorithms are as shown in below Figure 5.

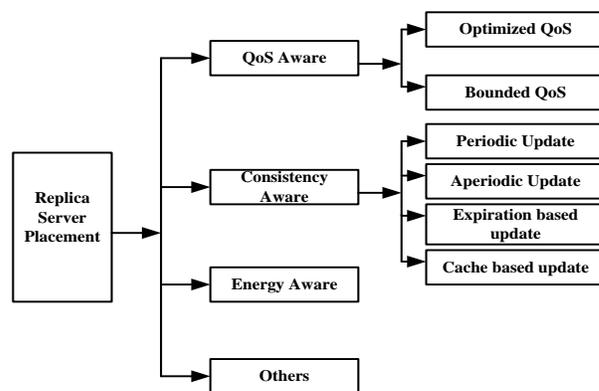


Fig. 5. Replica server placement techniques [32]

### D. Qualitative File Management with Resource Saving

A multimedia or real-time data are usually bigger and uses too much resource for its operation. It is because a typical CDN system fails to manage the video provisioning issues owing to different reasons i) the video clip may be accessed at a time by numerous online subscribers rather than separated by a specific user, ii) variability of multiple resource requirements, and iii) the server selection and video placement are directed

through online as a visitor or client arrive and departs the network. So, the overhead clip management will limit to provide real-time service. Hence file management with resource-saving or Adaptive Data Placement (ADP) system will improve power efficiency and resource in any CDN [35].

**E. Efficient Mobile Edge Computing and Caching**

The smart devices with different new application have increased exponentially in day by day. Due to the heavy burden on the backhaul links, latency, and other reasons, the traditional network cannot easily accommodate user demands. Therefore, mobile edge computation and caching at the edge of the cellular network is adapted in existing literatures [36], which bears the limitation and improve the QoS with high throughput [37-39]. Different communication technique with caching and computing as shown in Figure 6 below:

mm Wave Communication: Data Rate Transmission -High mobility, QoS	D2D Communication: Direct Data Transmission between End Devices -Higher Spectral Efficiency, Low Transmitter Power, Spectrum Reuse, One Hop Communication
Transmission Scheme: Content Centric Transmission, Multicast -Reduce Cost, Balance the Traffic	Interference Management: Reduce Backhaul Load, Overhead, Improve Throughput
Communication Resource Allocation and Scheduling: Efficiency of Caching, Higher Video Distribution Capacity	Synergy of Communication and Caching: Computation Capability, Reduce Transmission Bandwidth

Fig. 6. Communication techniques in edge computing and caching

**F. Content Distribution**

The future generation network (5G) offers promising efficient services for dynamic user requirements. However media entertainments, the IoT industrial devices contribute to the video traffic, which causes operation cost in Internet Service Provider and network load problems [40]. Therefore, Named Data Networking (NDN) is a more efficient form to support content distribution in IoT applications and NDN with network coding technique is also claimed to enhance the IoT throughput and efficiency of content transformations for 5G [40]. There are some of the works have done by Li et al.[41] and Kwak [42] presented 5G information-centric networking which shows the flexibility of mobility as a service in the future 5G network. Similarly, NDN based IoT study with different types of the application has been addressed which gives security of data [43]. The content distribution and NDN network using network coding theory proposed in the work of Zhao et al. [44], which benefits in storage efficiency and energy consumption.

**G. Caching as a Service**

The cellular network cannot efficiently handle the massive delivery of content that is continuously requested by multiple users, as each request is processed as a different End-To-End Connection (E2E) [41]. To exploding traffic demand, CaaS (Caching-as-a-Service) is utilized based on Cloud related Radio Access Network (RAN) with potential techniques. This process

exhibits the capability to cache anything at anytime and anywhere in the cloud-related 5G network system. It leads to improve the user quality of service in different mobile application [41]. Similar kind of work like Hybrid content caching in the 5G network for Efficient Content Exclusion and Replication in work of Kwak et al. [42] and Bilal and Kang [43]. Table 1 below shows the recent CaaS techniques and its performances.

TABLE I. CAAS APPROACHES AND PERFORMANCES

Authors	Solution Approaches, Algorithms	Performance
Zhao et al. [44]	Cluster content caching structure	Gain, cost
Aggarwal et al. [45]	The functional caching approach, a heuristic algorithm	Latency
Hu et al. [46]	Caching and Request Balancing (CRB) algorithm, Differential Provisioning and Caching (DPC) algorithm	Efficiency

Therefore, a closer look into the above outcomes shows that there are various literatures emphasizing on improving CDN performance. There is no doubt that cloud-based CDN is claimed to offer more advantages to the content delivery system; however, still, there are some problems that are found missing in any of the existing research approaches of cloud-based CDN system. The present state of cloud eco-system is already shrouded y security problems as well as performance issues at the same time. This primary problem can for sure offer challenges towards the efficient operation and productivity that can seriously affect the enterprise business. Another challenging problem observed in almost all the approaches discussed under this section is that existing scheme of CDN claims to offer supportability of the static data packet from one to another point, but there is no evidence to prove if it supports real-time transmission of data/service of dynamic origin. Apart from this, there is always a contradiction of a federation-based criterion where the cost factor of service is never found to be considered.

Apart from these, there is also the certain dedicated effort of improving the performance of CDN with a target to obtain a better form of scalability, mobility supportability, working with heterogeneous devices, interoperability, and supportability of federation criteria. However, the summary of Table-2 to Table 4 shows that almost all the studies have been associated with problems and there is no significant literature to offer comprehensive and full-fledged supportability of all such factors. Moreover, the methodology of carrying out the research (which is mainly prototyping) is another reason to narrow down the scope of work where the prominence of various network-related attributes on the cloud is found hypothetically considered in existing studies. These causes were reducing the applicability of existing schemes towards the CDN system that is claimed to be using either cloud or fog computing.

TABLE II. SCHEDULING SCHEMES ON CDN SYSTEM

Authors	Techniques	Advantages	Limitation
Giang et al. [47]	Prototyping, directed graph	Supports QoS	Doesn't support federation-based criteria
Hong et al. [48]	Simulation-based study	Offers diversified service handling	Doesn't offer QoS
Yangui et al. [49]	Prototyping, layer-based architecture	Induces better interoperability	Doesn't ensure supportability of diversified service handling

TABLE III. CDN SCHEMES TOWARDS COMMUNICATION PROBLEM

Authors	Techniques	Advantages	Limitation
Aazam et al. [50]	Prototyping, architecture-based	Offers interoperability	Doesn't address mobility issues
Shi et al.[51]	Prototyping, layer-based communication system	Simplified model	Significantly suffers from interoperability, mobility
Krishnan et al. [52]	Prototyping, architecture-based	Conceptual model	Significantly suffers from interoperability, mobility

TABLE IV. CDN SCHEMES FOR RESOURCE MANAGEMENT

Authors	Techniques	Advantages	Limitation
Cardellini et al. [53]	Prototyping, scheduler design	Comprehensive model	No benchmarking
Bittencourt et al. [54]	Analytical, layer-based architecture	Offers better QoS and mobility	Doesn't offer scalability
Kapsalis et al.[55]	Simulation-based, load balancing	Offers better QoS and mobility	Doesn't offer interoperability

## V. OPEN RESEARCH ISSUES IN CCDN

The prior set of sections has briefed the existing approaches and methods of improving the performance of CDNs. The existing approaches have beneficial points as well as limitations too. Although certain limitations could be overcome, there was no prominent discussion regarding its solution, which is now termed as research gaps. This section discusses the prominent research gaps as follows:

- *Security Prone:* There are very few research approaches that have been emphasizing on the prominent security breaches over CCDN. As exposures of CCDN towards adversaries are exponentially more, offering robust security is still a far away from successful real-time implementation.
- *Placement Strategies in Nascent Stage:* The placement strategy orients around exploring the suitable position of the server where the content, as well as the proxy server, should be residing on. It is quite a challenging task as it completely depends upon the frequency of accessing the specific resources as well as consumer-based unpredictable behavior. At present, the majority of the placement of data is highly static and very less dynamic, although some advancement has happened in the area of virtualization over the cloud.

*Routing among Surrogate-Server:* A surrogate server is also characterized by a certain set of resources and capabilities that are required to be utilized effectively. An efficient surrogate server should be able to identify the direction of routing that should not only save cost but should be highly faster. Unfortunately, existing literature towards CDN has no emphasized on this problem. With the increase of dynamic and complex

traffic behavior, there is a good probability that the decision of routing could go inappropriate that will result in unwanted cost levied by the user.

- *Request Redirection Method:* Proximity metric is one of the essential, dependable parameters to perform request redirection problem. Apart from security routing problem, it is yet a challenging task to take a precise decision of performing redirection of request on a specific set of CDN nodes ensuring lower cost.
- *Caching Techniques Localization of Surrogate-Server:* There are different forms of caching technique that is used in CDN; however, there are still certain problems in it. In the existing literature, the effects of diversified traffic behavior of HTTP header on the caching method while localizing the surrogate server are unexplained. The need for a change of default configuration on the existing CCDN in order to suit the traffic demand is not found to be discussed in existing kinds of literature.

## VI. CONCLUSION

This paper discusses various aspects of CDN where it could be found that it offers a significant advantage to a reliable and dedicated channel of data/service delivery, but it is also associated with a potential set of problems. Although, by migrating CDN over cloud has offered some tangible benefits but the problems have now doubled as cloud-based CDN exposes the CDN system to a large, dynamic, complex, and unsecure data. Since there is less number of studies carried out in the existing system; we focused on such problems and various research gaps in this study. Our next future work shall be to extend this review work further and make it more exhaustive as well as evolve up with some concrete solution to resist such critical research problems in CDN.

REFERENCES

- [1] Q. Jia, R. Xie, T. Huang, J. Liu, and Y. Liu, "The Collaboration for Content Delivery and Network Infrastructures: A Survey," *IEEE Access*, vol. 5, pp. 18088-18106, 2017.
- [2] K. Hosanagar, R. Krishnan, M. Smith and J. Chuang, "Optimal pricing of content delivery network (CDN) services," 37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the, 2004, pp. 10 pp
- [3] Pathan M, Buyya R. "A taxonomy of CDNs: In Content delivery networks," 2008 (pp. 33-77), Springer, Berlin, Heidelberg.
- [4] Stocker, Volker, Georgios Smaragdakis, William Lehr, and Steven Bauer. "The growing complexity of content delivery networks: Challenges and implications for the Internet ecosystem." *Telecommunications Policy* 41, no. 10 (2017): 1003-1016.
- [5] G. Ma, Z. Chen, J. Cao, Z. Guo, Y. Jiang and X. Guo, "A tentative comparison on CDN and NDN," 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC), San Diego, CA, 2014, pp. 2893-2898.
- [6] Pallis, George, and Athena Vakali. "Insight and perspectives for content delivery networks." *Communications of the ACM* 49, no. 1 (2006): 101-106.
- [7] Garmehi, Mehran, Morteza Analoui, Mukaddim Pathan, and Rajkumar Buyya. "An economic replica placement mechanism for streaming content distribution in Hybrid CDN-P2P networks." *Computer Communications* 52 (2014): 60-70.
- [8] Buyya, Rajkumar, Al-Mukaddim Khan Pathan, James Broberg, and Zahir Tari. "A case for peering of content delivery networks." *arXiv preprint cs/0609027* (2006).
- [9] Wang, Limin, Vivek Pai, and Larry Peterson. "The effectiveness of request redirection on CDN robustness." *ACM SIGOPS Operating Systems Review* 36, no. SI (2002): 345-360.
- [10] Hu, Han, Yonggang Wen, Tat-Seng Chua, Zhi Wang, Jian Huang, Wenwu Zhu, and Di Wu. "Community based effective social video contents placement in cloud centric CDN network." In *Multimedia and Expo (ICME), 2014 IEEE International Conference on*, pp. 1-6. IEEE, 2014.
- [11] Raciborski, Nathan F., and Bradley B. Harvell. "Write-cost optimization of CDN storage architecture." U.S. Patent 8,321,521, issued November 27, 2012.
- [12] Kovvali, Surya Kumar, Charles Boyle, Ravi Valmikam, and Krishnan Ramakrishnan. "Hierarchical Device type Recognition, Caching Control & Enhanced CDN communication in a Wireless Mobile Network." U.S. Patent Application 13/183,777, filed July 19, 2012.
- [13] Tariq, Muhammad Mukarram Bin, Ravi Jain, and Toshiro Kawahara. "Mobility aware server selection for mobile streaming multimedia content distribution networks." In *Web content caching and distribution*, pp. 1-18. Springer, Dordrecht, 2004.
- [14] Harahap, Erwin, Janaka Wijekoon, Rajitha Tennekoon, Fumito Yamaguchi, Shinichi Ishida, and Hiroaki Nishi. "Distributed algorithm for router-based management of replica server in next-CDN infrastructure." In *2013 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*, pp. 266-272. IEEE, 2013.
- [15] A. A. Haghighi, S. Shahbazpanahi and S. Shah Heydari, "Stochastic QoE-Aware Optimization in Cloud-Based Content Delivery Networks," in *IEEE Access*, vol. 6, pp. 32662-32672, 2018.
- [16] Papagianni, A. Leivadreas and S. Papavassiliou, "A Cloud-Oriented Content Delivery Network Paradigm: Modeling and Assessment," *IEEE Transactions on Dependable and Secure Computing*, vol. 10, no. 5, pp. 287-300, Sept.-Oct. 2013.
- [17] C. Dong, Y. Jia, H. Peng, X. Yang and W. Wen, "A Novel Distribution Service Policy for Crowdsourced Live Streaming in Cloud Platform," *IEEE Transactions on Network and Service Management*, vol. 15, no. 2, pp. 679-692, June 2018.
- [18] Mouradian, D. Naboulsi, S. Yangui, R. H. Glitho, M. J. Morrow and P. A. Polakos, "A Comprehensive Survey on Fog Computing: State-of-the-Art and Research Challenges," in *IEEE Communications Surveys & Tutorials*, vol. 20, no. 1, pp. 416-464, First quarter 2018.
- [19] Siracusano, Giuseppe, et al. "Re-designing Dynamic Content Delivery in the Light of a Virtualized Infrastructure." *IEEE Journal on Selected Areas in Communications* 35.11 (2017): 2574-2585.
- [20] M. A. Salahuddin, J. Sahoo, R. Glitho, H. Elbiaze and W. Ajib, "A Survey on Content Placement Algorithms for Cloud-Based Content Delivery Networks," in *IEEE Access*, vol. 6, pp. 91-114, 2018.
- [21] Y. Jin, Y. Wen, K. Guan, D. Kilper and Haiyong Xie, "Toward monetary cost-effective content placement in cloud centric media network," 2013 IEEE International Conference on Multimedia and Expo (ICME), San Jose, CA, 2013, pp. 1-6.
- [22] F. K. Khalaji and M. Analoui, "Hybrid CDN-P2P architecture: Replica content Placement Algorithms," *The 5th Conference on Information and Knowledge Technology*, Shiraz, 2013, pp. 7-12.
- [23] J. Chen and C. Chen, "Using Particle Swarm Optimization Algorithm in Multimedia CDN Content Placement," 2012 Fifth International Symposium on Parallel Architectures, Algorithms and Programming, Taipei, 2012, pp. 45-51.
- [24] K. Xu, X. Li, S. K. Bose and G. Shen, "Joint Replica Server Placement, Content Caching, and Request Load Assignment in Content Delivery Networks," in *IEEE Access*, vol. 6, pp. 17968-17981, 2018.
- [25] C. Yang, L. Huang, B. Leng, H. Xu and X. Wang, "Replica placement in content delivery networks with stochastic demands and M/M/1 servers," 2014 IEEE 33rd International Performance Computing and Communications Conference (IPCCC), Austin, TX, 2014, pp. 1-8.
- [26] Papagianni, Chrysa, Aris Leivadreas, and Symeon Papavassiliou. "A cloud-oriented content delivery network paradigm: Modeling and assessment." *IEEE Transactions on Dependable and Secure Computing* 10, no. 5 (2013): 287-300.
- [27] S. J. Stolfo, M. B. Salem and A. D. Keromytis, "Fog Computing: Mitigating Insider Data Theft Attacks in the Cloud," 2012 IEEE Symposium on Security and Privacy Workshops, San Francisco, CA, 2012, pp. 125-128
- [28] Y. Wang, T. Uehara and R. Sasaki, "Fog Computing: Issues and Challenges in Security and Forensics," 2015 IEEE 39th Annual Computer Software and Applications Conference, Taichung, 2015, pp. 53-59.
- [29] M. Mukherjee et al., "Security and Privacy in Fog Computing: Challenges," in *IEEE Access*, vol. 5, pp. 19293-19304, 2017.
- [30] A. Kapsalis, P. Kasnesis, I. S. Venieris, D. I. Kaklamani and C. Z. Patrikakis, "A Cooperative Fog Approach for Effective Workload Balancing," in *IEEE Cloud Computing*, vol. 4, no. 2, pp. 36-45, March-April 2017.
- [31] H. Zhang, Y. Qiu, X. Chu, K. Long and V. C. M. Leung, "Fog Radio Access Networks: Mobility Management, Interference Mitigation, and Resource Optimization," in *IEEE Wireless Communications*, vol. 24, no. 6, pp. 120-127, Dec. 2017.
- [32] Sahoo J, Salahuddin MA, Glitho R, Elbiaze H, Ajib W. A survey on replica server placement algorithms for content delivery networks. *IEEE Communications Surveys & Tutorials*. 2016 Nov;19(2):1002-26.
- [33] J. Sahoo and R. Glitho, "Greedy heuristic for replica server placement in Cloud based Content Delivery Networks," 2016 IEEE Symposium on Computers and Communication (ISCC), Messina, 2016, pp. 302-309.
- [34] Eslami G, Haghighat AT, Farokhi S. New replica server placement strategies using clustering algorithms and SOM neural network in CDNs. *Int. Arab J. Inf. Technol.*. 2017 Mar 1;14(2):260-6.
- [35] Sun G, Chang V, Yang G, Liao D. The cost-efficient deployment of replica servers in virtual content distribution networks for data fusion. *Information Sciences*. 2018 Mar 1;432:495-515.
- [36] Kuo WH, Lin YH. Resource-Saving File Management Scheme for Online Video Provisioning on Content Delivery Networks. *IEEE Transactions on Computers*. 2016 Jun 1;65(6):1910-20
- [37] Wang S, Zhang X, Zhang Y, Wang L, Yang J, Wang W. A survey on mobile edge networks: Convergence of computing, caching and communications. *IEEE Access*. 2017;5:6757-79.
- [38] G. Ma, Z. Wang, M. Zhang, J. Ye, M. Chen and W. Zhu, "Understanding Performance of Edge Content Caching for Mobile

- Video Streaming," in *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1076-1089, May 2017.
- [39] T. Taleb, S. Dutta, A. Ksentini, M. Iqbal and H. Flinck, "Mobile Edge Computing Potential in Making Cities Smarter," in *IEEE Communications Magazine*, vol. 55, no. 3, pp. 38-43, March 2017.
- [40] E. Baccelli, C. Mehlis and O. Hahm. "Information Centric Networking in the IoT: Experiments with NDN in the Wild," *ACM Conference on Information-Centric Networking*, pp. 77-86, 2014.
- [41] X. Li, X. Wang, K. Li and V. C. M. Leung, "CaaS: Caching as a Service for 5G Networks," in *IEEE Access*, vol. 5, pp. 5982-5993, 2017.
- [42] J. Kwak, Y. Kim, L. B. Le and S. Chong, "Hybrid Content Caching in 5G Wireless Networks: Cloud Versus Edge Caching," in *IEEE Transactions on Wireless Communications*, vol. 17, no. 5, pp. 3030-3045, May 2018.
- [43] M. Bilal and S. Kang, "A Cache Management Scheme for Efficient Content Eviction and Replication in Cache Networks," in *IEEE Access*, vol. 5, pp. 1692-1701, 2017.
- [44] Z. Zhao, M. Peng, Z. Ding, W. Wang and H. V. Poor, "Cluster Content Caching: An Energy-Efficient Approach to Improve Quality of Service in Cloud Radio Access Networks," in *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1207-1221, May 2016.
- [45] V. Aggarwal, Y. R. Chen, T. Lan and Y. Xiang, "Sprout: A Functional Caching Approach to Minimize Service Latency in Erasure-Coded Storage," in *IEEE/ACM Transactions on Networking*, vol. 25, no. 6, pp. 3683-3694, Dec. 2017.
- [46] M. Hu, J. Luo, Y. Wang and B. Veeravalli, "Practical Resource Provisioning and Caching with Dynamic Resilience for Cloud-Based Content Distribution Networks," in *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 8, pp. 2169-2179, Aug. 2014.
- [47] N. K. Giang, M. Blackstock, R. Lea, and V. C. M. Leung, "Developing IoT applications in the Fog: A Distributed Dataflow approach," in *Internet of Things (IOT)*, 2015 5th International Conference on the, 2015, pp. 155-162.
- [48] K. Hong, D. Lillethun, U. Ramachandran, B. Ottenwalder, and B. Koldehofe, "Mobile Fog: A Programming Model for Large-scale Applications on the Internet of Things," in *Proceedings of the Second ACM SIGCOMM Workshop on Mobile Cloud Computing*, New York, NY, USA, 2013, pp. 15-20.
- [49] S. Yangui, P. Ravindran, O. Bibani, R. H. Glitho, N. B. Hadj-Alouane, M. J. Morrow, and P. A. Polakos, "A platform as-a-service for hybrid cloud/fog environments," in *2016 IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN)*, 2016, pp. 1-7.
- [50] M. Aazam and E.-N. Huh, "Fog Computing and Smart Gateway Based Communication for Cloud of Things," in *2014 International Conference on Future Internet of Things and Cloud (FiCloud)*, 2014, pp. 464-470.
- [51] H. Shi, N. Chen, and R. Deters, "Combining Mobile and Fog Computing: Using CoAP to Link Mobile Device Clouds with Fog Computing," in *2015 IEEE International Conference on Data Science and Data Intensive Systems*, 2015, pp. 564-571.
- [52] Y. N. Krishnan, C. N. Bhagwat, and A. P. Utpat, "Fog computing - Network based cloud computing," in *2015 2nd International Conference on Electronics and Communication Systems (ICECS)*, 2015, pp. 250-251.
- [53] V. Cardellini, V. Grassi, F. L. Presti, and M. Nardelli, "On QoS-aware scheduling of data stream applications over fog computing infrastructures," in *2015 IEEE Symposium on Computers and Communication (ISCC)*, 2015, pp. 271-276.
- [54] L. F. Bittencourt, M. M. Lopes, I. Petri, and O. F. Rana, "Towards Virtual Machine Migration in Fog Computing," in *2015 10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC)*, 2015, pp. 1-8.
- [55] A. Kapsalis, P. Kasnesis, I. S. Venieris, D. I. Kaklamani, and C. Z. Patrikakis, "A Cooperative Fog Approach for Effective Workload Balancing," *IEEE Cloud Comput.*, vol. 4, no. 2, pp. 36-45, Mar. 2017.