International Journal of Electrical and Computer Engineering (IJECE)
Vol. 13, No. 5, October 2023, pp. 5588~5598
ISSN: 2088-8708, DOI: 10.11591/ijece.v13i5.pp5588-5598

A novel cost-based replica server placement for optimal service quality in cloud-based content delivery network

Priyanka Dharmapal[1], Channakrishnaraju[2], Chethan Bommalingaiahanapalya Krishnamurthy[3]
[1]Department of Computer Science and Engineering, Sri Siddhartha Academy of Higher Education University, Karnataka, India
[2]Department of Computer Science and Engineering, Sri Siddhartha Institute of Technology, Karnataka, India
[3]Department of Artificial Intelligence and Machine Learning, Vidyavardhaka College of Engineering, Visveswaraiah Technological University, Karnataka, India

Article Info

Article history:

Received Dec 10, 2022
Revised Apr 4, 2023
Accepted Apr 7, 2023

Keywords:

Cloud
Content delivery network
Cost
Replica server
Resource

ABSTRACT

Replica server placement is one of the crucial concerns for a given geographic diversity associated with placement problems in content delivery network (CDN). After reviewing the existing literatures, it is noted that studies are more for solving placement problem in conventional CDN and not much over cloud-based CDN architectures, which some few studies are reported towards replica selection are much in its nascent stages of development. Moreover, such models are not benchmarked or practically assessed to prove its effectiveness. Hence, the proposed study introduces a novel design of computational framework associated with cloud-based CDN which can facilitate cost-effective replica server management for enhanced service delivery. Implemented using analytical research methodology, the simulated study outcome shows that proposed scheme offers reduced cost, reduced resource dependencies, reduced latency, and faster processing time in contrast to existing models of replica server placement.

Corresponding Author:

Priyanka Dharmapal
Research Scholar, Department of Computer Science and Engineering, Sri Siddhartha Academy of Higher Education University
Tumkur, Karnataka-572017, India
Email: prsh2019@gmail.com

1. INTRODUCTION

With the rising demands of data and service availability, there is a need of an effective networking technology that assists in the offering the seamless services irrespective of traffic condition in dedicated manner. A content delivery network (CDN) is one such technology that assists in delivering the content efficiently and quickly using surrogate server and content placement [1]–[5]. Although, there are several reported advantage of adopting CDN system towards facilitating rate of data delivery, it is characterized by various challenging condition too [6]. The primary challenge in deployment of CDN is its inclusion of large sum of money with highly sophisticated process of deployment [7]. On the basis of multiple studies, it has been noted that enabling technologies of CDN in current times make use of cloud computing more [8]. This adoption offers beneficial characteristic of service provisioning, resource allocation, pay-as-per usage, and hosting service supportability [9].

It should be noted that in the complete operation of CDN, the role of replica server is highly important which is actually responsible for replicating the original content to content replica server. Irrespective of various approaches being carried out towards content placement algorithm, there are still a large trade-off between the demands of the user and the present state of operation of replica server. The primary issue associated with the placement of replica server is associated with the optimization problem, which could be either constraint or unconstrained type. The deployment cost of the replica server as well as its associated delivery cost and update cost is not much emphasized upon in existing system with respect to network topology, metric of network performance, latency, and count of hop. Apart from this, there are various theoretical models presented in existing studies which has presented discussion about the facility location with both capacitated and uncapacitated variant considering constraints as capacity of server. Existing studies have also used K-median, minimum k-center, and k-cache location. for the purpose of localization of replica server [10]. A thorough review of existing scheme also shows that majority of the approaches associated with replica server placement is carried out on the basis of conventional CDN and not towards cloud-based CDN. The prime difference between these two strategies is that conventional CDN are bit centralized while cloud-based CDN are highly distributed and still it could be centralized to formulate a large chain of network. It will mean that there are various inevitable challenges to be encountered when a conventional CDN is exposed to challenges of heavier traffic condition as well as uncertain behavior of various service and application running over it. In this entire scenario, it is also noted that cloud-based CDN system also offers a less emphasis towards harnessing virtualized environment for increasing the coverage of replica server. Although, it is nearly impossible to position the content server in multiple geographical places; however, the cost can be significantly minimized when the existing virtualized environment (or machine) can be classified into local and global form in order to offer connectivity in both smaller and larger scale. Hence, it is essential to carry out an investigation in this direction in order to understand the impact of using local virtual machine as well as proxies in order to facilitate better form of networking in cloud-based CDN system. Further, it is also essential to understand the impact of spatial attributes towards the placement of the replica server in cloud-based CDN and its possible impact in large scale network. Hence, in order to explore these possibilities, the proposed study also carry out a deeper investigation of current literatures where it is found that studies are more towards content placement and less towards replica server placement deploying the virtualized environment.

The discussion of the relevant literature is as: our prior review has discussed about different approaches and techniques required for improving the performance of CDN [11]. At present, there are various further attempts that directly or indirectly contributes towards leveraging the performance of CDN. The work carried out by Al-Abbasi *et al.* [12] have presented a probability-based model towards controlling the stalling event during streaming services in CDN. A model is developed on the basis of restricted spaces in caches along with allocation of CDN for better streaming delivery. Further, the work carried out by Chuan *et al.* [13] have presented an optimized model for content placement considering belief propagation system over cache in distributed network. The complete model is developed in a sequential flow where a selection mechanism of content helper is carried out followed by caching the contents and delivering them at proper destination. On the basis of states of communication channel and popularity, the modelling is carried out in iterative manner emphasizing towards controlling energy consumption in wireless network. The work carried out by Fan *et al.* [14] have presented an allocation model for replica server which is claimed to be more energy efficient as well as reliable. The study model contributes to concurrent task processing over the main server mapped with a dedicated storage point. The work carried out by Guerrero *et al.* [15] have developed a model for data replication with a core intention towards leveraging data availability over a sophisticated form of weighted network along with centrality factors. The study model also evaluates a graph partitioned attributed while the data replicates are stored in fog devices. According to this model, one unit stores fog information and a single replica of data while the other stores all file replica using greedy based approach.

The work carried out by Yovita and Syambas [16] have discussed about the importance of the caching mechanism which is highly essential for large distributed network. According to the study outcome, there are yet unsolved problems towards caching mechanism. Kusuma *et al.* [17] have emphasized towards the vertex markers in order to discuss possible scale of amendments towards grid methods. Liu *et al.* [18] have presented a unique replica placement scheme that uses the existing software framework for Hadoop in order to address the problems of large-scale voluminous data aggregated during data exchange. The study considers three-dimensional raster data in order to perform better determination of location associated with replica server placement while the study outcome show efficient performance with reduced network overhead. Adoption of machine learning is also witnessed in existing scheme towards solving the problem of selection of replica. This cadre of work is reported by Mostafa *et al.* [19] where a large scale environment is considered by deploying artificial neural network (ANN) for profiling the behavior of the location involved in the process. The outcome exhibits satisfactory predictive accuracy with sufficient optimization of channel capacity. The study towards cost estimation is carried out by Nazir *et al.* [20] while performing job scheduling over grid system. The work introduces a unique dynamic scheduling policy that is centralized towards replica placement. The prime intention of this model is to reduce the cost of replica placement by scheduling the data considering computing capacity of the node in order to explore the task to be processed. A case study of telco CDN is reported in work of Safavi *et al.* [21] where an on-line learning scheme towards replica placement is discussed considering the popularity of the contents. The spatial patterns are computed that is associated with the request of the content in order to perform predictive modelling on the basis of content popularity. As discussed by various work model, it was noted that content caching is one of the essential operations for leveraging the perform of CDN. Study in such direction was carried out by Saino *et al.* [22] focusing on both conventional and futuristic form of CDN where the authors contributes towards introducing a unique placement of contents along with higher degree of flexibility towards placement operation. The study outcome is exhibited to offer improved load balancing performance and being less sensitive toward any form of traffic fluctuation.

The study of content placement is also carried out over 5G network services as noted in work of Santos *et al.* [23] where the edge sites are targeted for placing the virtualized environment of CDN. The core objective of the model is towards splitting the heavier multimedia file followed by localizing those splitted files over dynamic and virtualized resources. Nearly similar form of approach is also carried out by Shankar and Chitra [24] using integrated machine learning process. According to this model, the latency is addressed by distributed storage of data followed by learning the placement coordinates using support vector machine (SVM) and radial basis function (RBF). This learning scheme is capable of classifying data centers along with prediction of possible traffic load by using either cloud or edge resources. Further, the optimization is carried out using graph partitioning in dynamic approach in order to cater up the latency demands with reduced cost of placement. Shao *et al.* [25] have carried out an investigation towards various processes of selection of replica followed by techniques towards itself placement in edge environment as well as in internet of things (IoT) environment. The review outcome stated the importance of provenance of data in order to improve accuracies. The author also stated the ongoing challenges of replica placement is quite complex to deal with especially in presence of dynamic CDN environment. Teng *et al.* [26] have presented a scheme of content placement considering the problem of delay minimization by harnessing biconvexity.

The work carried out by Tran *et al.* [27] have presented a framework using software defined network on CDN which primarily emphasizes on selection of server followed by constructing an intelligent controller system which is primarily meant for resisting all possible overloading condition in SDN. The work carried out by Xiong *et al.* [28] have presented a strategy for replication management in order to mitigate the delay during access considering data with spatial and temporal characteristics on the basis of caching system. According to the implementation model, the maximized popularity score is mined using correlational factor associated with an access of user by adopting beneficial aspect of location. It also uses all connected files obtained from the information of user access. This process finally yields replicates followed by selection of cache node in order to fulfil the purpose of placement. The work carried out by Xu *et al.* [29] have presented a scheme have presented an optimization framework using mixed integer linear programming where the complete model initiates by optimal selection of replica server followed by undertaking decision towards caching the content items within the replica server. Finally, all the varied servers are allocated with the loads of content request followed by constructing a heuristic method. Another unique work has been presented by Yu and Pan [30] where the problems associated with conventional hashing method for distributed data storage in Hadoop is addressed. According to this study model, a hypergraph-based methodology for solving the data placement has been discussed that takes into account varied metrics of performance connected with the location information of the target data in cumulative fashion. The study model uses this method in order to partition the selected dataset items and relocate them in different location of distributed storage units. The study is also claimed to be capable enough for exploring the presence of redundant data.

After reviewing the above-mentioned existing studies, the conclusive statement of problem is as: it is computationally challenging task to develop a generalized scheme of replica node placement over distributed environment of cloud-based CDN. The prime reason behind this statement are as: i) majority of the existing studies have emphasized on quality of service (QoS) attribute, which is essential; however, consideration of user's experience is also equally important which is not much reported to be emphasized, ii) existing approaches has deployed a mechanism considering varied forms of files; however, such schemes are not much effective when deployed under a situation of non-uniform distance between content server and replica server; and iii) further, non-utilization of virtualized environment is less seen in the existing scheme.

The proposed solution presented in this paper discusses about a unique replica server placement by harnessing the data availability potential of local virtualized machined and proxies connected to centralized content server that is further maintained in a distributed form. The core idea of this work is mainly towards offers similar user experience quality as well as service quality in presence of dynamic traffic condition in cost effective manner.

The value added in proposed scheme are as: i) the proposed scheme presents a cost-based optimal replica server placement strategy which equally contributes towards task allocation unlike existing schemes; ii) adoption of non-iterative scheme towards flow processing considering proximity importance attribute among all the nodes connection in large network; and iii) faster processing time to process the query irrespective of any form of traffic. The next section discusses about adopted method.

The organization of the proposed manuscript is as: section 1 discusses introduction while the section 2 discusses about methodology being undertaken, section 3 discusses about the result being accomplished while section 4 discusses about the conclusion of the paper.

2. METHOD

The prime aim of this stage of study is to develop a novel model that can perform provisioning of resources to ensure better replica server placement in dynamic distributed cloud-based CDN system. Adopting analytical research methodology, the proposed system initiates its design using communication model considering multi-cloud system as shown in Figure 1. For simplifying the topology design, the proposed system implements a tree structure that are highly clustered. According to this tree structure, each user is served by a dedicated cloud system using surrogate server that are further incorporated with operations associated with virtual machine, storage, and bandwidth. The next part of implementation is associated with developing a cloud-based CDN resource management system that will be further classified into planning stage and reallocation of user stage. The proposed system will use the cost model for assessing the cost incurred for resource allocation in present scenario. In the next stage of user reallocation, three set of operations are carried out toward resource provisioning considering formulation of practical constraints and achieving highest optimality in replica server placement. Finally, an objective function will be designed which can balance the resource provisioning demands with dynamic content delivery system for effective replica server placement in cloud-based CDN system.

This part of implementation is anticipated to achieve optimal performance in cloud-based CDN. A test environment will be constructed to assess the influence of various network attributes e.g., channel capacity, cost, and uncertain traffic situation on the model. The model is also anticipated to exhibited lower computational complexity. The next section discusses about the research methodology adopted for the purpose of implementation briefed with respect to system design and algorithm implementation.

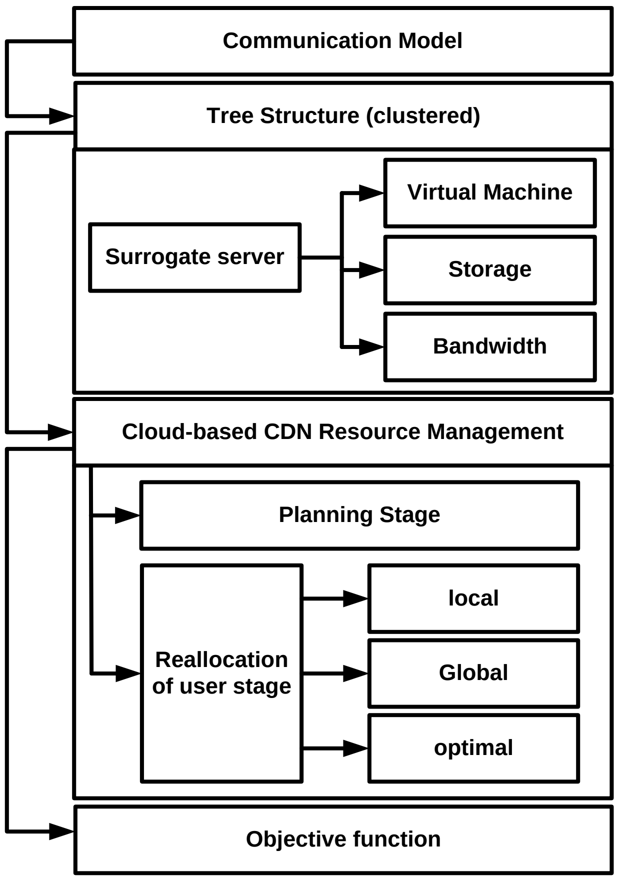

Figure 1. Architecture of 2nd module of implementation

2.1. System design

The core system design of the proposed scheme is based on the location and proximity of the replica server and the proxy cache in order to offer better availability of service via cloud networks. According to this scheme, a centralized content server (CCS) is considered which is positioned at uniform proximity of multiple local virtual machine (LVM) which retains information about replica server and proxy cache.

The Figure 2 highlights the placement strategy of replica server where i^{th} number of multiple observation zone O_z exists i.e., $O_{zi} = [O_{z1}, O_{z2}, \dots O_{zi}]$ where the placement of LVM is carried out. It will mean that each O_z consist of specific number of LVM (depending upon the geographic spread of O_z). Mathematically, it can be expressed as (1).

$$LVM_i = [P_i, R_i]^{\alpha\beta} \tag{1}$$

The (1) shows i^{th} number of LVM is positioned in O_{zi} and each LVM_i consists of proxy cache data P_i and replica server R_i considering α as total proxy server and specific bandwidth β associated with replica server. Each LVM are connected to each other by one-to-many relation in unique fashion. In order to simplify this process, the proposed scheme allocate a link coefficient γ for each of the possible connection from each LVM as shown in (2).

$$\gamma = \sum_{j=1}^{k} LVM_{j/l} \tag{2}$$

In the mathematical expression (2), the variable j and l will represent specific number of LVM and remaining number of individual LVM to be connected with current LVM. It should be noted that this is only possible when the proposed scheme is designed using tree-structure. The contribution of link coefficient is that it assists in formulating LVM connectivity with respect to reduced cost involved in data transmission over the assigned tree structure. One of the contributions of the proposed scheme is its placement strategy which is circular in its orientation that targets to offer a uniform performance of the data delivery services from the CCS. However, it should be noted that multiple CCS joins in the tree structure to formulate highly distributed and yet well-connected to each other targeting persistent quality of service delivery. Another contribution in this scheme is uniform utilization of bandwidth. It will eventually mean that each node of replica server in tree structure is characterized by unique orientation in the form of in-degree and out-degree links where each node bears information associated with service provider of CDN discretely defined for each direction in distributed manner. Another best part is its capability to identify the presence of any form of redundant data that may possibly reside in cache proxy.

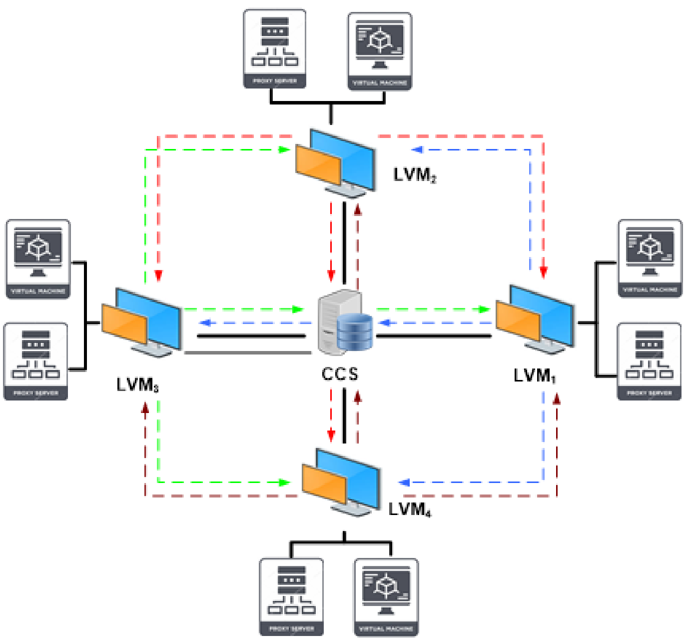

Figure 2. Scheme of replica server placement

2.2. Algorithm implementation

The proposed system introduces a very simple and yet novel algorithm which uses cost estimation in order to perform placement of replica server. It is to be noted that while performing the placement of replica server, the importance is given more towards LVM as it is the bridge between the replica server and the actual content server connected in distributed manner. The steps of proposed algorithm are as:

Algorithm 1. For cost-based replica server placement

Input: n_o, n_e, LVM, j, S
Output: Φ
Start
1. init net[n_o, n_e(LVM_j), S]
2. G→(n_o, n_e)^S
3. $\theta = \mu(G)$
4. formulate matrix $G=[\theta, I_L, O_L]$
5. Extract new tree=ω(G)
6. L_t=(G, cond($prob$))
7. $\Phi=\rho(\theta)$

End

The algorithm 1 takes the input of origin node n_o, end node n_e, local virtual machine *LVM*, number of LVM *j*, and service being hosted by the CDN *S* that gives the resultant outcome of cost Φ. The preliminary set of tasks of the proposed algorithm is to initialize specific number (i.e., j) of LVM considering the structure of its placement and connectivity with origin and end node (Line-1). The next part of implementation is associated with construction of graphical tree G for mapping the complete nodes and links in the form of a network topology in distributed and large-scale order (Line-2). It is to be noted that proposed scheme constructs distributed scheme of replicate server placement by connecting LVMs associated with centrally located CCS (Line-3). The computation of the central location is carried out by obtaining the score of each LVM that is divided with respect to available least quantity of nodes which has essential characteristics of edge information linking with the respective LVM. The outcome of this operation gives the score of importance attribute θ using a specific function μ doing the above-mentioned task by taking the input argument of G (Line-3). After computing the importance attribute, the algorithm formulates a two-dimensional matrix G that retains information associated with θ importance attribute, incoming link I_L, and outgoing link O_L (Line-4). This mechanism is deployed in order to construct a distributed network links connecting all the LVMs in different location. This assessment is performed for the purpose of assessing the consistency of the proposed algorithm in order to acquire reduced delay considering multiple position of the LVM associated with the replica server in specific O_z. After this task is accomplished, the algorithm deploys another function ω that constructs a diagraph tree structure G over the existing tree (Line-5). The system than generates a local tree L_t considering the newly constructed tree G obtained in prior step followed by selection of the LVM with an embedded statistical condition *cond* (Line-6). According to this condition *cond*, the construction of the L_t is formed only when its probability value *prob* is found to be statistically significant. In the last step, an explicit function ρ is formed that performs final estimation of the cost Φ associated with the placement of replica server at the time of data transmission (Line-7). The algorithm considers importance attribute θ considering the LVM as well as their location-specific data during this mechanism.

The complete operation of the cost function ρ is stated as:

a) The cost function ρ constructs a two-dimensional matrix of ($d_1 \times d_2$) where each elements of the matrix represents the cost attributes that is allocated with specific d_1 number of LVM with d_2 storing all the discrete jobs that it has to process. The construction of this matrix is carried out in specific form to retain least number of d_1 and d_2 in the form of rows and columns where allocation of bandwidth β is defined as $\beta = minimum(d_1, d_2)$.
b) The consecutive process is to look for least number of matrix element in complete d_1 rows followed by subtracting it from residual elements residing in the rows.
c) All the elements in the matrix are then looked for presence of any element with value zero. The system than flags all the element which is found to be non-zero elements in the matrix. Similar process is repeated for all the elements residing within the matrix.
d) The algorithm encapsulates all the flagged zero elements residing within the matrix in the form of columnar position. If the system finds β_{tot} number of columnar elements that are flagged with the zero elements than it represents this as a total allocation of bandwidth in unique form. This is the usual scenario of termination of total operation of an algorithm, otherwise it is resumed for next level of processes.
e) The system identified all the non-encapsulated zero elements followed by transforming into prime numbers. If further there is no flagged zero elements over the rows in matrix, the function proceeds to next step otherwise it encapsulates the specific elements maintained in row while it does not encapsulate

any columnar elements with elements flagged with zero. This process is continued until the function encapsulates all the zero elements followed by saving all the minimal value of non-encapsulated elements in matrix.
f) The function then constructs all the primed elements with zeros as well as flagged elements with zero.
g) The outcome of this process is yielded in 5th step and this outcome is appended complete elements in the matrix in the form of encapsulated form maintained over rows that is subtracted from all the non-encapsulated elements in columnar form. The function further performs operation highlighted in 5th step above without any need of changing flagging or primes.
h) Therefore, the function yields a set of scores that is required to be assigned to the node taking into account for location of all the elements with zero and flagging over the defined cost matrix. It will eventually mean that if matrix of dimension $d_1 \times d_2$ is found to be flagged as zero than all the scores associated with the respect to d_1 row will be assigned to scores that are associated with d_2 column. Therefore, the resultant outcome of the last function ρ is considered to be an optimal result with significantly least cost of placing replica server. The next section discusses about the accomplished outcome.

3. RESULTS AND DISCUSSION

From the prior section, it was noted that proposed scheme introduces a novel mechanism towards replica server placement in order to ascertain better performance delivery in cloud-based distributed CDN system. Hence, it is essential to chalk out a definitive assessment strategy in order to ensure better data transmission performance with respect to variable test environment in CDN. The algorithm of the proposed system is scripted in MATLAB emphasing towards accomplishing the tree structure deployment where cost for replica server placement is the prime analysis factor towards result analysis. Table 1 discusses about the evaluated outcomes of cost that is linked with individual LVM residing over multiple orientation considering central location of CCS. The assessment of proposed scheme is carried out by constructing a bipartite graph.

Table 1. Simulation parameters

Simulation parameter	Values assigned
Total deployment areas of LVM	9
Total replica server	10
Bandwidth assigned for replica server	5,000 Mbps
Total proxy (cache)	25
Bandwidth assigned for proxy	3,000 Mbps

Deploying the simulation parameters and its respective values exhibited in Table 1, the proposed system carries out cost estimation for individual LVM in 9 different circular position considering CCS at its center. The complete analysis is carried out over variable size of random data packets of 2,500 bytes over a simulation area of size 1,000×1,000 m^2 considering presence of 500 nodes. The assessment is carried out considering randomly selected origin node and end node. The individual outcomes of cost in 9 different locations are shown in Table 2 assessed over 50 iterations.

Table 2. Accomplished cost (Φ)

Origin	Estimated cost for nodal location of 9 LVMs								
	1	2	3	4	5	6	7	8	9
LVM_2	71	73	28	65	23	93	5	25	56
LVM_4	65	98	14	97	68	1	19	93	44
LVM_1	4	54	29	25	85	47	73	28	65
LVM_3	8	34	45	69	35	43	48	78	66
$LVM_2 \rightarrow LMV_1$	33	12	54	30	79	47	16	20	69
$LVM_2 \rightarrow LVM_3$	54	62	47	68	69	78	35	30	65
$LVM_4 \rightarrow LVM_1$	66	79	89	71	2	33	62	10	96
$LVM_4 \rightarrow LVM_3$	42	43	53	8	61	79	20	59	22
CCS	83	10	95	26	40	48	75	69	72

In order to assess the sustainability of the network, the analysis is carried out by further increasing the size of packet to 3,000 bytes where observation being carried out for 500 iterations in order to arrive at final averaged outcome of estimated cost of replica server placement as exhibited in Table 3. Comparing the difference of trends in cost in Tables 2 and 3, it can be seen the cost values get significantly reduced over a period of time with an inclusion of more number of traffic and more iteration compared to less number of traffic in CDN. The prime reason behind this is topology gets more branched with increasing traffic while allocation of resources is also distributive carried out which splits up the computational burden of each LVM using tree structure. Hence, adoption of tree structure can be seen as one contributory points towards minimizing the cost of replica server placement in cloud-based CDN system.

Table 3. Estimated cost

Location	LVM No	Φ
LVM_2	7	4
LVM_4	6	0
LVM_1	3	28
LVM_3	1	7
$LVM_2 \rightarrow LMV_1$	2	11
$LVM_2 \rightarrow LVM_3$	8	29
$LVM_4 \rightarrow LVM_1$	5	1
$LVM_4 \rightarrow LVM_3$	9	21
CCS	4	25

Figure 3 highlights the outcome associated with the estimated cost of replica server placement where a gain of approximately 45.2% is accomplished in contrast to existing schemes of QoS aware [31] and consistency-aware algorithms [31] that is frequently adopted in existing replica server placement. The existing scheme of QoS aware approach is noted to carry out pre-defined computation where the system ascertains its network topology in order to accomplish a known target QoS via LVM for processing generated query. This static characteristic of QoS-aware approach is therefore witnesses with discrepancies when exposed to increasing traffic that further increases the cost of replica server placement allocation. The consistency-aware algorithm is found to offer slightly efficient performance in comparison to QoS-aware approach.

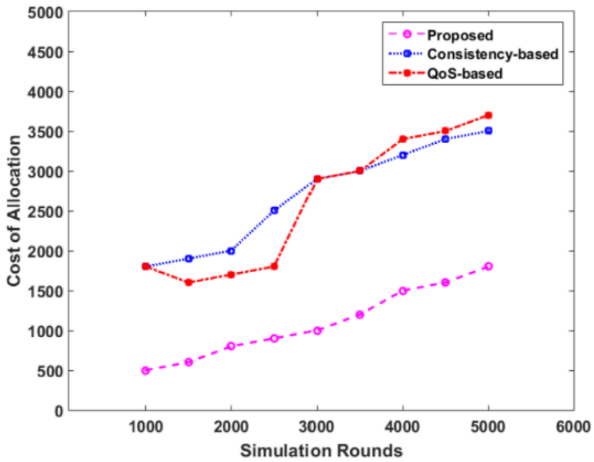

Figure 3. Comparative evaluation of cost of replica server placement

However, it was noted that such approach increasingly demands more computational effort in order to find an optimal condition which is not permitted by the nature of consistency-aware algorithm. It has to settle with the defined network attributes and hence fails to address the dynamic demands of traffic causing extra cost involved in replica server placement. On the other hand, proposed scheme is found to offer better capability of reducing the operational cost in presence of increasing traffic as it can handle multiple tasks using its symmetrical form of placement of both distributed CCS aligned with multiple LVM. Hence, the nearly uniform computational effort is used without introducing any form of staticness in the nature of the topology-building process. Further, the computed cost is also subjected to updating process with every event of topological alteration, which directly assists allocating the cost for other communicating nodes. Hence, proposed scheme offers reduced cost for replica server placement in cloud CDN environment.

Figure 4 showcase approximately 37.4% of improvement of better resource allocation compared to existing schemes. The prime reason behind this is the operation carried out by proposed scheme toward each incoming and outgoing traffic by maintaining equal balance between normal and priority task involved for every serviced hosted for analysis. This argument also contributes towards the reduction of latency as well as processing time as exhibited in Figures 5 and 6. The reduction in latency is found to be approximately 67.5% while the reduction processing time is approximately 37.2%. The conclusive remark of these outcome can be stated to be nearly similar pattern of outcomes associated with existing system which emphasizes more towards solving problem space of local networks whereas proposed scheme initiated with a large interconnected distributed network where each link is characterized by a weight while the nodes are allocated with cost factors. This assists in better deployment of decision making while performing data delivery in cloud-based CDN.

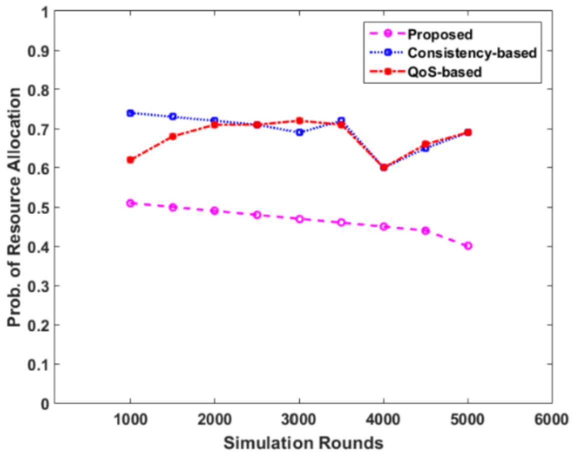

Figure 4. Comparative evaluation of probability of resource allocation

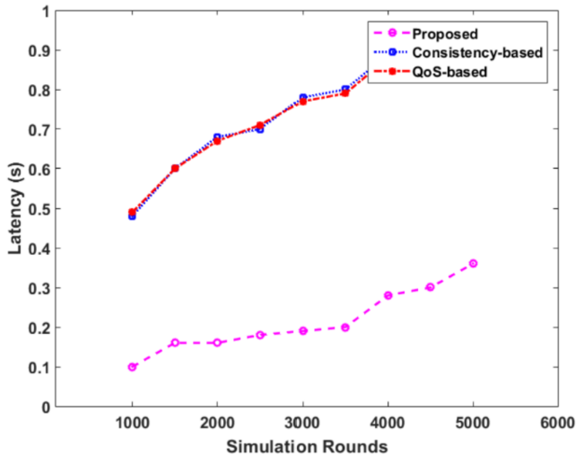

Figure 5. Comparative evaluation of latency incurred

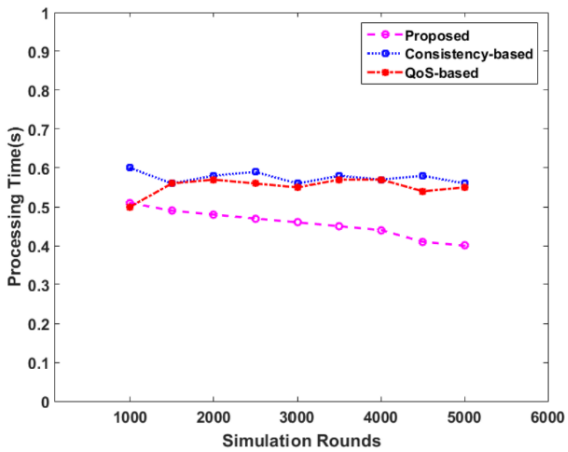

Figure 6. Comparative evaluation of processing time

4. CONCLUSION

The prime contribution of the proposed study model is to develop a unique modelling towards cost effective positioning of replica server considering the use case of cloud-based CDN system. The contribution of the proposed study model are as: i) a unique mechanism of computed cost based construction of structure of records of network towards reducing the dependencies of more number of replica server that indirectly reduces cost of data transmission; ii) the proposed scheme implements a unique concept of local virtual machine in order to replace the convention content server for cost reduction; this adaptation increases virtualization effectiveness creating a large chain of highly indexed structure resulting in higher availability of data and services; iii) the proposed scheme of replica server placement offers equal importance to user's experience and service quality; and iv) the study outcome shows that proposed system offers much better performance in contrast to existing scheme.

REFERENCES

[1] I. Kilanioti *et al.*, "Towards efficient and scalable data-intensive content delivery: state-of-the-art, issues and challenges," in *Lecture Notes in Computer Science*, Springer International Publishing, 2019, pp. 88–137.

[2] B. Zolfaghari *et al.*, "Content delivery networks," *ACM Computing Surveys*, vol. 53, no. 2, pp. 1–34, Mar. 2021, doi: 10.1145/3380613.

[3] T. Bilen and B. Canberk, "Deliver the content over multiple surrogates: A request routing model for high bandwidth requests," *Computer Communications*, vol. 146, pp. 39–47, Oct. 2019, doi: 10.1016/j.comcom.2019.07.009.

[4] S. Ul Islam *et al.*, "Leveraging utilization as performance metric for CDN enabled energy efficient internet of things," *Measurement*, vol. 147, Dec. 2019, doi: 10.1016/j.measurement.2019.07.042.

[5] G. Eslami and A. T. Haghighat, "A new surrogate placement algorithm for cloud-based content delivery networks," *The Journal of Supercomputing*, vol. 73, no. 12, pp. 5310–5331, Dec. 2017, doi: 10.1007/s11227-017-2088-5.

[6] Q. Fan *et al.*, "Video delivery networks: Challenges, solutions and future directions," *Computers and Electrical Engineering*, vol. 66, pp. 332–341, Feb. 2018, doi: 10.1016/j.compeleceng.2017.04.011.

[7] F. Ayankoya, O. Otushile, and B. Ohwo, "A review on content delivery networks and emerging paradigms," *International Journal of Scientific and Engineering Research*, vol. 9, no. 12, pp. 211–217, 2018.

[8] L. Yala, "Content delivery networks as a service (CDNaaS)," Universite Rennes 1, 2018.

[9] M. A. Salahuddin, J. Sahoo, R. Glitho, H. Elbiaze, and W. Ajib, "A survey on content placement algorithms for cloud-based content delivery networks," *IEEE Access*, vol. 6, pp. 91–114, 2018, doi: 10.1109/ACCESS.2017.2754419.

[10] D. Priyanka, Channakrishnaraju, and B. K. Chethan, "Insights on effectiveness towards research approaches deployed in content delivery network," in *Software Engineering Perspectives in Systems*, Springer International Publishing, 2022, pp. 224–243.

[11] V. Cohen-Addad, A. Gupta, L. Hu, H. Oh, and D. Saulpic, "An improved local search algorithm for k -median," in *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, Philadelphia, PA: Society for Industrial and Applied Mathematics, 2022, pp. 1556–1612.

[12] A. O. Al-Abbasi, V. Aggarwal, T. Lan, Y. Xiang, M.-R. Ra, and Y.-F. Chen, "FastTrack: minimizing stalls for CDN-based over-the-top video streaming systems," *IEEE Transactions on Cloud Computing*, vol. 9, no. 4, pp. 1453–1466, Oct. 2021, doi: 10.1109/TCC.2019.2920979.

[13] 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.

[14] Y. Fan, C. Wang, W. Wu, T. Znati, and D. Du, "Slow replica and shared protection: energy-efficient and reliable task assignment in cloud data centers," *IEEE Transactions on Reliability*, vol. 70, no. 3, pp. 931–943, Sep. 2021, doi: 10.1109/TR.2019.2923770.

[15] C. Guerrero, I. Lera, and C. Juiz, "Optimization policy for file replica placement in fog domains," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 21, Nov. 2020, doi: 10.1002/cpe.5343.

[16] L. V. Yovita and N. R. Syambas, "Caching on named data network: a survey and future research," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 6, pp. 4456–4466, Dec. 2018, doi: 10.11591/ijece.v8i6.pp4456-4466.

[17] W. T. Kusuma, A. A. Supianto, and H. Tolle, "Vertex markers: Modification of grid methods as markers to reproduce large size augmented reality objects to afford hands," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 1, pp. 1063–1069, Feb. 2020, doi: 10.11591/ijece.v10i1.pp1063-1069.

[18] Z. Liu, W. Hua, X. Liu, D. Liang, Y. Zhao, and M. Shi, "An efficient group-based replica placement policy for large-scale geospatial 3D raster data on Hadoop," *Sensors*, vol. 21, no. 23, Dec. 2021, doi: 10.3390/s21238132.

[19] N. Mostafa, W. H. F. Aly, S. Alabed, and Z. Al-Arnaout, "Intelligent replica selection in edge and IoT environments using artificial neural networks," *Electronics*, vol. 11, no. 16, Aug. 2022, doi: 10.3390/electronics11162531.

[20] B. Nazir, F. Ishaq, S. Shamshirband, and A. Chronopoulos, "The impact of the implementation cost of replication in data grid job scheduling," *Mathematical and Computational Applications*, vol. 23, no. 2, May 2018, doi: 10.3390/mca23020028.

[21] M. Safavi, S. Bastani, and B. Landfeldt, "Online learning and placement algorithms for efficient delivery of user generated contents in telco-CDNs," *IEEE Transactions on Network and Service Management*, vol. 17, no. 1, pp. 637–651, Mar. 2020, doi: 10.1109/TNSM.2019.2961560.

[22] L. Saino, I. Psaras, and G. Pavlou, "Framework and algorithms for operator-managed content caching," *IEEE Transactions on Network and Service Management*, vol. 17, no. 1, pp. 562–576, Mar. 2020, doi: 10.1109/TNSM.2019.2956525.

[23] D. Santos, R. Silva, D. Corujo, R. L. Aguiar, and B. Parreira, "Follow the user: a framework for dynamically placing content using 5G-enablers," *IEEE Access*, vol. 9, pp. 14688–14709, 2021, doi: 10.1109/ACCESS.2021.3051570.

[24] B. P. Shankar and S. Chitra, "Optimal data placement and replication approach for SIoT with edge," *Computer Systems Science and Engineering*, vol. 41, no. 2, pp. 661–676, 2022, doi: 10.32604/csse.2022.019507.

[25] Z.-L. Shao, C. Huang, and H. Li, "Replica selection and placement techniques on the IoT and edge computing: a deep study," *Wireless Networks*, vol. 27, no. 7, pp. 5039–5055, Oct. 2021, doi: 10.1007/s11276-021-02793-x.

[26] W. Teng, M. Sheng, K. Guo, and Z. Qiu, "Content placement and user association for delay minimization in small cell networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 10, pp. 10201–10215, Oct. 2019, doi: 10.1109/TVT.2019.2936182.

[27] H.-A. Tran, S. Souihi, D. Tran, and A. Mellouk, "MABRESE: a new server selection method for smart SDN-based CDN architecture," *IEEE Communications Letters*, vol. 23, no. 6, pp. 1012–1015, Jun. 2019, doi: 10.1109/LCOMM.2019.2907948.

[28] L. Xiong, L. Yang, Y. Tao, J. Xu, and L. Zhao, "Replication strategy for spatiotemporal data based on distributed caching system," *Sensors*, vol. 18, no. 1, Jan. 2018, doi: 10.3390/s18010222.

[29] K. Xu, X. Li, S. K. Bose, and G. Shen, "Joint replica server placement, content caching, and request load assignment in content delivery networks," *IEEE Access*, vol. 6, pp. 17968–17981, 2018, doi: 10.1109/ACCESS.2018.2817646.

[30] B. Yu and J. Pan, "A framework of hypergraph-based data placement among geo-distributed datacenters," *IEEE Transactions on Services Computing*, vol. 13, no. 3, pp. 395–409, May 2020, doi: 10.1109/TSC.2017.2712773.

[31] J. Sahoo, M. A. Salahuddin, R. Glitho, H. Elbiaze, and W. Ajib, "A survey on replica server placement algorithms for content delivery networks," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 1002–1026, 2017, doi: 10.1109/COMST.2016.2626384.

BIOGRAPHIES OF AUTHORS

Priyanka Dharmapal has 10 years of teaching experience for UG and PG courses in Computer Science and Engineering and is presently working as Assistant Professor in the Department of Computer Science and Engineering in Sri Siddhartha Institute of Technology, Tumkur. She obtained B.E degree from Visveswaraiah Technological University in the year 2009 and PG in Computer Science and Engineering in the year 2012 from Visveswaraiah Technological University. Her research interests are in the areas of content delivery networks, network security, mobile computing and cloud computing. She is currently pursuing doctoral degree in Sri Siddhartha Academy of Higher Education, Tumkur, under the guidance of Dr. Channakrishnaraju, Professor, Department of Computer Science and Engineering, Sri Siddhartha Institute of Technology, Tumkur. She can be contacted at email: prsh2019@gmail.com.

Channakrishnaraju has 26 years of teaching experience for UG and PG courses in computer Science and Engg, and is presently working as Professor in the Department of computer Science and Engg, in Sri Siddhartha Institute of Technology, Tumkur. He obtained B.E from Bangalore University in the year 1995 and PG in software systems in the year 2000 from BITS Pilani and doctoral degree in Sri Siddhartha Academy of higher Education, Tumkur. His research interests are in the areas of wireless sensor networks, network security and artificial intelligence. He can be contacted at email: rajuck@ssit.edu.in.

Chethan Bommalingaiahanapalya Krishnamurthy working as Associate Professor Department of CSE (AI and ML) in Vidyavardhaka college of Engineering, Mysuru, India. He obtained B.E degree from Visveswaraiah Technological University in the year 2005 and PG in Computer Science and Engineering in the year 2007 from Visveswaraiah Technological University and doctoral degree in the year 2020 from Visveswaraiah Technological University. He has around 16 years of teaching experience. His research domains are computer network, network security and software engineering. He can be contacted at email: chethan08@gmail.com.