Content Distribution in Cloud: A Survey

1st Nidhi Bhavsar, 2nd Prof. Jaimin Dave, 3rd Prof. Vijay Gadhavi

[1] M.E. Scholar, [2] Professor,[3] Professor
[1] Computer Engineering.,
[1] Alpha College of Engineering & Technology, Kalol, India
[1]nidhibhavsar2861988@gmail.com, [2]jaimin.dave@alpha-cet.in, [3]vijaybgadhavi@gmail.com

Abstract - Cloud Service has become a real trend in the development of network services. Content distribution network built on clouds have recently emerge and compared to conventional CDN, cloud based CDN have the benefit of cost efficient hosting services without owning infrastructure. Security is prime affecting factor in conventional CDN. So, we will propose a novel secure content distribution framework for cloud controller to securely disseminate their contents to legitimate users via Content distribution network. Use of our framework will enable legitimate users to receive encrypted content at nearby peer. Our framework utilizes in network caches for reducing delivery latency.

Index Terms - Content Distribution Network, Cloud Computing**,** Caching, Replica Placement, Cache Server

I. INTRODUCTION

Cloud computing has emerged as an enthralling paradigm for deploying Information on the Internet, because it enables solution providers to easily enlarge their offerings. More importantly, the elasticity of the "pay-as-you-go" paradigm enables solution providers to govern in operating costs, especially when demand is highly dynamic, or unpredictable. For instance, a cloud-based web hosting/caching solution can easily measure demand – and hence scale up or down its use of flexible cloud resources. User demand for popular and often high volume applications such as high-definition video, music, cloud-gaming, online social networks, and online-gaming is phenomenal; unbroken since years and still expected to grow. Content delivery networks have been widely used for many years providing service for millions of users. The goal of the cloud computing is to provide on-demand computing service with high reliability, scalability and availability in distributed environment. Lately, many of these networks are migrating to the cloud for its numerous advantages such as lower costs, increased performance, availability and flexibility. Compared to conventional CDNs, cloud-based CDNs have the benefit of cost efficient hosting services without owning infrastructure. It helps to achieve a high user satisfaction and resource utilization by ensuring an efficient and fair allocation of every computing resource.

II. BACKGROUND

2.1 CLOUD COMPUTING

Cloud Computing is currently one of the hottest topics in information technology (IT). Cloud computing is the evolution of a variety of technologies that have come together to change an organization"s approach for building an IT infrastructure. The cloud computing term describes a variety of different types of computing concepts that a large number of computers connected through a real time communication network. Cloud computing refers to manipulating, configuring and accessing the applications online. It offers online data storage, infrastructure and application. Building on compute and storage virtualization, cloud computing provides scalable, network-centric, abstracted IT infrastructure, platforms and applications as on-demand services that are billed by consumption. Cloud Computing relies on sharing of various resource(network, servers, storage, application and services) to achieve coherence and economies scale, and gives the highest interest to how to maximize the effectiveness of utilization of the shared resources. Cloud services are popular because they can reduce the cost and complexity of owning and operating computers and networks. Since cloud users do not have to invest in information technology infrastructure, purchase hardware, or buy software licenses, the benefits are low up-front costs, rapid return on investment, rapid deployment, customization, flexible use, and solutions that can make use of new innovations. In addition, cloud providers that have specialized in a particular area (such as e-mail) can bring advanced services that a single company might not be able to afford or develop.

Benefits

Cloud Computing has numerous advantages. Some of them are listed below:

- One can access applications as utilities, over the Internet.
- Manipulate and configure the application online at any time.
- It does not require installing a specific piece of software to access or manipulating cloud application.
- Cloud Computing offers online development and deployment tools, programming runtime environment through Platform as a Service model.

- Cloud resources are available over the network in a manner that provides platform independent access to any type of clients.
- Cloud Computing offers on -demand self – service. The resources can be used without interaction with cloud service provider.
- Cloud Computing is highly cost effective because it operates at higher efficiencies with greater utilization. It just requires an Internet connection.
- Cloud Computing offers load balancing that makes it more reliable.

2.2 CONTENT DISTRIBUTION NETWORK

"Content is King": Predicted by Bill Gates in an essay from 1996, this quote has become the latest buzz in the Internet economy. A content delivery network (CDN) is a system of distributed servers (network) that deliver WebPages and other Web content to a user based on the geographic locations of the user, the origin of the webpage and a content delivery server. The goal of a CDN is to serve content to end-users with high availability and high performance. CDNs serve a large fraction of the Internet content today, including web objects (text, graphics and scripts), downloadable objects (media files, software, and documents), applications (e-commerce, portals), live streaming media, on-demand streaming media, and social networks. Content Distribution network is simply a bunch of disks that are spread across the world in different datacenters. Content Distribution network is simply a network of servers that replicates your binary files so that they are served from geographically close locations. CDN works on the principle of delivering content from the nearest located server as per user location. A content delivery network (CDN) is a system of multiple computers that contains copies of data stored at various network nodes. A well designed and appropriately implemented CDN improves data access by enhancing bandwidth and minimizing access latency.

While content delivery networks also solve ancillary problems such as improving global availability and reducing bandwidth, the main problem they address is latency: the amount of time it takes for the host server to receive, process, and deliver on a request for a page resource (images, CSS files, etc.). Latency depends largely on how far away the user is from the server, and it‟s compounded by the number of resources a web page contains.

The typical functionality of a CDN includes:

- Request redirection and content delivery services to direct a request to the closest suitable surrogate server using mechanisms to bypass congestion.
- Content outsourcing and distribution services to replicate and/or cache content to distributed surrogate servers on behalf of the origin server.
- Content negotiation services to meet specific needs of each individual user (or group of users).
- Management services to manage the network components, to handle accounting, and to monitor and report on content usage.

2.2.1 CACHINNG AND REPLICATION

CDN providers use caching and/or replica servers located in different geographical locations to replicate content. CDN cache servers are also called edge servers or surrogates . CDNs distribute content to the surrogates in such a way that all cache servers share the same content and URL. Client requests are redirected to the nearby cache, and a selected cache server delivers requested content to the end-users. Thus, transparency for users is achieved. Additionally, surrogates send accounting information for the delivered content to the accounting system of the CDN provider. A cache server may be equipped with a streaming media cache. This enables on-demand content to be dynamically replicated locally, perhaps in an encrypted format. The surrogate may attempt to store all cacheable media files upon first request. When a cache server receives a client request for on-demand media, it determines whether the content is cacheable. Then it checks to see whether the requested media already resides in its local cache. If the media is not already in the cache, the cache server acquires the media file from the source server and simultaneously delivers it to the requesting client. Subsequent requests for the same media clip can be served without repeatedly pulling the clip across the network from the source server.

Replica placement mechanisms are needed to decide the replica server locations and to adaptively fill them with the proper content Adaptivity in replica placement is required to cope with changing traffic condition and is not related to a pull behavior like in traditional caching.

2.2.2 RESOURCE PROVISIONING

The cloud computing paradigm offers users rapid on-demand access to computing resources such as CPU, RAM and storage, with minimal management overhead. Recent commercial cloud platforms, exemplified by Amazon EC2, Microsoft Azure and Linode, organize a shared resource pool for serving their users. Resource provisioning means the selection, deployment, and run-time management of software (e.g., database management servers, load balancers) and hardware resources (e.g., CPU, storage, and network) for ensuring guaranteed performance for applications. This resource provisioning takes Service Level Agreement (SLA) into consideration for providing service to the cloud users. This is an initial agreement between the cloud users and cloud service providers which ensures Quality of Service (QoS) parameters like performance, availability, reliability, response time etc. Based on the application needs Static Provisioning/Dynamic Provisioning and Static/Dynamic Allocation of resources have to be made in order to efficiently make use of the resources without violating SLA and meeting these QoS parameters. Over provisioning and under provisioning of resources must be avoided. Another important constraint is power consumption. Care should be taken to

reduce power consumption, power dissipation and also on VM placement. There should be techniques to avoid excess power consumption. So the ultimate goal of the cloud user is to minimize cost by renting the resources and from the cloud service provider‟s perspective to maximize profit by efficiently allocating the resources. In order to achieve the goal the cloud user has to request cloud service provider to make a provision for the resources either statically or dynamically so that the cloud service provider will know how many instances of the resources and what resources are required for a particular application. By provisioning the resources, the QoS parameters like availability, throughput, security, response time, reliability, performance etc must be achieved without violating SLA.

Parameters for Resource Provisioning

- **Response time:** The resource provisioning algorithm designed must take minimal time to respond when executing the task.
- **Minimize Cost:** From the Cloud user point of view cost should be minimized.
- **Revenue Maximization:** This is to be achieved from the Cloud Service Provider‟s view.
- **Fault tolerant:** The algorithm should continue to provide service in spite of failure of nodes.
- **Reduced SLA Violation:** The algorithm designed must be able to reduce SLA violation.
- **Reduced Power Consumption:** VM placement & migration techniques must lower power consumption.

III. OVERVIEW OF CLOUD BASED CONTENT DISTRIBUTION NETWORK

Content Delivery Networks (CDNs) were introduced to allow for highly available and quick content delivery by keeping most recent content at servers near to users who usually request this data. This becomes possible by distributing the task of data delivery on multiple centers in order to offload origin servers by delivering data on their behalf. Recently, many CDN providers started migrating their networks into the cloud as the cloud provides numerous advantages for both CDN users and providers. The cloud helps reducing transmission latency as data is stored closest to the user. Operating costs are also reduced where resources could be rent from the cloud provider on demand. Cost reduction will also affect the users as they will no longer need to install physical storage devices to be part of the CDN, and will only pay for the content usage and content transfer. WITH the successful deployment of commercial systems and increasing user popularity, content distribution networks (CDNs) have received much attention in recent years. Conventional CDNs such as Akamai built hundreds of data centers to distribute the contents across the world. It has hence become financially prohibitive for small content providers to compete on a large scale by deploying new data centers. The emerging cloud vendors such as Amazon S3 are creating new opportunities to enable cost-effective CDNs. As the cloud vendors provide on-demand and cost-effective content storage and delivery capabilities, one can build CDNs upon the clouds without investments on installing and maintaining the infrastructure while providing scalable service. As a customer, a cloud CDN may benefit from elastic cloud charge models. Also, the cloud CDN can dynamically adjust the leases of bandwidth, virtual machine (VM), and storage resources based on the runtime demand rates to reduce the total rental costs without severely sacrificing the service performance [1].

A cloud - based CDN architecture can provide the following advantages:

- An elastic platform with ability to dynamically and easily scale capacity up and down
- Hides the infrastructure complexity from CDN applications and content provides
- Enable a QoS driven performance management
- Open standard approach to tap into the capabilities of public clouds to scale during peak demand

IV. RELATED WORK

Many research efforts have been made on those existing cloud based content distribution network, such as Bittorrent, VoD and live Streaming. Many cloud based CDN architectures are also available.

A recent study by Menglan Hu, Jun Luo, Member, IEEE, Yang Wang, and Bharadwaj Veeravalli, developed architecture as shown in Figure 1. In this architecture, a root site aims to serve K (types of) contents (e.g., videos) to users residing on M locations. Since the root site is far from the users, to meet QoS constraints, the root needs a CDN to help serve users‟ demands (or requests). The CDN is built on resources leased from N cloud sites located across the Internet[1]. They proposed a set of novel algorithms to solve the joint problem of resource provisioning and caching (i.e., replica placement) for cloud-based CDNs with an emphasis on handling the dynamic demand patterns. Firstly, they proposed a provisioning and caching algorithm framework called Differential Provisioning and Caching (DPC) algorithm, which aims to rent cloud resources to build CDNs and whereby to cache contents so that the total rental cost can be minimized while all demands are served. DPC consists of 2 steps. Step 1 first maximizes total demands supported by unexpired resources. Then, step 2 minimizes the total rental cost for new resources to serve all remaining demands. For each step we design both greedy and iterative heuristics, each with different advantages over the existing approaches. Moreover, to dynamically adjusts the placement of contents and route maps, they further proposed the Caching and

Request Balancing (CRB) algorithm, which is light-weight and thus can be frequently executed as a companion of DPC to maximize the total demands.

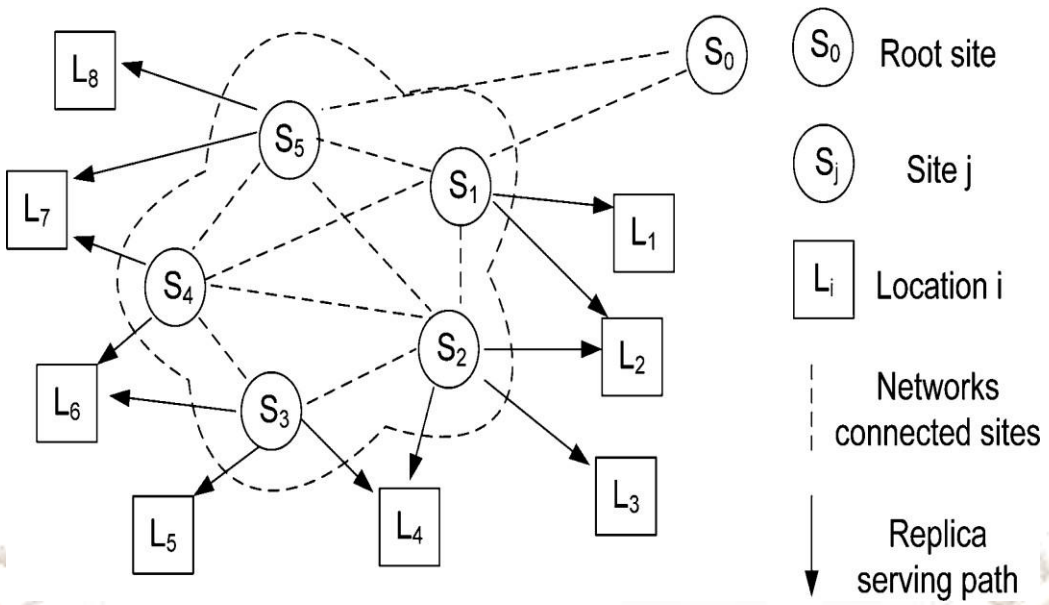
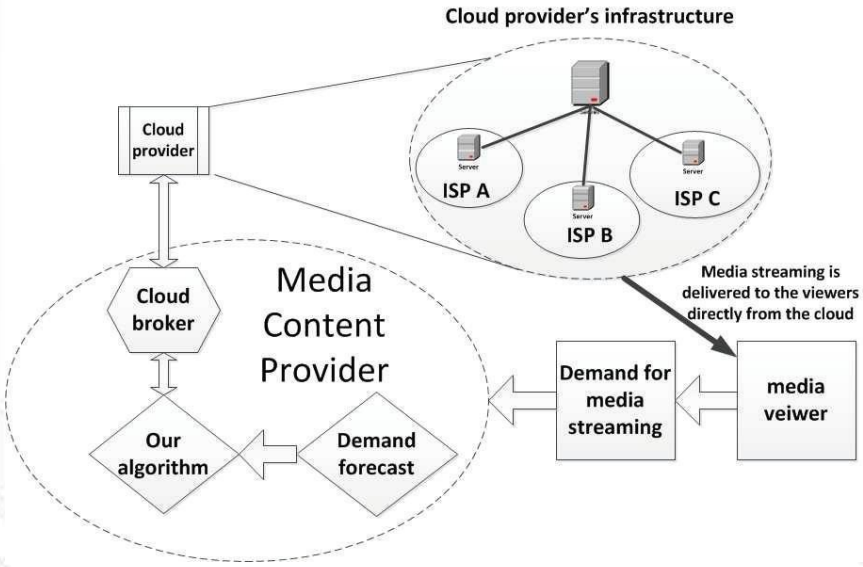
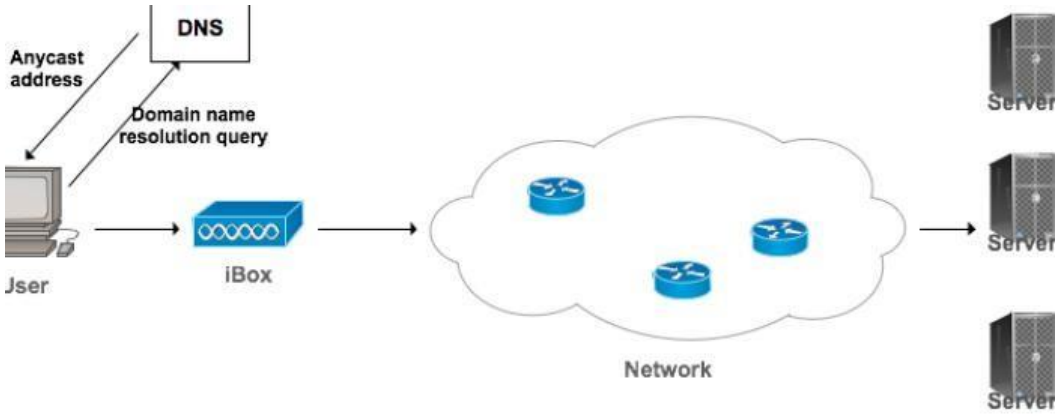


Fig 1: Cloud Content Distribution Network

In another research, Amr Alasaad, Kaveh Shafiee, Sathish Gopalakrishnan, and Victor C.M. Leung has proposed one architectures shown in figure 2. The system model consists of the following components. Media content provider, demand forecasting model, cloud broker, and cloud provider (Fig. 2). The cloud broker is located in the media content provider"s site and is responsible on behalf of the content provider for both allocating the appropriate amount of resources in the cloud, and reserve the time over which the required resources are allocated. Given the demand prediction, the broker implements proposed algorithm to make optimum decision for resource provisioning.



In this paper, Hai Anh TRAN, Abdelhamid MELLOUK, Said HOCEINI has proposed a Content Distribution Network Cloud Architecture (CDNCA). Our approach is based not only on Quality of Service criteria (QoS) (e.g. round trip time, network hops, loss rate, etc.) but also on the Quality of Experience that represents end-users perception and satisfaction. Experimental results show that CDNCA yields significant improvements over traditional approaches In this paper, we have proposed a deployable and scalable architecture, called CDNCA, for a Content Distribution Networks Cloud architecture. The main idea of CDNCA is to construct a cloud content distribution network that takes into account the end-user perception. In fact, as End-to-End QoS optimization is an NP-complete problem regarding two or more non additive criteria, CDNCA maintain a high level service regarding QoE criteria instead of optimizing multiple QoS criteria. It processes in two phases: selecting the "best" server among several replicated servers and routing requested data from this chosen server to end-user[4].



We consider QoE measured at end-user station in the system as metric to choose the "best" replicated server as well as route data content from these servers to end-users. This component works as a set-top box which we named here iBox (Intelligent Box) (Fig. 3). This iBox component has the functionality to select the server based on a server selection table. Fig. 3 illustrates CDNCA architecture, which consists of replicated servers, end-user station and iBox. The iBoxes placing at client side stocks *the server selection table* (Tab 1). The QoE measurement method (presented in section III-C) is integrated into the iBoxes.

V. CONCLUSION

Emerged within last decade, content distribution network based on cloud have quickly been recognized as a useful method for improving Internet content distribution scalability and quality. By replicating the content to a worldwide network of servers and redirecting the users to the nearest server based on the user request prediction and server load, CDN could improve the content access latency. So this paper includes the basic concepts related to cloud based content distribution network.

VI. REFERENCES

[1] Menglan Hu, Jun Luo, Yang Wang and Bharadwaj veeravalli , *"Practical Resource Provosong and Caching with Dynamic Resilience for Cloud Based Content Distribution Networks"*

[2] Amr Alasaad, Kaveh Shafiee, Sathish Gopalkrishnan and Victor C.M.Leung WiNMoS Lab , —*Prediction-based Resource Allocattion in Clouds for Media Streaming Applications* in 2012

[3] Zhi-Li Zhang , "*Feel Free to Cache Towards an Open CDN Architecture for Cloud-based Content Distribution"*

[4] Chrysa Papagianni, Aris Leivadeas, and Symeon Papavassilliou – *" Cloud- Oriented Content Delivery Network Paradigm : Modeling and Assessment "*– in September/October 2013.

[5] Hai Anh TRAN, Abdelhamid MELLOUK, Said HOCEINI —"*QoE Content Distribution Network for Cloud Architecture"* in 2011 First International Sympoisum on Network Cloud Computing and Application

[6] Shivadon Chaisiri, Bu-Sung Lee, Dusit Niyato – "*Optimization of Resource Provisioning Cost in Cloud Computing".*

[7] Balachander Krishnamurthy, Craig Wills and Yin Zhang —*On the Use and Performance of the Content Distribution Network".*

[8] Four Reasons We Choose Amazon‟s Cloud as Our Computing Platform,"" The Netflix „„Tech"" Blog, , Dec. 2010.