ISSN:0975-9646

State-of-the-art of Content Delivery Network

Manish Gupta, Dharmendra Kumar

Department of Computer Science & Engineering, Uttar Pradesh Technical University
United College of Engineering & Research, Allahabad, Uttar Pradesh, India

***Abstract*— Content Delivery Networks (CDNs) have evolved to overcome the inherent limitations of the Internet. CDNs approach is based on the observation that serving web content from a single location can present serious problems for site scalability, reliability and performance. Hence, a system is devised to serve requests of end-user from nearby optimal surrogate server. CDN replicates content from the origin server to cache servers scattered over the globe. This seminar reports provides the broad coverage of CDNs in terms of layered architecture, basic interactions, technical issues involved, applications, advantages and disadvantages. It also provides the analysis of state-of-arts of existing CDNs and their comparative study on the basis of content distribution and management and request routing technique.**

***Keywords*— Content Delivery Network (CDN), Architecture of CDN, Comparison of CDNs.**

I. INTRODUCTION

Applications that are accessible over the internet usually follow a client server paradigm, where the web browser act as the client program for accessing the web pages provided by the web server. This set of pages may consist of unchanging information or it may consist of dynamically generated information, the user-perceived response of the web-server is unacceptably large. In some cases the performance is poor due to an overload at the server. In other cases, the performance is poor due to the congestion within the internet.

When poor performance is due to an overload at the server, the right approach to handle the performance problem is to increase the capacity of the server. However, an approach that cans poor performance due to network congestion is still unclear.

An alternative approach to work around the performance problems in the internet is to have a set of cooperating edge servers distributed within the internet, with constant from web servers cached or replicated at the edge servers. This type of coordinated distributed scheme is known as a content distribution network.

When a user requests a web page that is part of a CDN the CDN will redirect the request from the originating site's server to a server in the CDN that is closest to the user and deliver the cached content. The CDN will also communicate with the originating server to deliver any content that has not been previously cached. The closer the CDN server is to the user geographically, the faster the content will be delivered to the user. CDNs also provide protection from large surges in traffic. [5]

The process of bouncing through a CDN is nearly transparent to the user. The only way a user know if a CDN has been accessed is if the delivered URL is different that the URL that has been requested.

This paper explores the CDNs in terms of architecture, basic interactions, technical issues involved, applications, advantages and disadvantages. It also provides the analysis of state-of-arts of existing CDNs and their comparative study.

II. OVERVIEW

A Content Delivery Network or Content Distribution Networks (CDN) consists of a collection of non origin servers that attempts to offload work from origin servers by delivering content on their behalf.

A CDN improve network performance by maximizing bandwidth, improving accessibility and maintaining correctness through content replication [1]. Now-a-days, CDN is commonly used by internet advertisement companies, data centers, ISP's, online music retailers, mobile operators and other companies. A typical CDN is combination of:

- *Content Delivery Infrastructure:* The content delivery infrastructure consists of a set of edge servers that deliver copies of content to end user [2].
- *Request Routing Infrastructure:* The request routing infrastructure is responsible for directing lien request to appropriate edge servers. It also interacts with the distribution infrastructure in order to keep track of up-to-date content stored in the CDN caches.
- *Distribution Infrastructure:* The distribution infrastructure moves content from the origin server to the CDN edge servers and ensures consistency of content in the caches [2].
- *Accounting Infrastructure:* The accounting infrastructure maintains logs of client accesses and records the usage of the CDN servers. This information is used traffic reporting and usage based billing [2].

A CDN is a collection of network elements arranged for more effective delivery of content to end users [7]. Content Delivery Networks (CDN), which evolved first in 1998 [7], replicate content over several mirrored web servers strategically placed at various locations in order to deal with the flash crowds. A CDN provides better performance through caching and replicating content over some mirrored web server [2].

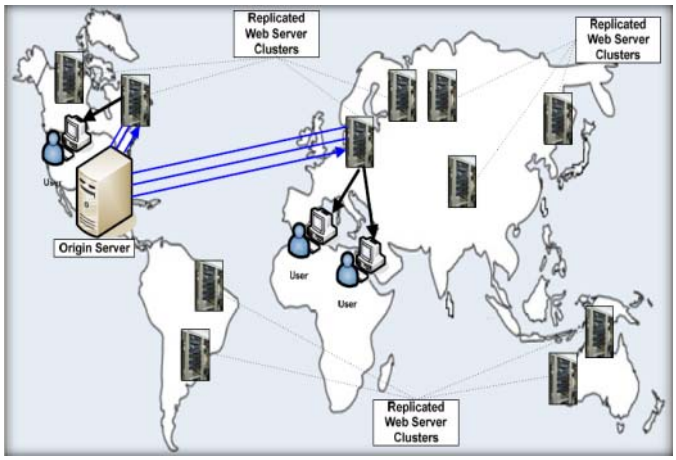

Fig. 1 Abstract architecture of a Content Delivery Network

Figure-1 shows a typical content delivery environment where the replicated web server clusters are located at the edge of the network to which the end users are connected. In such CDN environment, web content based on user requests are fetched from the origin server and a user is served with the content from the nearby replicated web server. Thus the users end up communicating with a replicated CDN server close to them and retrieves files from that server.

III. Basic interactions in a CDN

The high level view of the basic interaction flows among the components in a CDN environment. Here, *discovery.com* is the content provider and Akamai is the CDN that hosts the content of *discovery.com*.

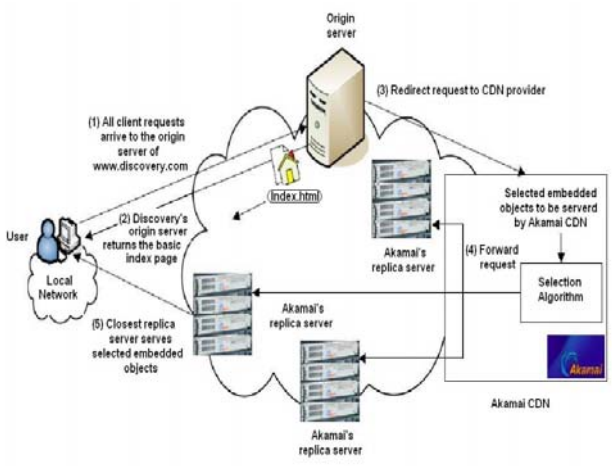

Fig. 2 Basic interaction flows in a CDN environment

The interaction flows are:

1. User requests content from www.discovery.com by specifying its URL in the web browser. User's request is directed to the server of discovery.com.
2. When discovery.com receives a request, its web server makes a decision to provide the basic contents (e.g. Index page of the site) that can be served from its origin server.
3. To serve the high bandwidth demanding and frequently asked contents, (embedded objects-fresh content, banner ads etc), discovery's origin server redirects user's request to the CDN provider (Akamai in this case).
4. Using the proprietary selection algorithm the CDN provide selects the replica server which is 'closest' to the end user, in order to serve the requested embedded objects.
5. Selected replica server gets the embedded object from the origin server, serves the end-user requests and caches it for subsequent request servicing.

IV. CDN: State of the art

With a CDN, end-user's requests are served from the web servers distributed around the Internet that cache the content originally stored in the origin server. A CDN is different from P2P networks and data grids in the sense that, CDNs are dedicated to caching web content so that users are able to access it faster; while P2P content sharing networks are vertically integrated to achieve a single goal (e.g. file sharing), and data grids provide a platform through which users can access aggregated computational, storage and networking resources to execute their data-intensive applications on remote data. The list of existing CDNs and there summary is given below.

A. Commercial CDNs

Table I
List of Commercial CDNs

Name	Web addresses
Accelia	www.accelia.net
Accellion	www.accellion.com
Activate	www.active.com
Akamai	www.akamai.com
AppStream	www.appstream.com
EdgeStream	www.edgestream.com
Globix	www.globix.com
LimeLight Networks	www.limelightnetworks.com
LocalMirror	www.localmirror.com
Mirror Image	www.mirror-image.com
Netli (Now acquired by Akamai)	www.netli.com
SyncCast	www.synccast.com
Tata Communications	www.tatacommunications.com
Value CDN	www.valuecdn.com
VitalStream	www.vitalstream.com

B. Academic CDNs

Table II
List of Academic CDNs

Name	Web addresses
CoDeeN	www.codeen.cs.princeton.edu
COMODIN	http://lisdip.deis.unical.it/research/index.html
Coral	www.coralcdn.org
Globule	www.globule.com

TABLE III
SUMMARY OF EXISTING COMMERCIAL CDNS

CDN Name	Service Type	Coverage	Products/Solutions (If any)
Accellion	Provide large file delivery service	Covers industries such as advertising/media production, healthcare, manufacturing, consumer goods, higher education etc.	Accellion Courier SFTA for on-demand and secure exchange of large-size files, and Accellion BRS as online desktop and server backup and recovery solution
Akamai	Provides CDN service, including streaming	Covers 85% of the market 20,000 servers in nearly 1,000 networks in 72 countries. It handles 20% of total Internet traffic today	Edge Platform for handling static as well as dynamic content, Edge Control for managing applications, and NOCC
AppStream	Provides on demand software distribution and software license management tools for extended enterprises	Small business to Fortune 1,000 corporations, educational institutions and government	AppStream software V5.0 including AppStream server, client and end-user application portal
EdgeStream	Provide disrupted video streaming application over the public internet	Provides video streaming over consumer cable or ADSL modem connections around the globe, even over paths that have 20 router hops between server and end-user	EdgeStream video on-demand and IPTV streaming software for video streaming
Globix	Provides Internet infrastructure and network services	Consists of both a trans-Atlantic/trans-continental IP backbone as well as an optical network throughout the Northeast and mid-Atlantic regions. It has more than 1200 customers	N/A
LimeLight Networks	Provides distributed on-demand and live delivery of video, music, games, and download	Surrogate server located in 72 locations around the world	LimeLight ContentEdge for distributed content delivery via HTTP, LimeLight MediaEdge streaming for distributed video and music delivery via streaming , and LimeLight custom CDN for custom distributed delivery solutions
Mirror Image	Provides content delivery, streaming media, web computing and reporting services	Surrogate server located in 22 countries around the world	N/A

TABLE IV
SUMMARY OF EXISTING ACADEMIC CDNS

CDN Name	Service Type	Coverage	Products/Solutions (if any)
CoDeeN	Provides caching of content and redirection of HTTP reuests	CoDeeN is an academic testbed content distribution network built on top of PlannetLab	N/A
COMODIN	Provides collaborative media playback service	COMODIN is an academic streaming CDN deployed on an international testbed	N/A
Coral	Provides content replication in proportion to the content's popularity	Coral is free peer-to-peer CDN. During beta testing, the Coral node network is hosted on PlanetLab, a large scale distributed research network of 400 servers, instead of third party volunteer systems. Of those 400 servers, about 275 are currently running Coral	N/A
Globule	Provide replication of content, monitoring of servers and redirecting client requests to available replicas	Globule is an open source collaborative CDN	N/A

V. COMPARISON OF EXISTING CDNS

TABLE V
COMPARISON ON BASIS OF CONTENT DISTRIBUTION AND MANAGEMENT

CDN Name	Surrogate placement	Content Selection and Delivery	Content Outsourcing	Cache Organization
Accellion	Single-ISP approach	*Content selection* • Partial-site content delivery *Content Clustering* • N/A	Non-cooperative pull-based	*Caching technique* • Inter-cluster caching *Cache update* • N/A
Akamai	Multi-ISP approach	*Content selection* • Full and partial-site content delivery *Content Clustering* • User's sessions based	Non-cooperative pull-based	*Caching technique* • Intra and inter-cluster caching *Cache update* • On-demand
AppStream	Single-ISP approach	*Content selection* • Partial-site content delivery *Content Clustering* • N/A	Non-cooperative pull-based	*Caching technique* • Intra-cluster caching *Cache update* • On-demand
EdgeStream	Single-ISP approach	*Content selection* • Partial-site content delivery *Content Clustering* • N/A	Non-cooperative pull-based	*Caching technique* • Inter-cluster caching *Cache update* • N/A
Globix	Multi-ISP approach	*Content selection* • Full and partial-site content delivery *Content Clustering* • N/A	Non-cooperative pull-based	*Caching technique* • Intra-cluster caching *Cache update* • On-demand
LimeLight Networks	Multi-ISP approach	*Content selection* • Partial-site content delivery *Content Clustering* • N/A	Non-cooperative pull-based	*Caching technique* • Intra-cluster caching *Cache update* • On-demand
Mirror Image	Multi-ISP approach	*Content selection* • Partial-site content delivery *Content Clustering* • URL based	Non-cooperative pull-based	*Caching technique* • Intra-cluster caching *Cache update* • On-demand
Netli	Single-ISP approach	*Content selection* • Partial-site content delivery *Content Clustering* • N/A	Non-cooperative pull-based	*Caching technique* • Inter-cluster caching *Cache update* • N/A
SyncCast	Single-ISP approach	*Content selection* • Full and partial-site content delivery *Content Clustering* • N/A	Non-cooperative pull-based	*Caching technique* • Intra-cluster caching *Cache update* • On-demand
CoDeeN	Multi-ISP approach	*Content selection* • Partial-site content delivery *Content Clustering* • N/A	Cooperative pull-based	*Caching technique* • Intra and inter-cluster caching *Cache update* • N/A
COMODIN	Single-ISP approach	*Content selection* • Full-site content delivery *Content Clustering* • N/A	Non-cooperative pull-based	*Caching technique* • Intra-cluster caching *Cache update* • Periodic update
Coral	Multi-ISP approach	*Content selection* • Full and partial-site content delivery *Content Clustering* • Users sessions based	Cooperative pull-based	*Caching technique* • Intra and inter-cluster caching *Cache update* • Cache invalidation
Globule	Single-ISP approach	*Content selection* • Partial-site content delivery *Content Clustering* • N/A	Cooperative pull-based	*Caching technique* • Intra and inter-cluster caching *Cache update* • N/A

TABLE VI
COMPARISON ON BASIS OF REQUEST-ROUTING TECHNIQUE

CDN Name	Request-Routing Technique
Accellion	HTTP redirection
Akamai	DNS-based request-routing
AppStream	HTTP redirection
EdgeStream	HTTP redirection
Globix	Global Server Load Balancing (GSLB) • Global awareness • Smart authoritative DNS
Mirror Image	Global Server Load Balancing (GSLB) • Global awareness • Smart authoritative DNS
Netli	DNS-based request-routing
SyncCast	Global Server Load Balancing (GSLB) • Global awareness • Smart authoritative DNS
CoDeeN	HTTP redirection
COMODIN	DNS-based request-routing
Coral	DNS-based request-routing
Globule	HTTP redirection DNS-based request-routing

VI. TECHNICAL ISSUES INVOLVED IN CDN

Since CDNs spread over multiple geographic location and involves multiple distributed components, there are several technical issues related to CDN content delivery are as follows:

A. Placement of Surrogates in CDN:

Since location of surrogate servers is closely related to the content delivery process, it puts extra emphasis on the issue of choosing the best location for each surrogate. Some theoretical approaches, server placement algorithms, and scalable replica management framework have been developed to model the surrogate server placement problem. For surrogate server placement, the CDN administrators also determine the optimal number of surrogate servers using single-ISP and multi-ISP approach.

B. Selection of Content:

The choice of content to be delivered to the end-users is important for content selection. Content can be delivered to the customers in full or in partial. In full-site content delivery the surrogate servers perform *entire replication* in order to deliver the total content site to the end-users. In contrast, partial content delivery provides only embedded objects – such as Web page images from the corresponding CDN.

C. Request Routing:

To select the most appropriate surrogate server for content routing several routing schemes can be used. They are: Global Server Load Balancing (GSLB), client multiplexing, URL rewriting, Anycasting and CDN peering. These routing schemes are stated in the following:

- In *Global Server Load Balancing (GSLB)*, service nodes (which serve contents to end-users) consisting of a GSLB-enabled Web switch and a number of real Web servers are distributed in several locations around the world. The GSLB-enabled switches are responsible for routing the client requests.
- In *Client multiplexing*, the client obtains the physical addresses of a set of physical replica servers and chooses one to send its request to.
- In *URL rewriting*, the origin server redirects the clients to different surrogate servers by rewriting the dynamically generated pages' URL links.
- The *Anycasting* approach, the client's request is sent to one server that serves the anycast address for a group of replicated Web servers.
- In *CDN peering* approach, Peer-to-peer content networks are formed by symmetrical connections between host computers. Peered CDNs deliver content on each other's behalf.

D. Outsourcing of Content:

Given a set of properly placed surrogate servers in a CDN infrastructure and a chosen content for delivery, it is crucial to decide which content outsourcing practice is to follow:

- Cooperative push-based- In *Cooperative push-based* approach, content is pushed to the surrogate servers from the origin and each request is directed to the closest surrogate server or otherwise the request is directed to the origin server.
- Non-cooperative pull-based- In *Non-cooperative pull-based* approach, client requests are directed (DNS redirection, HTTP redirection or URL rewriting) to their closest surrogate servers. If there is a cache miss, surrogate servers pull content from the origin server.
- Cooperative pull-based-The *cooperative pull-based* approach differs from the non-cooperative approach in the sense that surrogate servers cooperates each other to get the requested content in case of cache miss. Using a distributed index, the surrogate servers find nearby copies of requested content and store in the cache.

E. Content Replication and Caching:

Replicating content is common in large scale distributed environment like CDNs. Commercial CDNs

(e.g. Akamai) replicate content across the globe for high profile customers such as Symantec and Apple, that need to deliver large volumes of data in a timely manner. Content caching in CDNs can be

- Intra-cluster- For *Intra-cluster* caching of contents either of query-based scheme, digest-based scheme, directory-based scheme, and hashing-based scheme can be used.
- Inter-cluster basis-*Inter-cluster* content routing is necessary when intra-cluster content routing fails.

VII. APPLICATIONS

Some of the common applications for CDN are as follows**:-**

- Large File Downloads and Object Delivery
- Video (live and on-demand, HD)
- Music (live and on-demand)
- IPTV, set-top TV
- Software (including live online gaming)
- Images
- Flash files
- Social Media
- High-Traffic Large Library Web Sites
- Online Advertising [6]

VIII. ADVANTAGES

- *Improved end-user experience*- Increased content delivery speeds
- *Scalability*- Scalable delivery and storage
- *Investment*- Minimized costs of delivery versus a self hosting infrastructure build
- *Time to market*- Quick implementation, reduced time to market
- *Availability*- Redundant and resilient networks.[6]
- *Reliability*- If one servers fail, many more to step in its place
- *Security*- Additional layer of capacity and routing aiding security of the origin sever ensuring availability of content[6]

IX. DISADVANTAGES

- Some of the larger CDN's charge a setup fee, possibly to keep away smaller clients [4].
- Pricing is not often very obvious [4].
- Complexity in deployment procedure [3].

X. FUTURE SCOPE

- A unified content network – A unified content network, which supports the coordinated composition and delivery of content and services, is highly desirable.
- Towards a research Content Network (CN) – A test bed should have the ability to provide access control in order to limit concurrent experiments to a reasonable level.
- Load distribution for cooperative CDNs – The load distribution strategy for cooperative CDNs needs to be addresses on four important components:
    - Task assignment
    - Traffic congestion
    - Load dissemination and
    - Cooperative caching

Coordination among these core issues is another important consideration for successful exploitation of load distribution strategy.

XI. CONCLUSION

Content distribution networks (CDNs) are a mechanism to deliver content to end users on behalf of origin Web sites. Content distribution offloads work from origin servers by serving some or all of the contents of Web pages. Along with the proliferation, formation, and consolidation of the CDNs, new forms of internet content and services are coming into the picture. Consequently, content distribution, caching and replication techniques are gaining more attention in order to meet up the new technical and infrastructure requirements of the next generation CDNs. This may lead to new issues in the design and architecture of CDNs.

REFERENCES

[1] Rajkumar Buyya, Al-Mukaddim Khan Pathan, A Case for Peering of Content Delivery networks, IEEE Distributed Systems Online, October 2006.
[2] Mukaddin Pathan and Rajkumar Buyya-"A Taxonomy and Survey of Content Delivery Networks", Technical Report, GRIDS-TR-2007-4, Grid Computing and Distributed Systems Laboratoty, The University of Melbourne, Australia. 12 February, 2007.
[3] http://stackoverflow.com/questions/21452 77/what-are-the-advantages-and-disadvantages-of-using-cdncontent-delivery-network
[4] http://flowplayer.org/tutorials/int roduction-to-streaming-servers.html
[5] http://www.webopedia.com/TERM/C/ CDN.html
[6] http://www.vimeric.com/about-cdn.aspx
[7] F. Douglis, and M. F. Kaashoek, "Scalable Internet Services", IEEE Internet Computing, Vol. 5, No. 4, 2001, pp. 36-37.