Université d'Ottawa · University of Ottawa

Université d'Ottawa · University of Ottawa

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTDOCTORALES

FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

Yaxin CHAO

AUTEUR DE LA THÈSE - AUTHOR OF THESIS

M. Sc. (Systems Science)

GRADE - DEGREE

Systems Science

FACULTÉ, ÉCOLE, DÉPARTEMENT - FACULTY, SCHOOL, DEPARTMENT

TITRE DE LA THÈSE - TITLE OF THE THESIS

Content Delivery Networks

D. Wright

DIRECTEUR DE LA THÈSE - THESIS SUPERVISOR

CO-DIRECTEUR DE LA THÈSE - THESIS CO-SUPERVISOR

EXAMINATEURS DE LA THÈSE - THESIS EXAMINERS

M. Benjoucef

J. Nash

J.-M. De Koninck, Ph.D.

LE DOYEN DE LA FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTDOCTORALES

SIGNATURE

DEAN OF THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

Content Delivery Networks

By

Yaxin Chao

A thesis submitted to the Faculty of Graduate and Postdoctoral Studies in partial fulfillment of the degree of Master of Science

Supervisor: Dr. David Wright

University of Ottawa

Ottawa, Ontario, Canada K1N 6N5

Your file Votre référence
ISBN: 0-612-90043-6
Our file Notre référence
ISBN: 0-612-90043-6

Acknowledgements

I would like to express my heartiest appreciation and gratitude to my supervisor, Dr. David Wright, for his support, encouragement and guidance during my completion of the thesis.

Many thanks go to my grandparents, my parents and my wife Liou, for their unfailing love, encouragement and support.

Abstract

Since bottlenecks and congestion often happen with the phenomenal growth in the use of the network, the demand for improving network delivered performance is very necessary. Deploying Content Delivery Networks (CDNs) technology to improve delivery performance has caused more and more people's attention. Content Delivery Networks employ many geographically distributed sites to improve the scalability and improves network performance by reducing the client's response time.

In this thesis, we initially present the background knowledge on CDNs, which includes CDNs concept, function, methodology, components and market analysis. The thesis emphasizes analyzing, comparing and evaluating various aspects of CDNs. We divide the whole comparison and evaluation into three parts. Firstly, we analyze and compare CDNs to other methods for improving performance and congestion control. Secondly, we overview different technologies that can be used within CDNs, then compare and analyze various CDNs components and implementations. Thirdly, we select current representative CDNs companies or providers to compare and evaluate their products and services. In this thesis, we also describe some main applications of CDNs and do a case study of Cisco ECDN and Volera Velocity CDN for E-learning. Finally, we conclude this thesis by summarizing our analysis and giving recommended trends for CDNs development.

Keywords:

Content Delivery Networks, Improvement, Performance, Scalability, Reliability, Response Time

Contents

List of Figures

List of Tables

Acronyms

ACL	Access Control List
ASP	Application Service Providers
BGP	Border Gateway Protocol
CA	Certificate Authority
CAP	Content Access Point
CLI	Command Line Interface
CDNs	Content Delivery (or: Distribution) Networks
CDSP	Content Distribution Service Providers
CPU	Central Processing Unit
CSR	Certificate Signing Request
DSSP	Distributed Site State Protocol
DNS	Domain Name System
DoS	Denial of Service
DRP	Director Response Protocol
ECDN	Enterprise Content Delivery Network
E-Business	Electronic Business
E-Commence	Electronic Commence
E-Learning	Electronic Learning
FTP	File Transfer Protocol
GSLB	Global Server Load Balancer
GUI	Graphical User Interface
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
ICMP	Internet Control Message Protocol
ICP	Internet Content Providers
IDC	Internet Data Centre

IGP	Internal Gateway Protocol
IP	Internet Protocol
ISP	Internet Service Provider
IT	Information Technology
LAN	Local Area Network
LB	Load Balancer
LDNS	Long Distance Network Services
MAC	Media Access Control
MPEG	Moving Pictures Experts Group
MTBF	Mean Time Between Failures
NNTP	Network News Transfer Protocol
PC	Personal Computer
POH	Power On Hours
POP	Post Office Protocol
QoS	Quality of Service
RAID	Redundant Array of Independent Disks
RAM	Random Access Memory
RED	Random Early Detection
RSA	Rivest-Shamir-Adleman
RSVP	ReSerVation Protocol
RTOS	Real-Time Operating System
RTSP	Real Time Streaming Protocol
RTT	Round Trip Time
SLB	Server Load Balancer
SNMP	Simple Network Management Protocol
SSL	Secure Sockets Layer
TCP	Transmission Control Protocol
TTL	Time To Live
UDP	User Datagram Protocol
URL	Uniform Resource Locator

VPN	Virtual Private Network
WAN	Wide Area Network
WAP	Wireless Access Protocol
WML	Wireless Markup Language
WRED	Weighted Random Early Detection
WWW	World Wide Web
XML	Extensible Markup Language

Chapter 1

Introduction

1.1 Motivation

For the last twenty years, computer networks have been growing explosively. They have become an essential part of our infrastructure. Networks are used almost in every area, including business, education, government, even entertainment. At the same time, global Internet has also grown continuously. Internet has grown into a production and communication system that reaches millions of people in all populated countries of the world, which also bring enormous economic margin.

Although this increase in the size, number, and complexity of computer networks has enabled the deployment of many new applications and brought huge benefits, it has a downside as well. As networks become larger, more complex, and support more traffic, the performance of network applications becomes more unpredictable and often unacceptable [1]. For example: the rapid increase in the number of users often causes congestion, which can result in high delays and loss rates in the packet transmission.

Therefore, we imminently need a new technology or method to improve the networks. Recent developments in Content Delivery Networks (CDNs), also know as Content Distribution Networks have caused more and more people's attention. Content Delivery Networks employ many geographically distributed sites to improve the scalability and improves network performance by reducing the client response time.

1.2 Objective

In this thesis, we focus on the following points:

1. Overview the concept of CDNs. Explain the function and methodology of CDNs, and describe the components that make up CDNs and do the market analysis

2. Compare and analyze CDNs to other methods for improving performance and congestion control.

3. Overview different technologies that can be used within CDNs. Compare and analyze various CDNs components and implementations.

4. Select current representative CDNs companies or providers to compare and evaluate CDNs products and services.

5. Describe some main applications of CDNs and do a case study of Cisco ECDN and Volera Velocity CDN for E-learning.

6. Conclude and summarize CDNs features, application criteria, and some recommended trend.

1.3 Thesis Organization

The thesis is organized as follows:

Chapter 2 provides an overview of the concept of CDNs, the reason why CDNs are needed and the goals of CDNs. It discusses how CDNs work and basic architecture of CDNs. It also analyzes different types and various components of CDNs. Finally it provides a brief CDNs market analysis.

Chapter 3 expatiates on the methodology of this thesis. Comparing and evaluating are the core in this thesis. This chapter provides the graphical overview of different content and determines objects to compare and analyze.

Chapter 4 compares and analyzes CDNs to other methods to congestion avoidance and network improving, which includes Congestion Control, Capacity Planning and Quality of Service (QoS), etc.

Chapter 5 overviews and analyzes different technologies and approaches for CDNs components and implementations. It consists on how to build original or caching sites, client redirection, Specified DNS server and Global Server Load Balancer (GSLB), interconnecting within CDNs, security and management of CDNs.

Chapter 6 selects current representative CDNs companies and providers to compare and evaluate their products (Caching Products, GSLB, SSL Accelerators) and CDNs services.

Chapter 7 firstly discusses the objects of CDNs application, then analyzes some main applications of CDNs, and finally, selects real cases of Cisco ECDN and Volera Velocity CDN for E-learning to perform a case study.

Chapter 8 summarizes the analysis mentioned earlier and provides conclusions. These consist of conclusions drawn from Chapter 4, Chapter 5 and Chapter 6, overall evaluation of CDNs (strength & weakness of current CDNs approaches), CDNs applications environment (suitable & unsuitable application for CDNs), and future Issues (hot Points & threats to future CDNs).

Appendices list more information of these CDNs vendors products and service for readers reference.

Chapter 2

Background

In this chapter, we will describe CDNs, why we need, goals of CDNs, how CDNs work, types of CDNs, components of CDNs and CDNs market analysis.

2.1 What are CDNs

Content Delivery (Distribution) Networks (CDNs) are thought of as an overlay network on top of the network infrastructure that is built to deliver content to a distributed audience.

CDNs add management and quality of service to the Internet [4]. They improve Internet performance through caching or replicating content. They provide scalability and availability of network content delivery by distributing many servers across the Internet "close" to customers. Customers obtain content from these edge servers directly rather than from the origin servers. A server is considered "good" for a client based on many possible criteria, such as network distance to the server, network conditions, and the load on the server. [6] CDNs also provide other value-added services, such as secure access to content, etc. CDNs share three integrated technology characteristics [11]:

- A dedicated server network with some degree of shared (cached) content.
- A dedicated and intelligent distribution mechanism to streamline data delivery to the server networks, end users or both.

- A mechanism to intelligently match the requesting user with the most efficient distribution server or distribution path, and in some cases, customize the information distributed to the requesting user.

2.2 The Need for CDNs

Why need CDNs ? Because the Internet is inherently latent and has no guarantees of service quality. The nature of what makes the Internet into the great enabling technology of our lifetime is the same thing that make it flawed in its capability to deliver content. Because Internet is a collection of disparate networks linked together through public and private peering relationships, it has latency because of slow links, slow backbones, many network hops, and even distance [5]. A number of factors affect internet performance, but the primary bottlenecks today occur in four areas [7]:

- The "first mile" – getting the content from a provider's site out and onto the Internet. The first mile can represent the greatest cost to the provider's site operation, as it is often where they have to maintain the majority of their own infrastructure.
- Peering points – those network traffic exchange points where traffic tends to get stuck instead of being delivered efficiently.
- Network backbones – the long-haul segments that move traffic, but often at a snail's pace.
- The "last mile" – just before content reaches the end-user (from the ISP to the modem and finally onto the PC).

The Internet users is increasing every day unceasingly. "In the US alone, 220.4 million people are projected to be online by 2003" [7]. The following Figure 2.1 is the comparative estimates of worldwide Internet users by region from 1999 to 2003 [7].

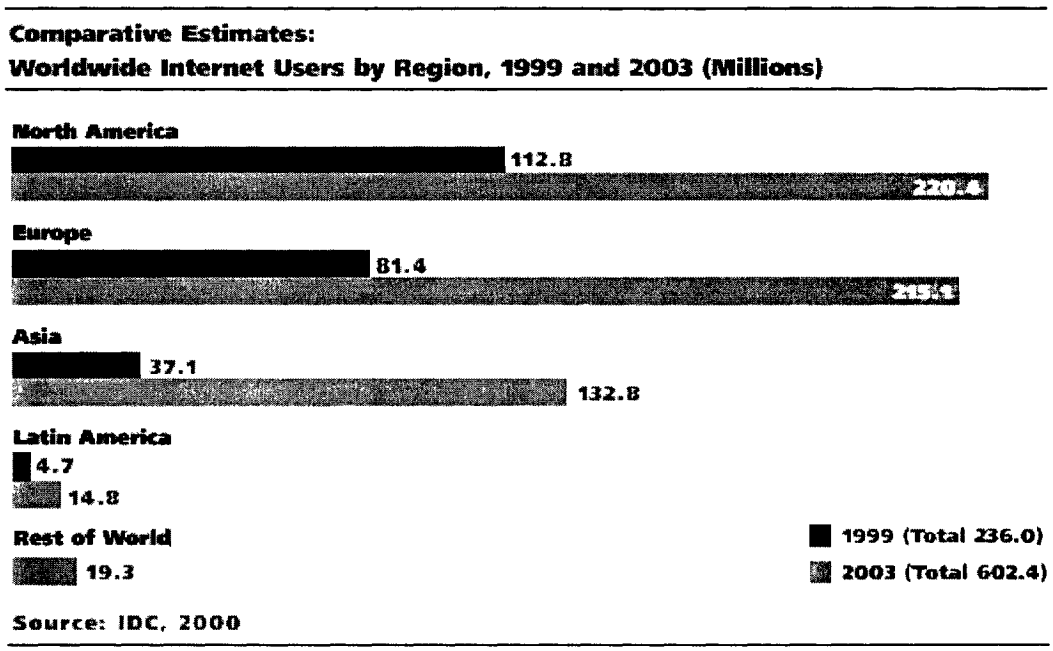

Figure 2.1 Comparative Estimates of Worldwide Internet Users by Region [7]

According to the statistic of Jupiter Communications - Site Operation Strategies 2000 [7]: "The top 10 content sites, currently averaging 12.2 million page views per day, are expected to grow to 27.9 million daily page views over the next four years. Smaller content sites, currently averaging 464,000 daily page views, should easily pass one million daily page views during that time."

Obviously, the congestion of Internet is not likely to abate automatically as the number of users increasing continually. It urgently needs new technologies to change this phases. CDNs were born out of these challenges. CDNs redirect clients get content from proxy servers directly rather than from the origin servers through reasonable determining. They will significantly reduce delivery distance to minimize the response time. For example, if we are in Ottawa and try to visit a website in

Vancouver. In the traditional solution, we have to surf to the original website in Vancouver. The large number of routers in the long distance between Ottawa and Vancouver will increase response time. However, If the website use CDNs which has lots of proxy servers over Canada, one is in Toronto, CDNs will redirect us to the nearest proxy servers in Toronto instead of the original servers in Vancouver. Less routers in shorter distance will let us get data required directly and faster. The following Figure 2.2, 2.3 shows this case without CDN and with CDN. The blue line is the path it goes through.

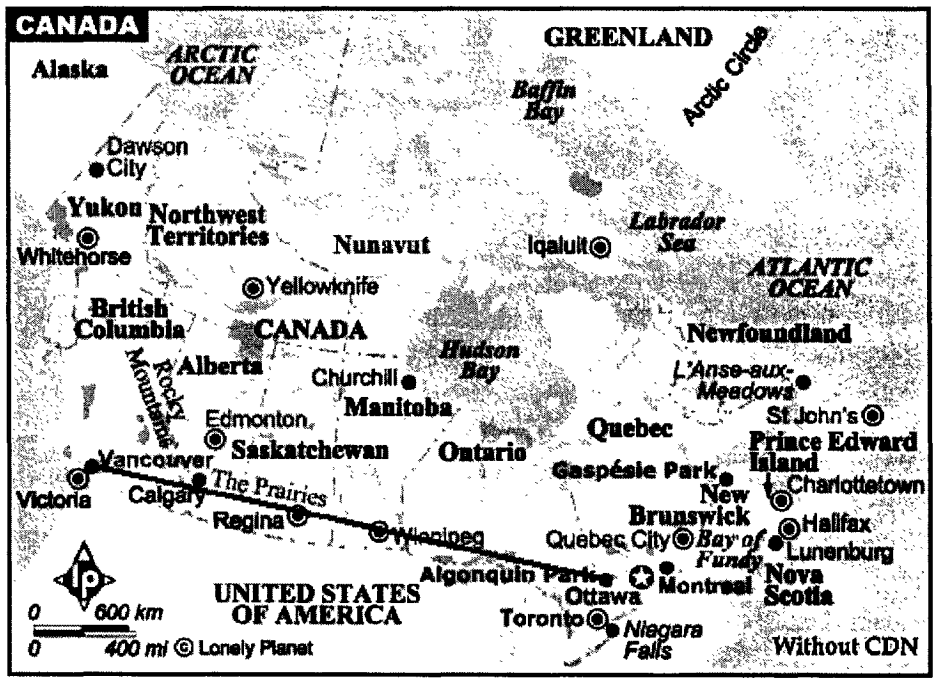

Figure 2.2 Without CDN

Figure 2.3 With CDN

According to the measurement of Keynote System Inc. (www.keynote.com), it shows that network performance significantly improve through comparing CDNs and Non-CDNs. [8]

- **Comparative object download measurements** Keynote did the measurement of download performance of two identical objects (one retrieved from the main Web server and one retrieved via the CDN) at regular intervals of 4 hours (8:00pm, 12:00am, 4:00am, 8:am, 12:00pm, 4:00pm) from April 14th to 17th, 2000. The X-coordinate is the time of day for measurement, the Y-coordinate is the Mean Value (In Seconds) of download the object.

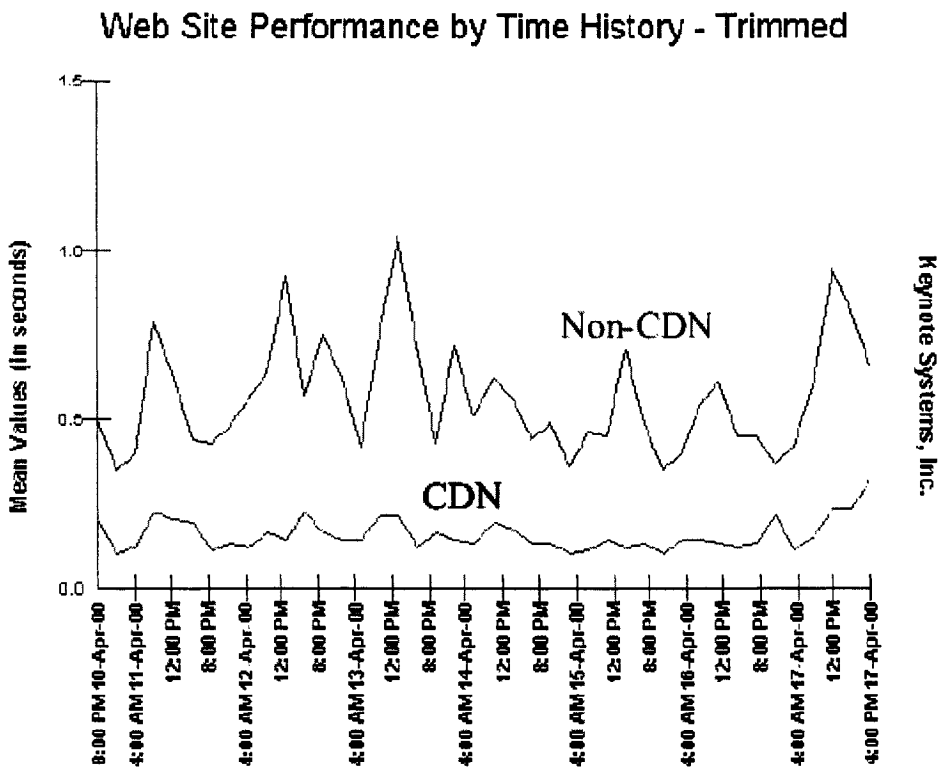

Figure 2.4 Comparative Object Download Measurement [8]

Note there is less download time and relative evenness download quantity by CDN than by Non-CDN. It shows the improvement in download speed and smoothing of peak loads on the CDN-hosted object compared to the same object hosted on the main Web server. This is typical of well-implemented CDN systems.

- **Comparative Error Chart** (Figure 2.5). Reducing network error is very important to ensure network performance, reliability and availability. Keynote also measures and compares error of CDNs and Non CDNs by reporting the percentage of agent requests that resulted in a successful download.

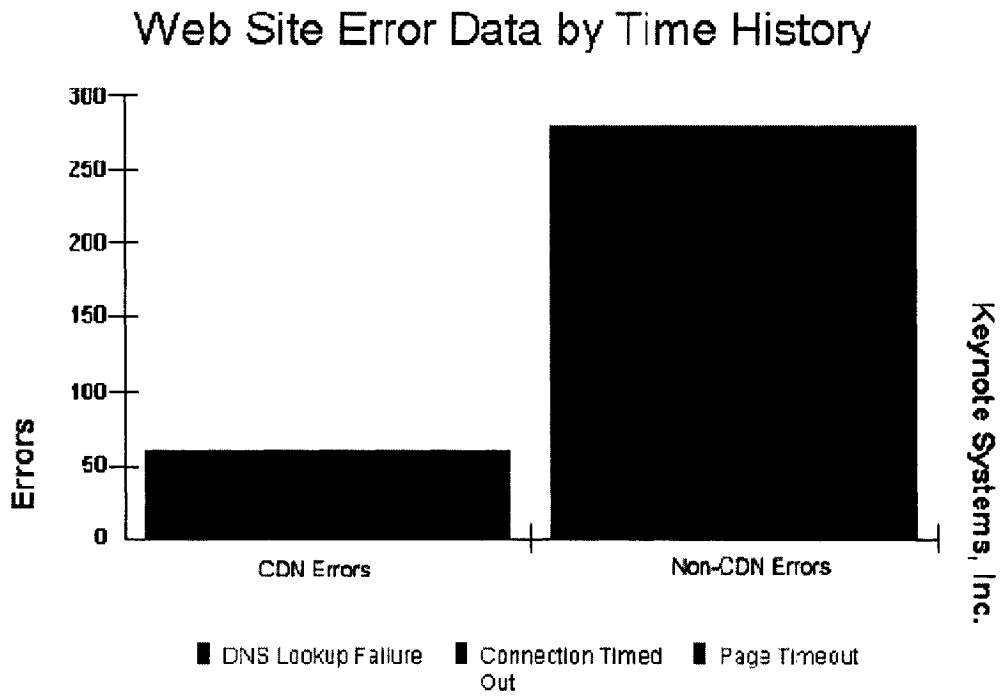

Figure 2.5 Comparative Error Chart [8]

No matter how impressive of the website, it is useless if it can't be brought to the customers. By using CDNs, it not only improves network performance, reliability and availability, but also reduces infrastructure costs and enhances the company competitiveness. [8] "Akamai has found that Web sites using a CDN have been able to increase clicks-through by 20%, reduce abandonment rates by 10-15%, and boost order completion by 15%. "[7]

2.3 Goals of CDNs

The first goal of CDNs is to improve performance. The web pages download faster as the end user's browsers providing them with superior online experience. Faster response time due to factors, such as geographical proximity, network proximity and etc.

The second goal of CDNs is to increase availability. With distributed content around the world at hundreds of servers, failures would be less common and far less damaging than in a centralized content environment. Simultaneously, the web site experiences less downtime as CDNs can route the user's requests among the hundred of CDNs servers connected to various networks.

CDNs also provide the content providers with much greater scalability, and flexibility to grow and change over time. The website performance doesn't degrade significantly as the load on the site increases.

In addition, CDNs can protect the network system by preventing unauthorized accessing, filtering out inappropriate or offensive content and provide support for different types of content distribution.

2.4 How CDNs Work

Most CDNs have similar work processing, despite a few differences in architecture and operating standards. In simple terms, CDNs work through copying, distributing, and storing content on local servers, ensuring the data is fresh, and directing users to

the most appropriate servers. Generally there are three steps as follows to be able to show how CDNs work .

The first step in content delivery is to replicate, distribute and store content on application-specific proxies. Examples are HTTP proxies (for regular Web traffic) and RTSP (Real Time Streaming Protocol) proxies (for multimedia streaming). Stored content can be distributed from the original server to the network edge via the Internet, or satellite. Internet distribution is the simplest and most commonly used method. However, it is satellite distribution that may hold the greatest potential as a cost-effective, high-performance and reliable method. Satellite distribution can broadcast high-bandwidth content to areas that lack of sufficient data communications infrastructure, such as rural or remote areas.

The second step CDNs systems update, store and serve replica copies at the edges of the Internet. Some systems push content into a cache under the direction of a centralized control system, or push updates as changes are made on the origin server. Alternatively, the centralized system directs requests for content to specific servers, and the servers retrieve the content on a needed basis.

The third step CDNs systems direct incoming requests to the appropriate proxies (caching) servers. This routing optimizes access times, and lowers the cost of content delivery. CDNs systems use one of several mechanisms to direct content requests, including HTTP redirection, IP redirection, or DNS redirection.

The sketch map of how CDNs work is depicted in Figure 2.6 [10]

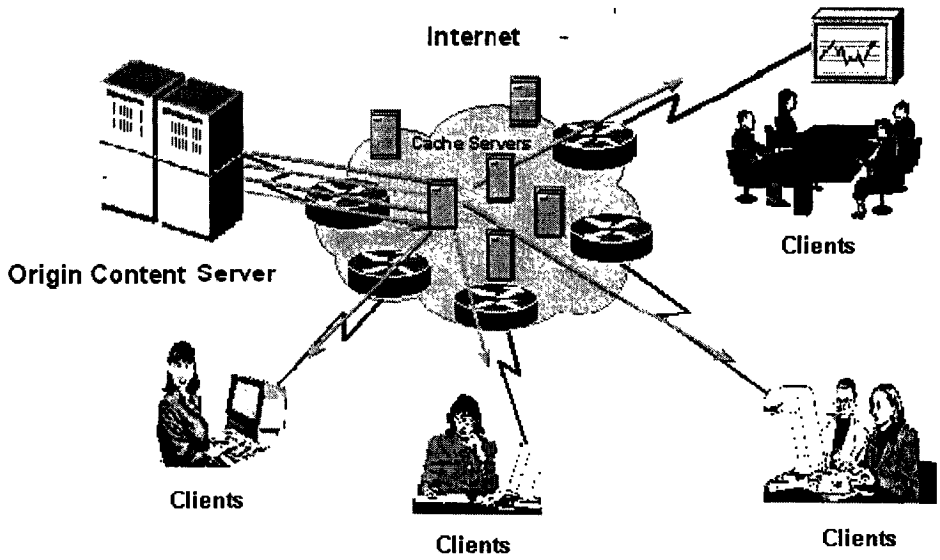

Figure 2.6 Sketch Map of How CDNs Work

These (caching) proxies are located at the edge of the network to which end-users are connected. Therefore each of the nodes in the CDNs is located close to the user (access network), which makes it much easier to adapt to varying qualities of end-user equipment or their preferences. Typically, the CDNs provide the following functions [10]:

- Redirection services to direct a request to the cache server that is the closest and most available.
- Distribution services, e.g., a distributed set of proxy servers that cache content on behalf of the origin server, mechanisms to bypass congested areas of the Internet or technologies like IP-multicast, and replication services.
- Accounting services to handle, measure, and log the usage of content.

Please see Chapter 5 to get more information.

2.5 Types of CDNs

Although the basic concept and service provided by CDNs are essentially the same, there are several different types of CDNs. Three basic types of CDNs are Internet CDNs, Subscriber CDNs and Enterprise CDNs [5].

Internet CDNs is the generic name for the most commonly thought of Content Delivery Networks, which use a distribution of servers or caches at various points close to users. Internet CDNs services are sold to the content providers for ensure their website and content will be readily accessible and perform well. There are three slightly different subtypes within Internet CDNs implement and approach, which are overlay CDNs, Peering CDNs and Hosting CDNs. Each one is briefly described as follows.

- **Overlay CDNs** commonly create a virtual network of servers scattered across many other providers' layer 3 network POPs ("Post Office Protocols: An e-mail protocol used to retrieve e-mail from a remote server over an Internet connection" [12]) and data centre instead of building or maintaining its own. By Overlays CDNs on top of the backbones of other companies, Overlay can provide network value-adding service for the content providers. Akamai is a good example of Overlay CDNs.

- **Peering CDNs** provide the same service as Overlay CDNs, but its equipment is placed at the peering points between the service provider's network and its peers ("Any of the devices on a layered communications network that operate on the same protocol level" [3]) instead of positioning content at POPs or data centres. The approach works well for service providers that have excess backbone speed and capacity, as well as substantial peering connections with other networks.

- **Hosting CDNs** are to use their existing multiple data centres to house content for the CDNs service. In many cases, Hosting CDNs can be much less expensive to build and maintain by using the existing investment in data centre facilities.

Subscriber CDNs are similar to Internet CDNs on technology. However Subscriber CDNs are targeted to different markets. End Users are the primary target market for Subscriber CDNs comparing to Content Owners & Providers for Internet CDNs. End Users will pay small money to be able to get better network services instead of a free piece of standard Internet access. The primary target customers for a subscriber CDNs are consumers and some businesses.

Enterprise CDNs place caching infrastructure in each major enterprise locations and positions content for commonly used items such as files, training presentations, corporate audio or video announcements and any other commonly used content objects. The target market of Enterprise CDNs are enterprises or corporations with distributed campuses in many different locations. By implementing the Enterprise CDNs, corporations can use more robust media, provide faster access to common files, and reduce wide area network (WAN) transport costs.

The following Table 2.1 compares Paying Customer, Cache Locations of Different CDNs Types [5].

CDN Type		Clients	Cache Locations
Internet CDNs	Overlay CDNs	Content Owners and Providers	Multiple data centres, POPs, Peering points of various partners
	Peering CDNs	Content Owners and Providers	Network provider's multiple, Peering points
	Hosting CDNs	Content Owners and Providers	Hosting provider's multiple data centres
Subscriber CDNs		End Users	ISP's multiple POPs
Enterprise CDNs		Enterprise and Corporations	Corporation's central and branch offices within the enterprise network

Table 2.1 Different CDNs Types

2.6 Components of CDNs

The CDNs have five essential components, which are Origin, Caches, Load Balancers/Redirectors, Security and Management. Some items overlap, depending on the specific technology selected. But, conceptually, these five elements are required to provide CDNs services effectively.

Origin is a server or set of servers, which contain the "master" content for that site. Copies of that content are then either pushed or pulled to the various locations around the globe. Changes to the content are only made on the Origin. In the most case, the Origin resides in a data centre and can't be managed directly by the CDNs service providers [5].

Caches act as the non origin content servers. Caches locate in many places within network, hold the content locally, waiting for users to request content. Caches work through a proxy server or set of proxy servers. When the Caches receive a new request, they will check to see if it is available in the cache, if it is, one of the Caches will respond and distribute content to the clients directly, if it isn't, the Caches will contact the origin and get the content.

Load Balancers/Redirectors Load Balancers and some Redirectors, such as Global Server Load Balancer (GSLB) or Domain Name System (DNS) work as a decision maker in the CDNs to select the right site and redirect the client to the most appropriate site. It reduces response time and improves network performance by improving proximity between the client and site. It also improves scalability and off loads the load of the network by finding the least loaded site at a given time.

Security of CDNs is to ensure the security of the original or caching sites, communication between different sites and content in CDNs sites. It includes some relative product and implements, such as Firewall or SSL Accelerator. One of important functions of CDNs is protect the network system by preventing unauthorized accessing.

Management To function effectively CDNs must be carefully managed and monitored [5]. Three important types of management are involved with CDNs: Machines Management, Content Management, and Accounting & Billing Management.

- **Machines Management** includes monitoring link saturation and cache server hit ratios, and expanding or contracting the cache server farm and network

connections based on findings. Good Machine Management also includes verifying that the related equipment such as switches, routers, peering connections, and appliances are all working properly [5].

- **Content Management** involves making many of the decisions discussed earlier regarding pulling content or pre-positioning it. Determining how to ensure consistency scheme needs when content is replicated or cached from the original site to caching site. There are different content management schemes that can be maintained the consistency of the content distributed across multiple CDNs sites depending on the different requirement for the content, such as Periodically Update Scheme, Update Notification Scheme and Consistent Update Scheme (Please see 5.6.2 to get more information).

- **Accounting and Billing Management** is for diagnosing performance problems in order to avoid errors and improve performance of network [5]. In addition, how to decide charger for CDNs service provided is also depending on Accounting and Billing Management. Because the price of CDNs service can depend on various factors, including how many different distribution points are used, how much bandwidth was consumed at those different points, how many errors or delays happened in distribution processing and over what period of time all these factors occurred, etc. Accounting and Billing Management can be used to answer these questions for CDNs service providers and their customers.

In Chapter 5, we will compare and analyze these components of CDNs deeply and extensively. Please see Chapter 5 to get more detailed information.

2.7 CDNs Market Analysis

CDNs market mainly consists CDNs products market and services market. Both of these two CDNs market have significantly increased with amount of content and the number of online users increases. According to HTRC Group, LLC. (http://www.htrcgroup.com) study [9]: "The world-wide CDNs product market will grow from $122M in 2000 to an estimated $1.4B by 2004." Please see Figure 2.7 [10]

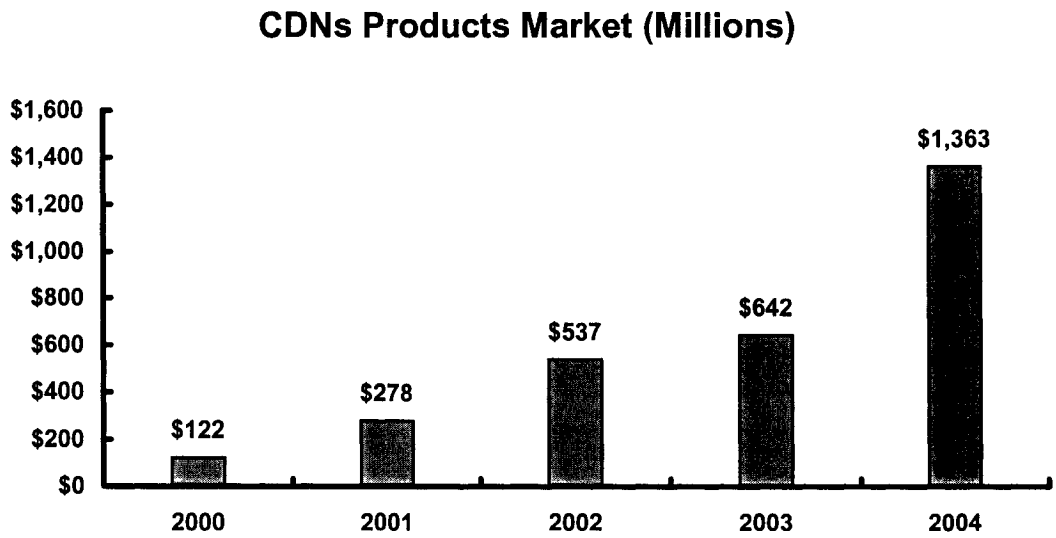

Figure 2.7 World-wide CDNs Products Market Forecast [9]

CDNs service market started off at $62 million in 1999. HTRC Group, LLC. 's Content Delivery Service market study shows: " the new CDNs service market will grow significantly to an estimated $2.2 billion by 2003." The expected growth of the CDNs service market is shown in Figure 2.8 [10]

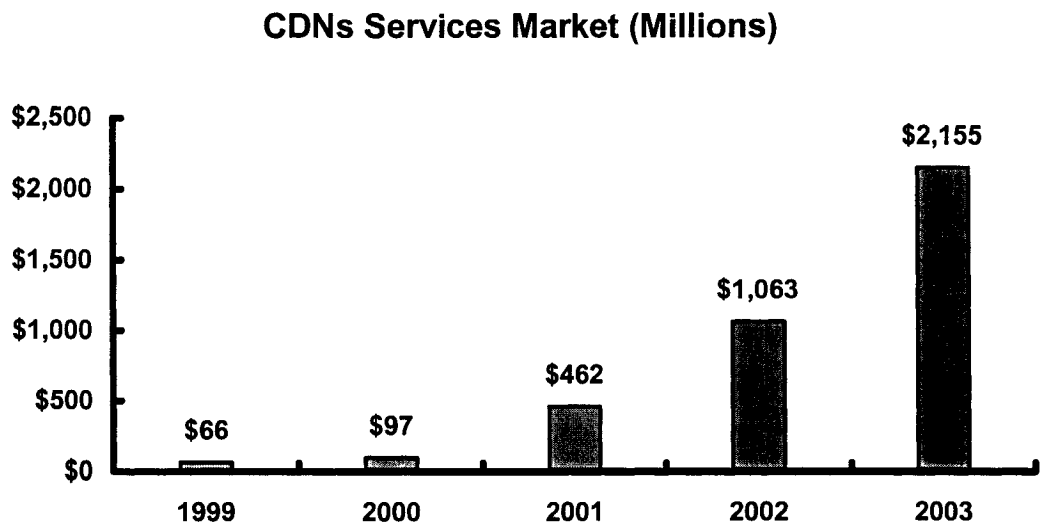

Figure 2.8 Growth Estimation of The CDNs Service Market [9]

The market for CDNs has exploded over the past few years, and will continue to grow in the near future. Since CDNs bring huge margin and business value, more and more companies have been or began to research and develop CDNs technology, produce CDNs products, or provide CDNs services. In the Chapter 6, we will select some representative CDNs companies which are either CDNs product manufacturers, or CDNs service providers, to compare and analyze their CDNs products and services.

Chapter 3

Methodology

The key point of the thesis is describing and analyzing new technology for improving network performance – Content Delivery Networks (CDNs). Comparing and evaluating are the main methodologies used in this thesis. We will divide whole comparison and evaluation into three parts. In these parts, we will select representative congestion control and network improving methods, various CDNs components and implements, or different CDNs products and services to compare and evaluate their features and characters, discuss and analyze their strong points and shortcomings. The best benefit of CDNs is to improve performance, scalability and reliability. At the same time, it will filter out inappropriate or offensive content. So we will use these relative points above as the criteria to compare and evaluate different features and characters, discuss and analyze strong points and shortcomings in this thesis. Because almost all companies regard publishing details of their technology as confidential. We are only able to get limited information, describe and evaluate systems by the general publications written by others, or by CDNs companies' own white paper and product manual.

Figure 3.1 provides a graphical overview of these three parts of comparison and analysis.

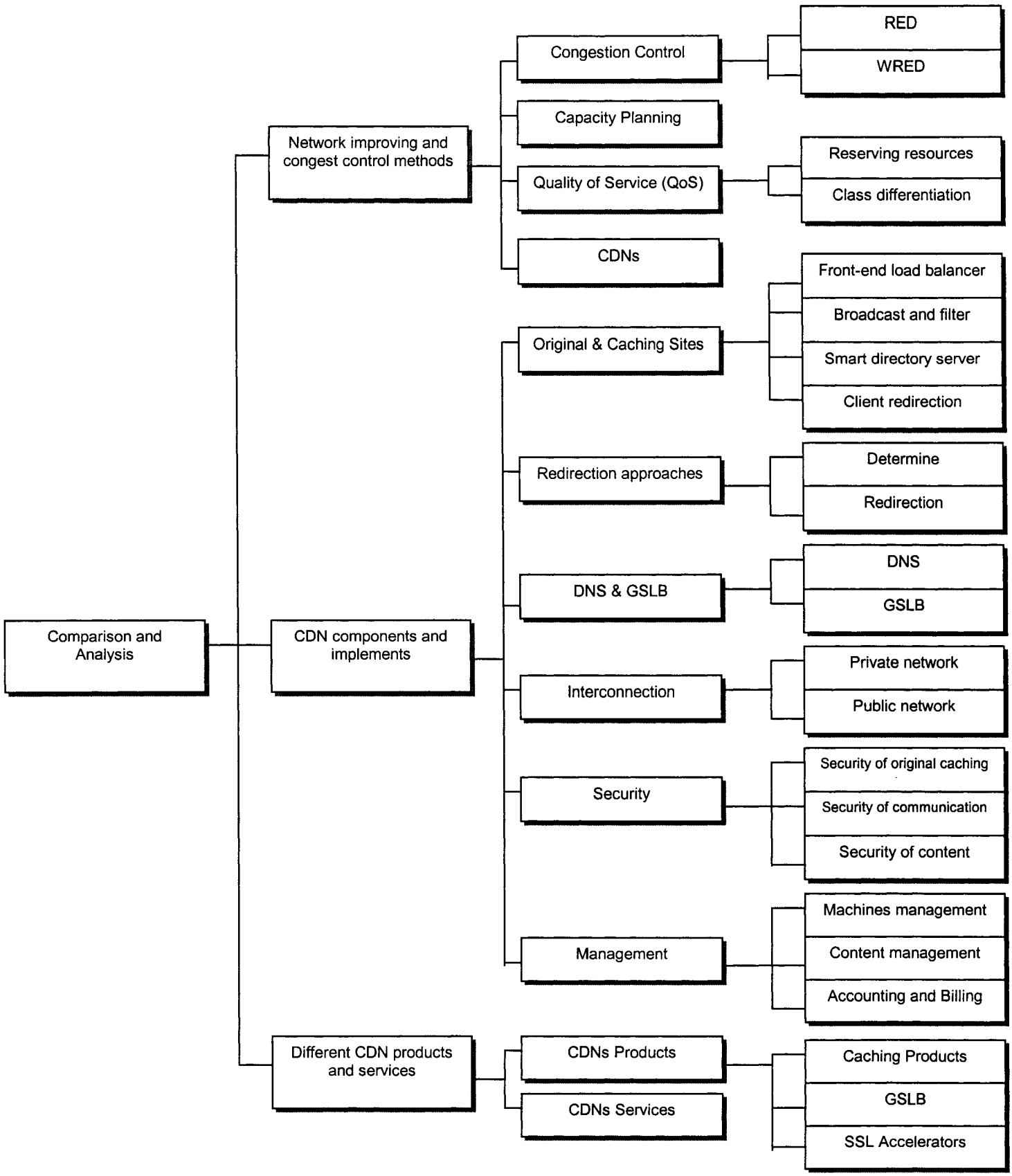

Figure 3.1 Graphical Overview of Three Parts

Part 1

In this part, we will compare CDNs with other congestion avoidance and network improving methods, such as Congestion Control (RED & WRED), Capacity Planning and Quality of Service (QoS) (Reserving Resources & Class Differentiation).

Part 2

In this part, we will compare and analyze various CDNs components and implementation. This part consists six subparts as follows:

- How to build an original site and caching sites
- Different methods for client redirection approaches
- Specialized DNS servers and global sever load balancers
- Interconnection within CDNs
- Enduring security
- CDNs management

Each subpart will be described and analyzed in detail.We also use illustrations and tables to explain them.

Part 3

In this part, we will describe and compare different CDNs products and services operated by current main CDNs companies. We will compare among the alternative commercial offerings from the viewpoints of various components, and their strong points & shortcomings. This part consists of two subparts, which are comparing and evaluating different CDNs products and services.

On one hand, we will select CDNs products includes Caches, Global Server Load Balancer and SSL Accelerators. Table 3.1, 3.2, 3.3 shows these CDNs companies that we select as study objects separately.

Caching Products

The following Table 3.1 shows some representative Caching Products and their companies which we select as study objects.

Product	Company	URL
Array 1000	ClickArry Network	http://www.clickarray.com
SA-7000	Cacheflow	http://www.cacheflow.com
Netcache C6100	Network Appliance	http://www.netapp.com
Cache Engine 590	Cisco	http://www.cisco.com
Inktomi Traffic Server	Inktomi	http://www.inktomi.com
Volera Excelerator	Volera	http://www.volera.com

Table 3.1 Referenced Caching Products, CDNs Companies and URLs

Global Server Load Balancers

The following Table 3.2 shows some representative Global Server Load Balancers and their companies which we select as study objects.

Product	Company	URL
3-DNS	F5	http://www.f5.com
Arrowpoint CSDNS	Cisco	http://www.cisco.com
Distributed Director	Cisco	http://www.cisco.com
Web Server Director	Radware	http://www.radware.com
Alteon Web OS GSLB	Nortel	http://www.nortelnetworks.com
Server Iron	Foundry Network	http://www.foundrynet.com

Table 3.2 Referenced GSLB, CDNs Companies and URLs

SSL Accelerators

The following Table 3.3 shows some representative SSL Accelerators and their companies which we select as study objects.

Product	Company	URL
Array 1000	ClickArray Network	http://www.clickarray.com
iSD-SSL 2.0	Nortel Networks	http://www.alteonwebsystems.com
SA-700	CacheFlow	http://www.cacheflow.com
NetStructure 7115	Intel	http://www.intel.com
e-Commerce 540	F5	http://www.f5.com
SSL-R	SonicWall	http://www.sonicwall.com

Table 3.3 Referenced SSL Accelerators, CDNs Companies and URLs

On the other hand, we will choose top 10 CDNs providers to compare and evaluate their services. In addition, we use illustrations and tables to explain relative content Table 3.4 shows these CDNs providers that we select as study objects separately.

Service Solution	Company	URL
GlobalWise	Adero	http://www.adero.com
EdgeSuite	Akamai	http://www.akamai.com
EdgeServer	CacheWare	http://www.cacheware.com
Streaming Media Service	Cidera	http://www.cidera.com
FireSite	Clearway	http://www.clearway.com
GlobePort	Digital Island	www.digitalisland.com
eXT Technology	EpicRealm	http://www.epicrealm.com
iBEAM Distribution Network	IBeam	http://www.ibeam.com
instaDelivery Services	Mirror Image Internet	http://www.mirror-image.com
Pushcache Push Software	Pushcache.com	http://www.pushcache.com

Table 3.4 Referenced CDNs Service Solutions, Providers and URLs

Please see the detail comparison and analysis in Chapter 3, Chapter 4 and Chapter 5.

We will also list more information of these CDNs vendors products and services on Appendices for readers reference.

Chapter 4

Compare CDNs to Other Congestion Control and Network Improving Methods

There are several methods for eliminating networks bottlenecks, improving distribution and performance within the network. These main methods for avoiding congestion and network improving include Congestion Control, Capacity Planning, Quality of Services and Content Delivery Networks. We analyze and evaluate their strong points and shortcomings separately as follows.

4.1 Congestion Control

Congestion Control is to avoid congestion in Internet based on the situation of networks by relative algorithms or approaches, such as Random Early Detection (RED) and Weighted Random Early Detection (WRED). We overview them as follows.

Random Early Detection (RED) is a congestion avoidance mechanism that takes advantage of TCP's congestion control mechanism [16]. It uses algorithms that are designed to avoid congestion in Internet works before it becomes a problem. RED works by monitoring traffic load at points in the network and stochastically discarding packets if the congestion begins to increase. The result of the drop is that the source

detects the dropped traffic and slows its transmission. RED is primarily designed to work with TCP in IP Internet environments [14].

RED has the following strong points and shortcomings [16]:

Strong points:

- RED reduces the chances of tail drop (Tail drop is caused by a full queue causes, which is dropped packets that could not fit into the queue because the queue was full) by selectively dropping packets when the output interface begins to show signs of congestion. RED drops some packets early rather than wait until the buffer is full.
- RED avoids dropping large numbers of packets at once and minimizes the chances of global synchronization. Thus, RED allows the transmission line to be used fully at all times.
- RED algorithm is simple, which could be implemented with moderate overhead in current networks.

Shortcomings:

- RED randomly drops packets prior to periods of high congestion without caring for the priority level of packets. So it is possible to lead the important packet to be dropped, and unimportant packet to be kept.
- RED statistically drops more packets from large users than small. Therefore, traffic sources that generate the most traffic are more likely to be slowed down than traffic sources that generate little traffic.

Weighted random early detection (WRED) combines the capabilities of the RED Algorithm with IP precedence. This combination provides for preferential traffic handling for higher-priority packets. Packets with a higher IP precedence are less likely to be dropped than packets with a lower precedence. Thus, higher priority

traffic is delivered with a higher probability than lower priority traffic [16]. WRED works by monitoring traffic load at points in the network and selectively discards lower-priority traffic when the interface starts to get congested as well as provide differentiated performance characteristics for different classes of service (Please see Figure 4.1) [14]. However, it is also possible to configure WRED to ignore IP precedence when making drop decisions so that non-weighted RED behaviour is achieved.

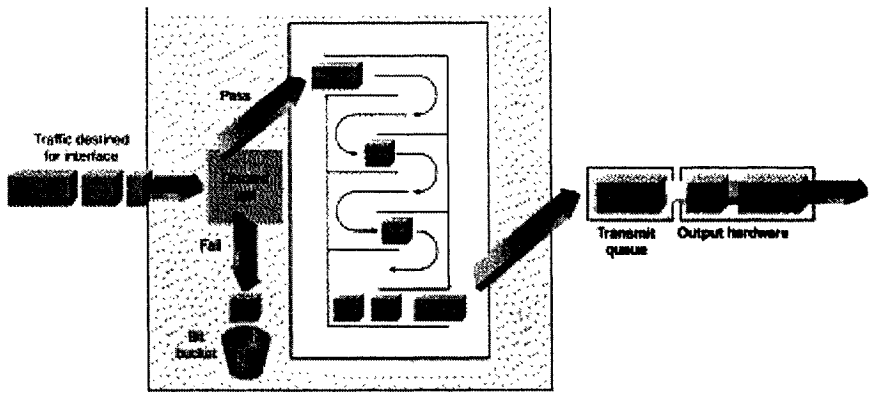

Figure 4.1: WRED Provides A Method for Avoiding Network Congestion

WRED working as a approach for congestion control has several similar features as RED. We analyze its strong points and shortcomings as follows.

Strong points:

- WRED combines the capabilities of the RED Algorithm with IP precedence. It provides separate thresholds and weights for different IP precedence and different qualities of services for different traffic. Standard traffic may be dropped more frequently than premium traffic during periods of congestion [16].
- WRED also helps prevent overall congestion in an Internet works. WRED uses a minimum threshold for each IP precedence level to determine when a packet can be dropped [16].

- WRED is also RSVP-aware and can provide an integrated services controlled-load QoS [14].

Shortcomings:

- WRED is only useful when the bulk of the traffic is TCP/IP traffic. With TCP, dropped packets indicate congestion, so the packet source will reduce its transmission rate. But with other protocols, packet sources may not respond or may resend dropped packets at the same rate. Thus, dropping packets do not decrease congestion [16].
- WRED treats non-IP traffic as precedence 0, the lowest precedence. Therefore, non-IP traffic will, in general, be more likely to be dropped than IP traffic [16].
- Within each queue, a finite number of packets can be housed. A full queue causes tail drops. Tail drops are dropped packets that could not fit into the queue because the queue was full. Although WRED always drop the packets with a lower precedence more than higher IP precedence, when the queue is full the packet discarded may have been a high-priority one, [14].

Summary:

Both RED and WRED are main approaches for Congestion Control. They have some similar features, such as congestion avoidance through dropping various packets. They also have their own characters.

RED is a congestion avoidance mechanism that takes advantage of TCP's congestion control mechanism by randomly dropping packets prior to periods of high congestion. RED tells the packet source to decrease its transmission rate. Assuming the packet source is using TCP, it will decrease its transmission rate until all the packets reach their destination, which indicates that the congestion is cleared [16]. RED can work for congestion avoidance, reduce the chances of tail drop, and

minimize the chances of global synchronization. But RED perhaps drop important packet since it randomly drops packets.

WRED combines the capabilities of the RED Algorithm with IP precedence, which generally drops packets selectively based on IP precedence. WRED can distinguish different priority traffic. Higher priority traffic is delivered with a higher probability than lower priority traffic. WRED also helps prevent overall congestion in an Internet works. But WRED is only suitable for TCP/IP traffic than other protocol, and the packet discarded may have been a high-priority packet if the queue is full. WRED is usually used in the core routers of a network, rather than the edge. Edge routers assign IP precedence to packets as they enter the network. WRED uses the precedence to determine how it treats different types of traffic.

4.2 Capacity Planning

Capacity Planning includes two key points. They are predicting the expected load on the different network elements and furthest reducing the elements to be overloaded. Capacity Planning tries to predict enough capacity to support network link so that least link ever becomes congested, at same time, applications are executed at servers that are fast enough so as to furthest reduce overload.

The key of Capacity Planning is eliminating bottlenecks [1]. If one or more performance bottlenecks happen within the system, the network performance will be depressed. If the bottleneck is a network link, one cannot increase its capacity until it eliminates a bottleneck. Similarly, if the bottleneck is a machine, it can be upgraded to a faster one or augmented by another machine in order to avoid overload.

For capacity planning, one first needs to make an estimate of the total amount of traffic expected within the system. If a reasonable accurate estimate of the traffic at each network link can be obtained, one can then plan to have a new link in place with adequate capacity to satisfy the traffic demand. Similarly, if a reasonable estimate of the processing cycles needed by an application is obtained, one can have appropriate servers or clusters of servers that can satisfy the requisite demand [1].

Capacity Planning has both obvious strong points and shortcomings, we analyze as follows:

Strong points:

- As long as the growth in bandwidth requirements or processing cycles required for any installation can be predicted reasonable accurately, capacity planning can be applied successfully. For environments in which the demands for traffic and processing power changes at predictable rate, one can plan ahead to have adequate capacity within the network and avoid bottleneck.
- As long as there is sufficient link capacity and processing power within the network, most applications can be expected to have a performance reasonably close to their best-case scenario.

Shortcomings:

- The effectiveness of capacity planning depends significantly on the accuracy of the traffic prediction. If the traffic has an unanticipated surge or incorrect expected, the capacity provide within the network may not be adequate to ensure good performance. For example, many online stores or websites perhaps encounter unexpected growth when they have engaged in promotional activity that has suddenly drawn a large number of customers. Because Capacity planning is based on an estimate of how much traffic is expected at a site, and if the estimate proves to be incorrect or insufficient, the performance experienced by the site is likely to degrade.

- Sometimes the traffic load is predictable. But if upgrade in the network capacity cannot satisfy the requirement of fast growing, the bottleneck will perhaps happen. For example, one might be able to estimate that the traffic growth is increasing so rapidly that an upgrade in the link capacity is needed within 3 days. However, if the link provider takes a week to install the new line, the new line cannot be upgraded quickly enough to avoid the anticipated overload.

Summary:

Capacity Planning is fundamentally a technique for estimating the normal operating environment and planning for resources accordingly. In order to plan adequate capacity in advance, accuracy of the traffic prediction is required. Besides this, upgrade in the network capacity cannot be too fast. Obviously, Capacity Planning is not very efficient at taking care of things that are not normal expected.

4.3 Quality of Service (QoS)

Quality of Service (QoS) is another method to depreciate network bottleneck and improve performance. Sometimes, no matter how good a job one does at capacity planning, excess capacity will also perhaps be used up for one or more reasons. The surplus capacity may be used up by new applications coming on line or new users coming on line. So the basic assumption behind the Quality of Service is that one can never have enough resources within the network.

The QoS approach is an attempt to protect the performance of some subset of flows in the network. There are different levels in QoS. According to different level, QoS will give different priority in order to improve performance for important network distribution. Different levels are as follows [14]:

- **Best-effort service (also called lack of QoS)** - Best-effort service is the basic connectivity with no guarantees.
- **Differentiated service (also called soft QoS)** - Some traffic is treated better than the rest (faster handling, more average bandwidth, and lower average loss rate). This is a statistical preference, not a hard and fast guarantee.
- **Guaranteed service (also called hard QoS)** - This is an absolute reservation of network resources for specific traffic.

Two approaches to QoS are commonly used, which are Reserving Resources and Class Differentiation [1]. Reserving Resources is within the network in order to ensure adequate capacity to meet the performance requirement. Class Differentiation provides preferential treatment to some users over others. We will simply introduce them as follows.

Reserving Resources is an application that inform network reserve amount of resources (processing capacity at the node and bandwidth on the links) for special high performance distribution requirement. The resource requirement information is exchanged using a signalling protocol. The signalling is normally done before actual data transfer begins. The signalling processing sends message over the nodes that would be involved in the communication and states a mount of resources that would be needed by a flow. Each node that received the signal will reserve resources or decline this request. If all of nodes along the path accept signal and reserve resources for special network distribution, it will ensure the flow would have good performance. (Please see Figure 4.2) [1][14]

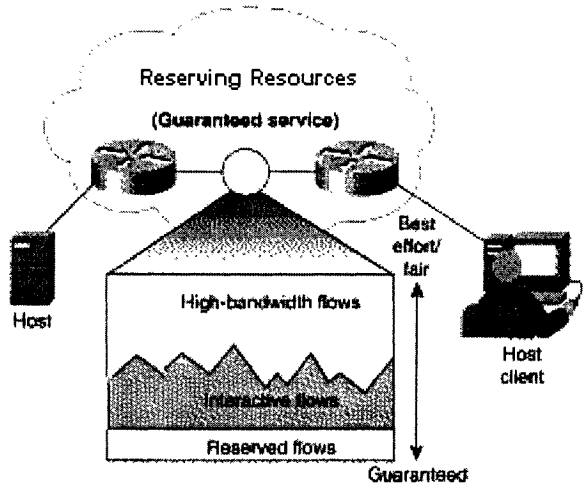

Figure 4.2: Reserving Resources Implemented in Network [14]

Class Differentiation classifies all the packets entering into the network into different classes of service [1]. The simplest instance of differentiated classes is networks that offer different priority level. Priority queuing can flexibly prioritize according to network protocol, incoming interface, packet size, source/destination address, and so on. Each packet can be placed in one of four queues-high, medium, normal, or low-based on an assigned priority. During transmission, the algorithm gives higher-priority class absolute preferential treatment over low-priority ones. So there are always some classes taking precedence over others. When there is congestion at one of the links, the performance of the high-precedence class is less likely to be degraded. However the lower-precedence class is more likely to be degraded. Class Differentiation is useful for making sure that mission-critical traffic traversing various WAN links gets priority treatment. (Please see Figure 4.3) [14]

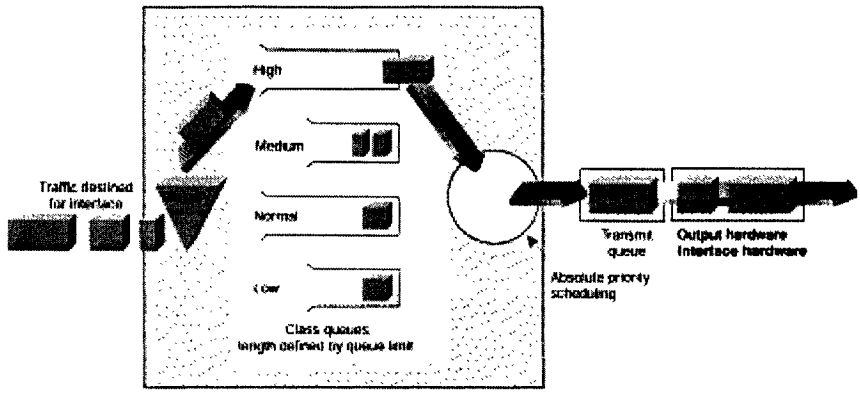

Figure 4.3: Class Differentiation Implemented in Network [14]

After we introduce two mainly applications of QoS to improving network distribution. Then we analyze its strong points and shortcomings.

Strong points:

- Control resources - QoS can control resources such as bandwidth, equipment, wide-area facilities, etc, and give priority to an important database access.
- More efficient use of network resources - QoS technologies can help users know what their network is being used for and they are served the most important traffic to their business. Additionally, QoS makes certain that WAN is used efficiently by mission-critical applications that are most important to the business, and that other applications using the link get their fair service without interfering with mission-critical traffic.
- Tailored services - the control and visibility provided by QoS enables Internet service providers (ISP) to offer carefully tailored grades of service differentiation to their customers.

Shortcomings:

- In reality, it is a daunting task to decide which set of traffic flows should get preferred performance and others should not.

- QoS requires an upgrade of the entire networking infrastructure; all switches need to support reservations or class differentiation in order to be effective. The upgrade needs huge cost and is a significant challenge in any operational network.

Summary:

QoS is a good method to eliminate network bottlenecks and improve distribution performance. QoS has several strong points, such as control resources, more efficient use of network resources and tailored services. Through Reserving Resources and Class Differentiation, etc approaches, QoS avoids and controls network congestion in order to improve distribution performance. However, QoS also has some shortcomings, such as difficult decisions which traffic flow ought to get preferred performance, significant challenge with upgrade entire network infrastructure, etc. These are big challenges for QoS in real deployment.

4.4 Content Delivery Networks (CDNs)

Content Delivery Network (CDN) is also a main method to eliminate network bottlenecks and improve distribution performance. Its key is to avoid the congest links within the network. If there is no congestion within the path of client-server communication, it is likely to have a good performance. If there is congestion within the path, CDNs will let client to get its services from a secondary server (Caching Server) with better performance.

In the Chapter 2, we have already introduced CDNs' components and types. Here we use Figure 4.4 to describe the architecture of CDNs. There are several clients in the network; they try to access the original server across a backbone network. If the path that is used between clients and the original server is congested, clients have to

suffer from poor performance. However, there are several caching servers within the network; clients can obtain all the information they need by contacting the caching servers instead of the original ones. Clients will avoid the congested path and get better performance. Any caching servers only have good contacting with a few of clients, so they need more than one caching servers in networks in order to satisfy requirements of clients in different places. Figure 4.4 illustrates a typical Content Delivery Networks (CDNs). It includes several caching servers located in different places within the network. Each client is directed to one of caching servers, which is the nearest to it, and gets good performance even if the path between the client and original server is congested.

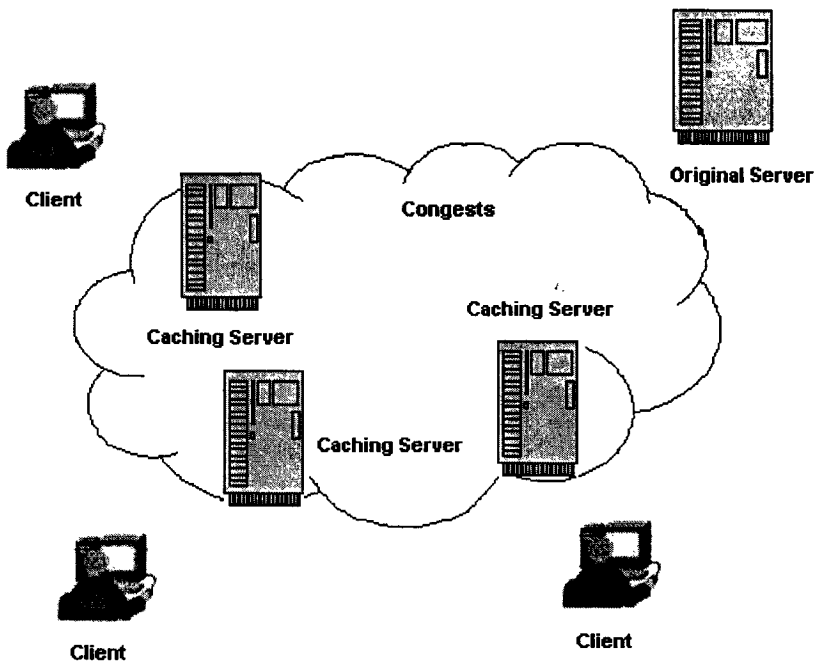

Figure 4.4: Caching Server Instead of Original Server in CDNs

Content Delivery Networks (CDNs) have many significantly strong points. At the same time, CDNs also have their own limitations. We will analyze them separately as follows:

Strong points:

- CDNs approach avoids congestion in the backbone of the network, and also result in better network performance for the clients.
- Caching servers help original server to handle many more clients than only the original server. So CDNs can significantly improve the scalability of network.
- CDNs also improve network reliability. Even when the main server is down, one of the caching servers may be able to provide data to clients. Thereby CDNs will improve the availability and reliability of networks.

Shortcomings:

- CDNs consist of many caching servers that are geographically distributed in different sites. These sites and servers need to be managed, which is maybe a complex and costly work.

Summary:

Content Delivery Network (CDN) is a new technology for improving performance of network, it can reduce the effect of network bottlenecks and improve distribution performance, and it also can significantly improve scalability and availability of network. Although CDN also has its own limits such as strong manageability which caching severs required and a cost challenge for application CDNs, Content Delivery Network (CDN) is a promising technology for performance, scalability and availability improving.

In this Chapter, we compared and evaluated CDNs to other avoiding congestion and network improving methods, Congestion Control, Capacity Planning and Quality of Service. RED and WRED were mainly Congestion Control approaches by dropping packets when congestion begins. Capacity Planning estimated the normal operating environment and planning for resources in advance. QoS worked through Reserving

Resources and Class Differentiation, and differentiates packets into different priority level to avoid network congestion and improve distribution performance. Then CDNs used caching sites instead of original site to deliver content to a distributed audience. These methods mentioned above are all very useful technology in network design for congestion control and network improving, however in this thesis, we mainly focus on CDNs technology. In some case, more than one technology can be combined to apply together. For example: When we apply CDNs to improve network performance, we still need to forecast traffic in order to calculate the number of cache servers to deploy, which is used similarly as Capacity Planning. We also could implement Differentiation Servers within CDNs to ensure that streaming content is delivered on time.

In Chapter 5, we will compare and analyze various CDNs components and implementations.

Chapter 5

Compare and Analyze Various CDNs Components and Implementations

In Chapter 4, we compared CDNs to other congestion control and network improving methods, such as RED, WRED, Capacity Planning and Quality of Service (QoS). In this Chapter, we will compare and analyze various CDNs components and implementations. As described in Chapter 2, CDNs have various essential components, which are origin, caches, load balancers/ redirectors, security and management. This Chapter includes six sections.

In section 1, we will compare and analyze different implementations to build an original site and caching sites. They include front-end load balancer, broadcast and filter, smart directory server and client redirection implementations, each one also include various methods to realize it. For example, there are surrogate severs, network address translator boxes, share IP address and back-up balancer implementations to belong to front-end load balancer implementation. Through comparison and analysis, we try to let readers understand what is the difference between these various methods to build original site & caching sites and improve the scalability of networks. We also use figures to illuminate these implementations and use tables to show the comparisons.

Section 2, we will discuss different methods to client redirection approach in order to explain how to determine the most appropriate site for the client and how to redirect client to the selected site according to former decisions. We use tables to compare these approaches or scheme separately.

Section 3, specified DNS server and global server load balancer are important components for client redirection and load balance of network to improve the scalability and performance of network. In this section, we compare and analyze the strong points the shortcomings of DNS sever and global sever load balancer separately. We also use figures and tables to show the features.

Section 4, we will analyze various interconnection implements within CDNs. There are two common approaches to provide the interconnection of CDNs. One is private network that consists wired and wireless private connection. The other is the public network connection. We also use tables to show their features.

Section 5, we will analyse three aspects of ensuring security of CDNs, which are ensuring the security of original or caching sites, security of communication between different sites within CDNs and security of content in CDNs sites. Ensuring the security of CDNs is the precondition of CDNs reliability.

Section 6, we will focus on the different management issues of CDNs. There are mainly three issues for CDNs management. The first is machines management at the sites of CDNs. The second is content management to ensure the consistency through three schemes to realize this aim. They are periodically update, update notification and consistent update schemes depending on different requirement for consistency. The last one is accounting and billing management, which includes analysis for statistic data of archive and error & urgency situation in order to diagnose performance problems to avoid error and improve performance of network.

5.1 Different Methods to Build Original and Caching Site

In order to support more clients, there are several methods to build original site or caching site to improve network scalability and performance. CDNs consist one original site and multiple caching sites. Each of the caching site and original site may consist of a single server or a large cluster of servers. Here, we analyze various techniques that can be used to build original site or caching site.

5.1.1 Front-End Load Balancers Implementation

The use of a front-end load balancer before a set of servers is the most common technique to cooperate multiple servers in one site [1]. The multiple servers over a high-speed server LAN build an original site or a caching site. Each of the multiple severs all has its address according to network requirement. However there is only one single address advised for the entire site and to client. Clients via external network, first come to the load balancer. The load balancer determines which of various servers at the site should serve client's request. Then request will be forward to that server. Once a client has been directed to a special server of the site, the state information will be reserved in order to redirect the same client to the same server subsequently. The criterion, which the load balancer makes distribution, is based on traffic load among different servers. Some load balancer look at transport and network headers in order to mark their distribution, others look at application-level information to make a more information decision. Now, there are different ways to use front-end load balancer to build an original site or caching site.

5.1.1.1 Surrogate Server Implementation

When the front-end load balancer is used as surrogate server, it represents only one address that is considered as address of entire site to external network. The real web server has its own IP address. Please see the Figure 5.1 as follows:

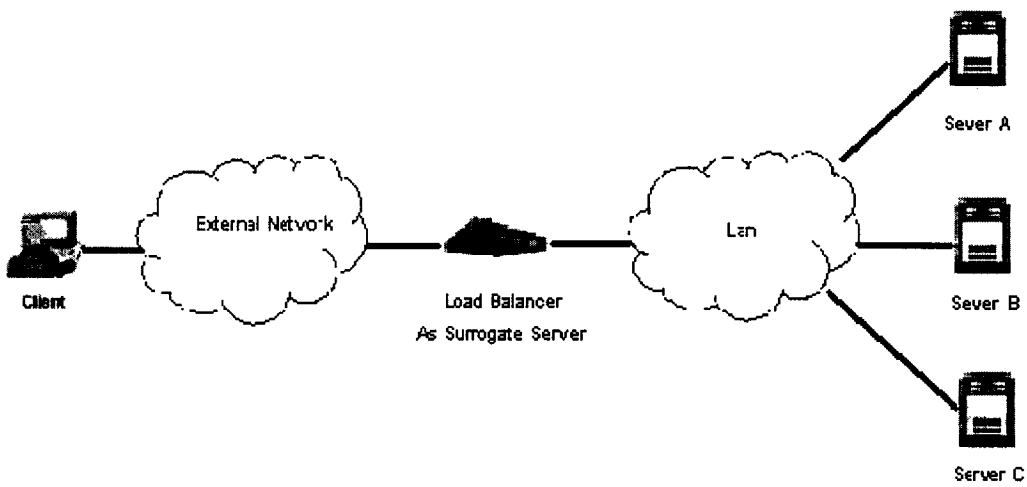

Figure 5.1: Surrogate Server Implementation

When the client wishes to connect to the original site or caching site, they will connect the front-end load balancer first. When the load balancer accepts the connection and completes the TCP processing, then the front-end load balancer, as a surrogate server will decide which of the real server A, B, or C to serve this request. It establishes a special connection between the selected server and client. The load balancer copies any data that it receives from the client to the selected real server. As the surrogate server, the load balancer has the maximum flexibility in determining how to direct the client to different real servers. It can direct client to different servers according to different content they require. For example, if the clients require image

file, they will be redirected to sever A, if the clients require video file, they will be directed to sever B, other clients requirement will be directed to server C.

Front-end load balancer, which is used as surrogate server, has strong points and shortcomings as follows:

Strong points:

The load balancer is transparent between clients and server, which means both of client and server not being aware of the load balancer 's existence.

Shortcomings:

The load balancer has to process all the requests (a large number of packets) coming into the site, so it maybe results in reduced efficiency, and become a performance bottleneck. In addition, load balancer is single point of failure, which perhaps leads failure.

5.1.1.2 Network Address Translator Box Implementation

Beside front-end load balancer could be taken as surrogate server, it can also be as a network address translator box. Network address translator box can be used as a way to handle the network address single exit if site has more machines than the set of legal addresses, which has been allocated on the Internet. In general, The number of packets flowing from the clients towards the servers are much less than the number of packets flowing from the servers to clients. So the multiple network address translator box used will allow the packets flows from the servers to the clients through different path. The way will release more processor cycles to be used for load balancer in order to improve more scalability. Please see Figure 5.2 as follows:

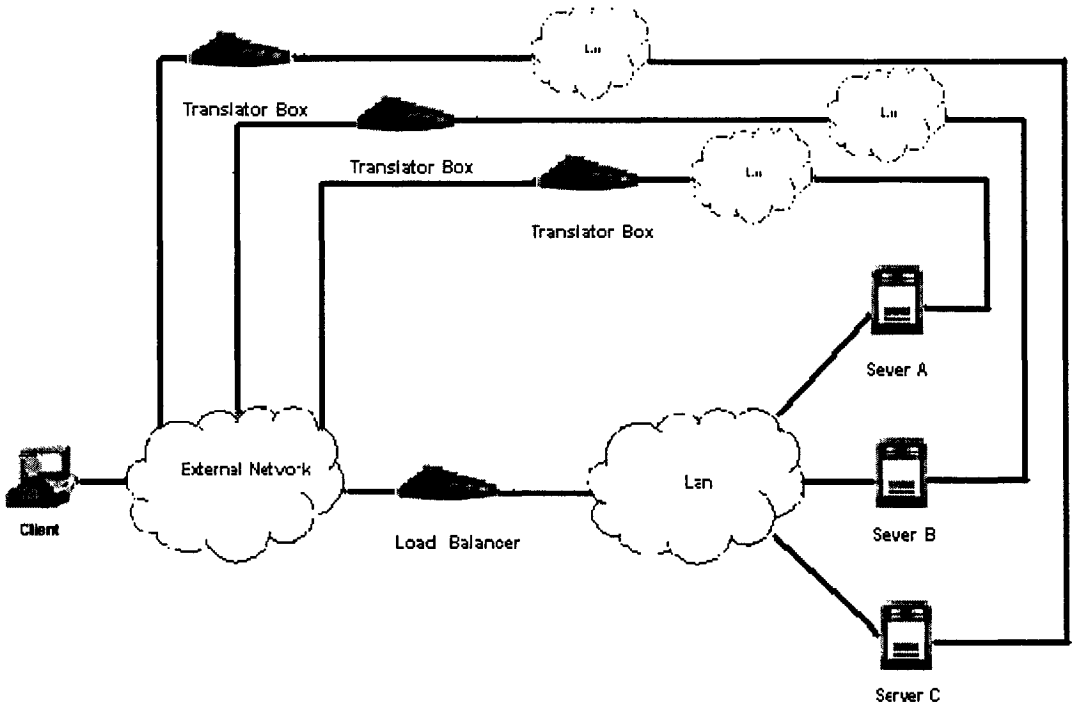

Figure 5.2: Network Address Translator Box Implementation

When the client tries to connect to the original site or caching site, the packets from client will be addressed to the load balancer through external network. The load balancer does not do any TCP processing, just selects one of the back-end servers and forward the packets to it. The back-end server will do the transport protocol processing. Then the packets will flow from one sever to the client through one of choosing translator box. At this point, all the different IP address of different back-end server will be translated to same IP address to external client.

Front-end load balancer, which is used as network address translator box, also has strong points and shortcomings as follows:

Strong points:

The packets needn't flow back through the equipment. It will reduce the work pressure of front-end load balancer, and avoid the network bottleneck.

Shortcomings:

This implementation needs extra boxes, which will lead to additional costs in managing and operating those boxes. In this implementation, although translator boxes are backup, load balancer is single point of failure, which perhaps leads to whole failure.

5.1.1.3 Shared IP Address Implementation

Share IP Address Implementation is a solution to allow all the back-end severs of a site use the same IP address for the external DNS. This implementation no longer needs network address translator boxes. It will simplify the configuration as well as keep the same external IP address. Because all the real servers have the same IP address, so the load balancer has to use other identifier (MAC address) to distinguish among the different real server. The configuration of Share IP Address Implementation is as the following Figure 5.3:

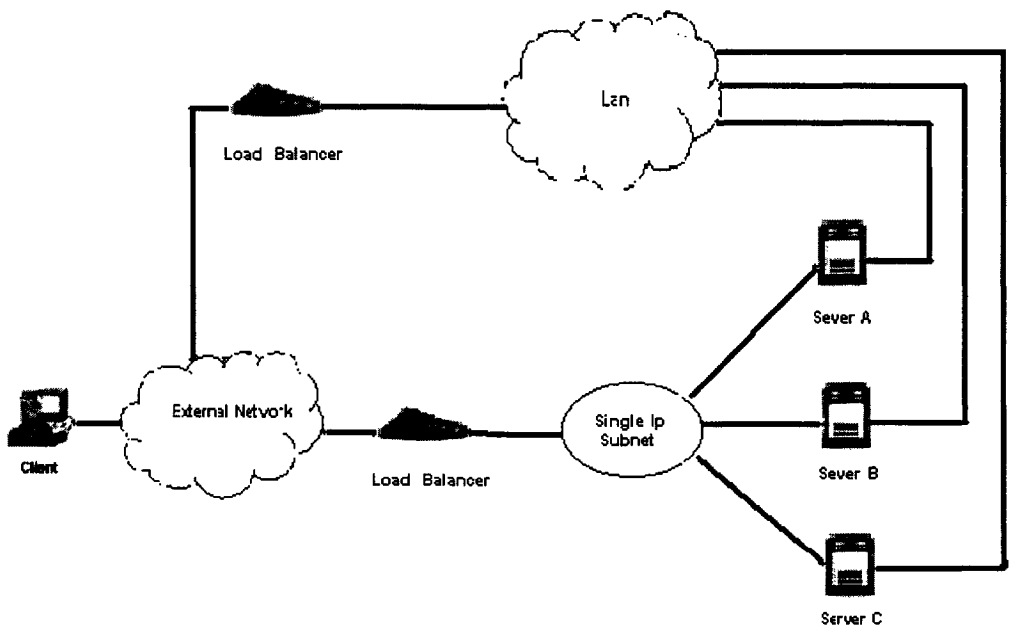

Figure 5.3: Share IP Address Implementation

When the clients access the load balancer, the load balancer will determine the right server to forward the packet to it according to different MAC address. The back-end server, which is selected on the basis of its MAC address, will accept the packet and process, and send the packet out through other load balancer in order to keep the same IP address to external network. This special implementation will improve the efficiency of load balancer.

Sharing IP Address Implementation has strong points and shortcomings as follows:

Strong points:

To use MAC address allows all the web servers use the same IP address; the network address translation boxes on the reverse path can be eliminated. This would simplify the configuration of the web server farm as well as reduces extra cost.

Shortcomings:

Like other single point of failure load-balancing schemes, this implementation also perhaps suffers from the similar failure.

5.1.1.4 Back-up Load Balancer Implementation

When a single front-end load balancer (also called a "single point of failure") is used within a server farm, its failure perhaps can cause the entire site to be inaccessible, although the all back-end servers of the site are good. In order to ensure the efficiency of site when the failure happens, we can use back-up Load Balancer. Please see the Figure 5.4 as follows:

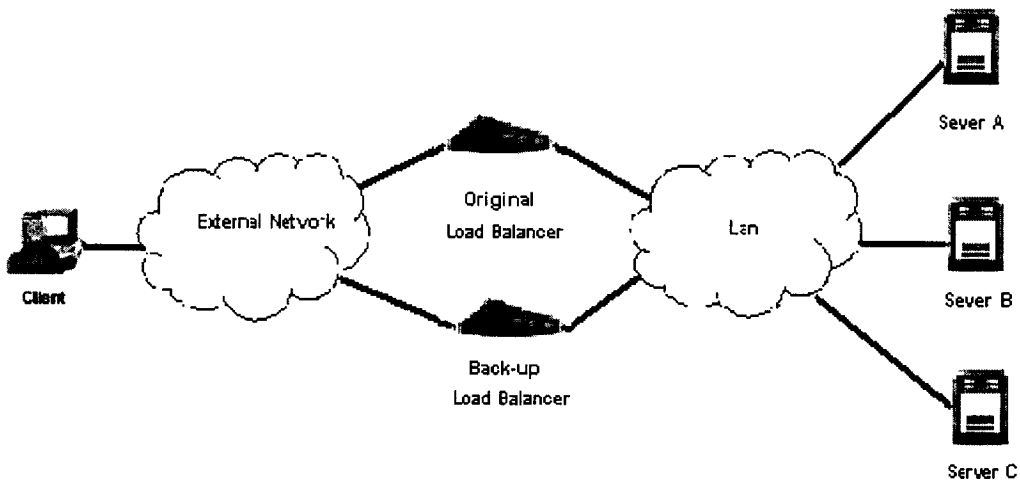

Figure 5.4: Back-up Load Balancer Implementation

When the client wishes to connect to the original site or caching site, they will connect the original front-end load balancer first. However if the original load balancer does not work, the connection will access back-up load balancer automatically. That means the back-up load balancer may operate as a stand-by, in which case it is only used when the original load balancer fails. The back-up load balancer will, instead of original load balancer, accept the connection and complete the TCP processing, then decide which of the real server and forward the request.

Back-up load balancer has strong points and shortcomings as follows:

Strong points:

The back up load balancer guarantee efficiency and operation of site, in case original load balancer fail. This implementation can make the failure minimized.

Shortcomings:

This implementation needs extra load balancer and will lead to additional costs. At the same time, it also needs more effective management and operation.

5.1.2 Broadcast and Filter Implementation

The broadcast and filter implementation provides an alternative method for multiple servers to work together without a special front-end load balancer. In this approach, the request is simply sent to all of back-end server in the site, a collaborative protocol is used among the different back-end servers in order to make only one of them really process the request. The front-end load balancer will be replaced by a standard router with a minor configuration. When the clients wish to connect the original site or caching site, they access the standard router and broadcast request to all the back-end servers. According to the static mapping of back-end servers' IP address and MAC address [2], although each of back-end servers received all the request, each sever only accepts a subset request to future process and filters other request. Please see the Figure 5.5 as follows:

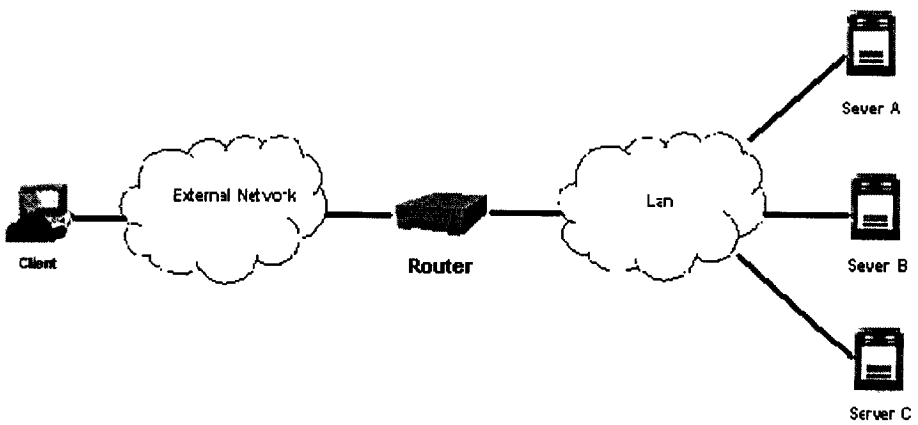

Figure 5.5: Broadcast and Filter Implementation

There are some strong points and shortcomings within broadcast and filter Implementation.

Strong points:

- The broadcast and filter implementation does not require a special front-end processor and therefore avoids a single point of failure.
- There is no load balancer, no need for address translator box, so it will reduce the cost of configuration

Shortcomings:

- Since there is no load balancer, maybe it will lead unbalance to deliver load to different back-end severs.
- Since each sever has to receive all the requests from client coming to site, the servers also have to decide whether to discard or process this request. For all the filter requests, it is an effort waster. As a result, this implementation perhaps cannot handle more traffic as front-end load balancer used [1].

5.1.3 Smart Directory Server Implementation

Smart directory server implementation is the other approach to improve the scalability of network. The key of this implementation is to direct clients to different server machines by translating from machine name to machine address. Before a client connects with sever, it needs to create a directory to translate the machine name to machine address. So different clients will be provided different machines address in order to direct them to different server machines. Replacing of the front-end load balancer, this implementation use domain name sever. It is the DNS-based schemes to improve the scalability of site. There are three mainly methods of the implementation to distribute traffic to different machines.

- The domain name server simply maintains all the public IP address associated with a domain name [1]. The client will use one of these names to initiate the connection. The domain name server can swap the order in which it sends the IP address around for different requests. A client only picks up the first IP address in the list of communication. Different clients will be directed to different servers.
- The site only has a single domain name to the outside Internet. But each sever of the site has its own IP address. When client 's request is translated from domain name to IP address, the domain name server hands out only one of these different IP address.
- The site has different domain name and IP address. One of them is published and used by clients. The others are used inside. Each server of the site has its own domain name and IP address. When a query for the published domain name access, the domain server choose one of the different back-end server' names as the alias of published domain name, then get appropriate IP address for communication.

The primary issue in using different methods is to control how much traffic load is directed at each server on the original site or caching site. The strong points and shortcomings are as follows:

Strong points:

This implementation does not need a separate box in front of the different severs. So the serves can be located in different locations. It has no limitation of zone. Domain name system has other advantages, such as binding the caching of name with address, which result in a quicker response on the name resolution to the client.

Shortcomings:

This implementation is more complex to control how much traffic is directed at each of servers, in some cases it perhaps leads to unbalance load on different back-end severs because not all name servers generate equal amount of load on the site. This

implementation may not be the best suitable for distributing load among multiple local servers.

5.1.4 Selection by Client Implementation

If the client is given more initiative, some of the onus of selecting the right server can be moved to the client. There are mainly three methods of selection by client implementation. We will compare and analyze them as follows.

5.1.4.1 Smart Selection by Client

Smart selection by client implementation will provide clients all the possible severs that they can be used for communication. The client will select one of the many severs which is suitable for communication requirement. The selection can be made either at the stage at which the domain name severs translate the machine name to IP address, or at the stage the actual communication is established. After the clients sent a request out, the request will be received by all the servers at the site and be responded to the clients. The client will select the first come response to future process, and stop communicating with other servers.

5.1.4.2 Client Redirection Protocol

When severs selections entirely rely on the behaviour of the client, there is perhaps a big risk for original or caching site. If the clients are not fully secured, a malicious client can easily misuse the scheme, resulting in unbalanced load on the different servers of the site.

At this point, there is a combine of client selection and server site control. When the client selects a suitable server to establish communication, the site also keeps the

right to control selection decision and redirects the client to new server according to the traffic load of different servers.

5.1.4.3 Concurrent Parallel Access by Client

Concurrent Parallel Access by Client is another option to maintain communication with more than one server concurrently. If the client sends a request to all server, the client will receive information about the multiple servers that are available at a site. Using the Concurrent Parallel Access by Client Implementation can let all the available severs work at the same time. Servers that are light loaded and respond to the client quickly can get a larger share of the requests from the client. And servers that are heavy loaded and respond to the client slowly can get a smaller share of the requests from the client. This implementation can keep the load balanced.

In this section, we compared and analyzed different implementations and methods to build scalable original site or caching site. It included Front-End Load Balancers Implementation, Broadcast and Filter Implementation, Smart Directory Server Implementation and Selection by Client Implementation. They havd their own strong points and shortcomings. Generally, depending on the nature and actual needs, we can combine one or more of the techniques described above to come up with a solution in order to improve the scalability and performance of network.

Here, we use the following Table 5.1 shows how those alternatives compared. It consists four points to compare.

- Impact on congestion: How does this alternative reduce network and/or server congestion
- Cost: cost of equipment

- IP address implications: how are IP addresses organized at the site
- Reliability: what happens if one piece of equipment fails

Implementation		Impact on congestion	Cost	IP subnet implications	Reliability
Front-End Load Balancers	Surrogate Server	Reduces server congestion by choosing a lightly loaded server	Cost of a Load Balancers	IP address of site is IP address of Load Balancers. TCP processing done by Load Balancer	Load Balancers is single point of failure
	Network Address Translator Box	Reduces congestion at both back-end server and Load Balancers	Cost of a Load Balancers plus Translator boxes	IP address of site is IP address of Load Balancers and translator boxes. TCP processing done by back-end servers	Translator boxes is backup, Load Balancers is single point of failure
	Shared IP Address	Reduces congestion at both back-end server and Load Balancers	Cost of Load Balancers	All severs of one site have the same external IP address. TCP processing done by end servers	Load balancer is single point of failure
	Back-up Load Balancer	Reduces server congestion by choosing a lightly loaded server	Cost of origin plus backup Load Balancers	IP address of site is IP address of Load Balancers. TCP processing done by Load Balancer	Load Balancers is back-up
Broadcast and Filter		Reduces server congestion by choosing a lightly loaded server	Cost of standard router with a minor configuration	All severs of one site have the same external IP address. TCP processing done by end servers	No require a special front-end processor and avoids a single point of failure
Smart Directory Server		Reduces server congestion by choosing a lightly loaded server	Cost of DNS	There are various organizations for IP addresses at the site. It can has same or different IP address	DNS is single point of failure. No LB, it maybe lead unbalance
Selection by Client		Reduces server congestion by client select initiatively	Without requiring add cost of equipment	Different server has its won IP address	Risk come from the malicious clients

Table 5.1: Compare Different Methods to Build Original or Caching Site

5.2 Different Methods to Client Redirection Approach

In last part, we have already compared and analyzed different technologies to build an original or caching site and how to direct clients to one of many servers within the site. A content delivery networks consists of many geographic sites [8]. What is the most appropriate site for client and how to redirect client to this site is one of kernel function of CDNs. In this part, we will compare and evaluate different methods to client redirection approach based on how to determine the most appropriate site and how to redirect client to it.

5.2.1 How to Determine The Most Appropriate Site for the Client

Different requirement has different definition, so there are different scheme for CDNs design. On one hand, if CDNs is designed to minimize the response time from a site to a client, so the most appropriate site would be defined by its most proximity to the client. On the other hand, if CDNs is designed to improve the scalability of the site, so the most appropriate site would be defined by least current load at a given time. In this part, we will look at different methods to how to determine which site is the most appropriate one.

5.2.1.1 Active Schemes for Client Redirection

An active scheme is a scheme that generates the extra packets within the network in order to measure response time or load within sites. The type of extra packets is based on the requirement for CDNs.

If the goal of CDNs is to improve the performance of the sites and reduce the response time of a client, it needs to select a site that would be closest to the user. In this case, the performance-monitoring server would monitor the distance of the client from its site and sends this information to the decision maker (Such as a specialized DNS server or a Global Server Load Balancer, Detail information, please see 5.3). The decision maker would then select the site with the smallest distance and then inform the client of the selection.

If the goal of CDNs is to improve the scalability of the sites and handle more clients, the requests of clients should be routed to the site that is the least loaded one. An active scheme would determine the least load site by having the decision maker send a probe message to each of the sites. A performance-monitoring server at each of sites would respond to the probe message by including an estimate of current load at the site [28]. The decision maker could then select the site with the least load.

When the decision maker received the request from clients, it will send a probe message to all the sites in the CDNs. After the performance monitor at the sites gets a probe message, it can apply Internet Control Message Protocol (ICMP) (The protocol that IP uses to report errors and exceptions) echo [2] [3] or Ping (A program used to test network connectivity by the process of sending the request and measuring the response) [2], and other mechanisms to connect the client machine to get the network delay from the client in order to measure the distance to the client. At same time, the performance monitor can measure site's load explicitly. The response to the probe message can also include a measure of how much load the site is currently experiencing. The performance monitor will send such measurement to the decision maker; then the decision maker can choose the site with least delay to the client, the least load of site, or an appropriate weight between the two [1]. The following Figure 5.6 will show how it works.

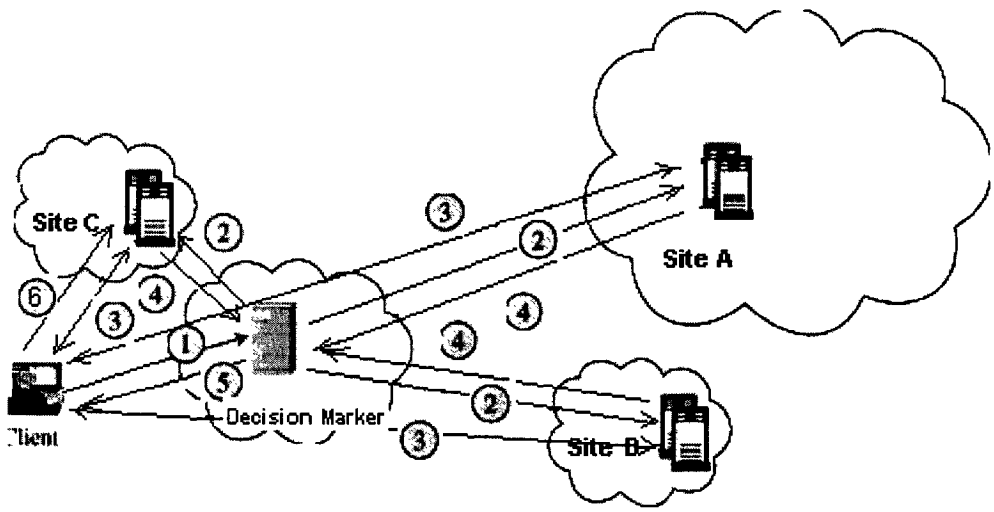

Figure 5.6: Active Schemes for Client Redirection

The strong points and shortcomings of Active Schemes for Client Redirection are as follows:

Strong points:

Active Schemes can get up to date information on network performance. It can determine exactly which site is closest to the client and which site has the least load under the current state of the network.

Shortcomings:

This implement itself perhaps brings delivery delay when the number of CDNs site is large. Active Schemes for Client Redirection includes sending probe message to all the sites of CDNs and collecting all the responses. It maybe generates significant traffic and adds delay time for client to wait for result from decision maker.

5.2.1.2 Passive Schemes for Client Redirection

Passive Schemes for Client Redirection is another scheme to redirect clients to appropriate site by a routing table. Within the Internet, both client machine and sites servers are identified by their IP address. The routing table is typically based on the group of IP address of clients and those of the CDNs sites. When decision maker receives the request, it will determine the most appropriate site and return the location of site to client according to the routing table. The routing table is made up of a routing information matrix that includes the round trip latency between client and the site, the current load of the site, or a combination of them. Routing table has two types, they are static routing table and dynamic routing table. We will analyze them separately as follows.

Static Routing Table

A static routing table is created in a static manner, which can be constructed if the topology of the network the clients would use to communicate with the sites in the CDNs is known. If the connectivity of the clients to the different sites in known, it can determine a priori which sites the client ought to be directed to. The address structure within enterprise Intranets as well as the Internet tends to follows a hierarchical structure. This hierarchy helps in defining the routing table and keeping it manageable. The main task in determining a static routing table is to determine which of the CDNs sites is closest to each of client within the network [1].

An enterprise Intranet usually consists of several campus sites that are connected together by the wide area links. Some of the campus sites host the servers that participate in the CDNs. The other campus sites are the clients accessing the CDNs. Generally, it uses assigning the cost of each link between a CDNs site and a client to

determine which link is closest. The cost of each link can be defined in many ways, e.g., it can be a constant, proportional to the delay on the link, or be a function of link capacity and delay. [1] A constant cost function will minimize the number of links a client traverses to the selected CDNs site. The use of the link propagation delay would minimize the latency between the client and selected CDNs site. Once such a cost metric has been defined and the topology of intranet is known, it is relatively straightforward to compute the lowest-cost path between each CDNs sites and clients within network. This provides the routing information matrix and creates a static routing table, which can be used to select the most appropriate CDNs site for each set of clients within the Intranet.

A static routing table can also be used in the Internet. One concern here is that the subnets that make up the Internet are much more numerous than those within the largest enterprises Intranet. The other more important is not as readily available as an enterprise Intranet topology. It should discover the subset of the Internet topology in order to know the distance from each of the CDNs sites to the different subnets that make up the Internet. There are also many ways to get Internet topology. For example, one way to discover this limited Internet topology is by looking at the BGP (Border Gateway Protocol: The main exterior Gateway Protocol used in the Internet., BGP provides routing among Autonomous Systems [2]) tables of the ISPs that provide connectivity to each of the sites. The BGP routing table contains information about the different subnets known to the ISP routes and number of different administrative domains (AS hops) that are to be crossed to reach that subnet. This provides the routing information matrix from the vantage point of a single site. Combining the information from all the sites and routing each subnet in the BGP tables to the site with the shortest AS hops, it can create a static routing table for the clients on the Internet. [1]

Over a period of several days, the status of network is measured again and updated routing table once depending on the Intranet or Internet dynamics. However, the result would be keep as a static map that could be used to redirect client to the most appropriate site efficiently. It has strong points and shortcomings as follows:

- **Strong points:** This scheme makes redirection easy, just by means of a simple look at static routing table.
- **Shortcomings:** If the Internet delays, congestion, and routing tables continuously change due to variations in traffic and usage pattern. Static routing table that is created by specify time information cannot be as an effective scheme to redirect client to the most appropriate site.

Dynamic Routing Table

A dynamic routing table is a routing table that is continuously updated to reflect the most appropriate site for the clients. All the performance monitors on the sites maintain dynamic routing matrix information. The performance monitors not only base on passive observation of the client's traffic, but also do active scheme by sending probe packets from its CDNs site to the clients. Dynamic routing table should typically be created and modified with information obtained while the CDNs is on line. Resource of clients, performance statistics of each site and packet trace, etc provide the information to build dynamic routing table. There are strong points and shortcomings of dynamic routing table as follows:

- **Strong points:** Dynamic routing table is much flexible and suitable. Since the performance and information of network is always changed, dynamic routing table can work better to redirect client to the most appropriate site by continuously updating according to newest status of network.

- **Shortcomings:** Dynamic routing table also has the common limit of the Passive Schemes for Client Redirection. If a client has visited the CDNs before, the routing table is formed easily. However when a client first visits to the decision maker, no proximity information for the client available, so it seems redirection application can not work as well as former.

We show their features to compare by the Table 5.2, it consists three points.

- Avoid congestion: How to avoid congestion
- Grade of methods: Redirect method is easy or complex to operate
- Flexibility and suitability: The ability to determine the redirection information promptly

Schemes		Avoid congestion	Grade of Methods	Flexible and suitable
Active Schemes		Active through probe message	When CDNs sites is large, it is complex and it maybe generates significant traffic and delay	Has high ability to provide exact and prompt information
Passive Schemes	**Static Routing Table**	Passive through looking at static routing table	After CDNs gets information of client first arrived, the subsequent direction is easy	It is period updated, so it maybe can't provide information promptly if the routing tables need continuously changed
	Dynamic Routing Table	Passive through looking at dynamic table	After CDNs gets information of client first arrived, the subsequent direction is easy	It has high ability to provide exact and prompt information

Table 5.2: Compare Different Schemes to Determine Right Site

More times will be needed in a combination of active and the passive schemes for client redirection. Decision Maker can get information of resource of clients and performance statistics of each site. It can use either static routing table or dynamic routing table to direct client to the most appropriate site. If no any information is found, decision maker also can use active approach or select a default site.

In this subsection, we compared and analyzed different methods how to determine the most appropriate site. Next subsection, we will introduce how to redirect client to this site.

5.2.2 How to Redirect Client to The Most Appropriate Site

Wide area redirection is the emphasis of discussion in this part. We compare and analyze various approaches which is how to redirect client to the most appropriate site. The various approaches mainly are Load Balancer Triangular Routing, Special Directory Server, Client Redirection Protocols and Wide Area Routing. We will analyze them separately as follows.

5.2.2.1 Load Balancer Triangular Routing Approach

Load balancer triangular routing approach is an approach that is a front-end load balancer located at the original site, which selects the most appropriate site to direct client to it. Since this figure of approach like triangular, it is called Load balancer triangular routing approach. Figure 5.7 illustrates how the process operates. The inner cloud in the network is the region that is congested or performed poorly. The client first contacts the front-end load balancer at the original site. The load balancer has selected the site that is closer to the client. It forwards the packets received from

the client to the caching site. The caching site sends the packets back to the client directly. Using triangular routing can avoid client to connect the load balancer multiple times if there is congestion between client and origin. This approach will avoid congestion in the network. With the load balancer triangular routing approach, when the load balancer choose the most appropriate site, it forwards the request received from the client to the selected caching site. When the client receives the notification to change address, it will use address of the selected site, This would then send request directly to the selected site instead of contacting the original site. In the case, the client just need contact the load balancer only once. If there is congestion between client and original site, the request just only access congestion in one round trip.

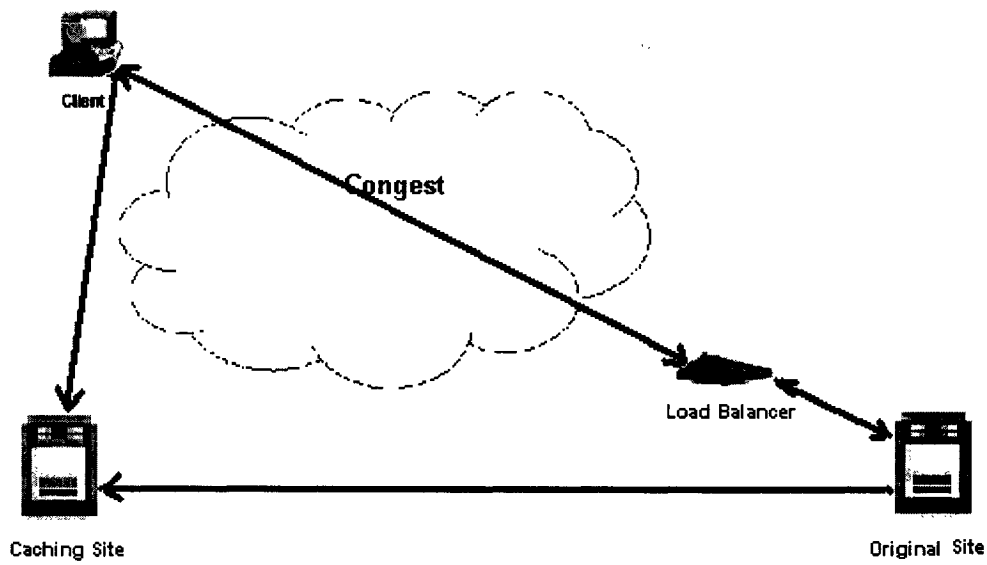

Figure 5.7: Load Balancer Triangular Routing Approach

Load Balancer Triangular Routing Approach has strong points and shortcomings as follows:

Strong points:

Load Balancer Triangular Routing Approach avoid the client access congestion multiple times. It is suitable to apply in the case where content distribution is the primary solution being deployed in the network. Load Balancer Triangular Routing Approach may work as a short-term solution to get some extra processing capacity in a temporary manner [1].

Shortcomings:

Since Load Balancer Triangular Routing Approach can't really increase the capacity at the server site, it is not suitable for a long-term solution. As a long-term solution, it is much better to increase the capacity at the server site than to deploy a content distribution solution.

5.2.2.2 Special Directory Servers

Special directory server approach can map a client to the most appropriate site of the many different sites. In this scheme, the special directory server is used to do the translation from the server name to a network address whenever a client wants to communicate with the server. In the wide area environment, the directory server has to take information of load on the different servers, and additional factors into account, namely, the geographic location of the client. The directory system in the case of the TCP/IP protocol suit is the domain name system, and the use of domain name system in this case means giving out address of the appropriate caching site when a client tries to resolve its name. When the query tries to resolve an address , it comes to a special DNS server (Please see 5.3.1 to get detail information of the

Special DNS Server). The server will choose the right site and provide the its address depending on where the client is located [1][24]. The process is described as follows.

When the client sends request, the query first comes to a special directory server. The special directory server will choose the most appropriate site by active or passive scheme mentioned early and give the address of this site as the correct result. The special directory sever gets the DNS query request, examines which site can make the right answer and returns relative IP address. The client uses it to connect to the appropriate site. If the special directory server gets the request from a client for the first time, it can use active scheme (The process please see 5.2.1.1) to find which site is the closest to client. If the request is more than once and the special directory sever has already query previously, it knows which site is the most appropriate. When the same client sends the request, the special directory sever can redirect the client to the selected site directly. Special directory server also has strong points and shortcomings as follows:

Strong points:

Special directory server supports wide area load distribution without requiring a change in the system infrastructure. In addition, if the special directory server knows the location of the client, it could choose the most appropriate site so quickly and with a reasonable degree of accuracy.

Shortcomings:

In some network protocol, the special directory server cannot know the location of client, it can't decide easily where the query is originated from and perhaps do a random selection first. This can reduce the effectiveness of this approach.

5.2.2.3 Client Redirection Protocols

Client redirection protocol is a way for a server to tell the client that it is redirected to another server at one of the caching site instead. The client redirection protocol can be developed at any of the different levels in the networks stack. In general, client redirection protocol is build by URL rewriting. URL rewriting relies on the fact that most clients would tend to visit a first page, and then follows links in that page to other information at a site [1][23]. For example, when a client visits a site for a the first time, e.g., to the site http://www.cdns.com, the client is given a page in which all the embedded links that normally would be of the form http://www.cdns.com/links is changed to http://www.cdns.com/cache/links. If the client wants to download a graph, the client redirection protocol will change the embedded link http://www.cdns.com/uo.gif to http://www.cdns.com/cache/uo.gif in the most appropriate caching site. So the client can download the graph in the caching site instead of original site to reduce response time. There are strong points and shortcomings of client redirection protocol as follows:

Strong points:

The biggest advantage of the client redirection protocol is that it can redirect client to most appropriate site by URL rewriting to improve network performance. When the client was redirected to the selected caching site by the redirection protocol URL rewriting, the connection will be kept if the selected caching site is always able to serve well.

Shortcomings:

The main issue of this scheme is that it is not transparent to the client, and the clients can see that they have been redirected to a caching site rather than the original site. In addition, client redirection protocol should determine right sites for each client and

satisfy its requests. However, when the number of client increases, there will be a main challenge for client redirection protocol.

5.2.2.4 Wide Area Routing

Wide area routing is to exploit the routing infrastructure within the network to direct the client to the nearest CDNs site. Routing within the IP network is generally divided into two categories: routing with a single administrative domain and routing across multiple administrative domains. These are also known as the interior gateway protocols and the exterior gateway protocols. A gateway is the same thing as a router.

Routing with a single administrative domain is always used in an enterprise environment. If the CDNs solution is going to operate in an enterprise environment where the interior gateway protocols are running. It can exploit the routing scheme to direct clients to the closest server. The operator of the network needs to agree to maintain the subnet to run the CDNs sites as separate entries in its routing table. If so, it can advertise the connectivity information of the different CDNs sites to the routers, which are most closely attached to these sites within the network. The routers will propagate the connectivity information, and each client will be directed to the nearest site.

In more cases in Internet, there are routing across multiple administrative domains within the network. It uses a similar scheme to work with the exterior gateway protocols. The exterior gateway protocols are defined to operate across multiple administrative domains and multiple network operators. At this level, most network operators like to merge the number of destination IP address in order to reduce the size of their routing tables. It can achieve much of the same result by connecting to

more than one network operators at each of the CDNs sites. Most network operators prefer to carry as much of the traffic as they can by themselves, sending packets to another operator only when there is no other choice. Therefore, if the CDNs sites are connected to multiple ISPs at each of the site, there will be multiple paths advertised within the network for these sites, and the clients will be distributed among the many sites to arrive to the nearest ones finally [1].

We analyze its strong points and shortcomings as follows:

Strong points:

Using the wide area routing is very easy to work with existing network infrastructure. It also brings a client to the closest site. Especially when there are small numbers of sites and each site is located in a different geographic location, this approach is very useful.

Shortcomings:

Wide area routing does not take into account the load of sites. So this scheme will not be able to result in spreading the load evenly across different sites. Another shortcoming is that it needs the cooperation of the network operator, which limits the environments under which it can be applied. Especially if there are large numbers of sites and multiple administrative domains are involved, this approach will become more complex. For example, it is unlikely to persuade all of the network operators in the Internet to leave the subnet corresponding to the wide area routing scheme alone, the scheme may not be as effective as in the case when dealing with a single network provider [1].

Here we use following Table 5.3 to show how those alternatives compare. It consists four points.

- Impact on congestion: How does this alternative reduce network congestion
- Cost: cost of equipment
- Advantages: The main strong points of this approach
- Limits: The main shortcomings of this approach

Approach	Impact on congestion	Cost	Advantages	Limits
Load Balancer Triangular Routing	Reduces congestion by Triangular Routing approach	Cost of a Load Balancer	It avoid the client access congest multiple times. It is suitable to work as a short-term solution	It can't really increase the capacity at the server site.
Special Directory Servers	It is DNS-based to reduces congestion by choosing a lightly loaded site	Cost of a DNS	It supports wide area load distribution without requiring a change the system infrastructure. If location of the client known, it can choose the right site quickly	If the location of client unknown; it will reduce the effectiveness of this approach.
Client Redirection Protocols	It is URL rewriting to reduce congestion by choosing a lightly loaded site	Without adding cost of equipment	It can redirect client to the right site by URL rewriting to improve network performance. It needn't add extra equipment.	It is not transparent to the client. It is complex when the number of client increases.
Wide Area Routing	Reduces congestion by interior gateway protocols and the exterior gateway protocols	Without adding cost of equipment	It is easy to work with existing network infrastructure. Especially when there are small numbers of sites and each site is located in a different geographic location, this approach is very useful.	It can't result in spreading the load evenly across different sites. And it needs the cooperation of the network operators, which limits its application

Table 5.3: Compare Different Approach to Redirect Client

In this subsection, we compared and analyzed many different approaches used to redirect client to the most appropriate site from many sites that makes up CDNs. According to different status and environment, we can combine these techniques to come up with a better solution. Here, the sticking point for client redirection is the decision maker of CDNs. It can be used to decide the most appropriate site for the client and redirect the client to the site. The decision maker could be either a specialized DNS server or a Global Server Load Balancer. Next section, we will emphasize on the specialized DNS server and Global Server Load Balancer.

5.3 Specialized DNS Server and Global Server Load Balancer

Both Specialized DNS Server and Global Server Load Balancer work as a decision maker in the CDN to select the right site and redirect the client to the site. Since they are major and important CDN components. So here we use an separate section to analyze them in detail.

5.3.1 Specialized DNS Server

DNS (Domain Name System) Server is used to translate machine names into equivalent IP address [2]. A DNS server responds to a query by looking up the name and returning the address. When a client requests a URL of a website, their browser queries their local Domain Name Server (DNS) to resolve the host name of this website to an IP address. The local DNS performs queries repeatedly, until reaching the authoritative DNS for the given domain. The authoritative DNS replies with one or more IP addresses for the given domain name. Once the local DNS has

the IP address, it responds back to the customer's browser, which in turn opens a TCP connection with the given IP address and downloads the web pages. The local DNS caches the authoritative DNS response and provides the same address for future requests to the domain until the time-to-live (TTL) parameter of the domain's IP address specified by the authoritative DNS expires [18].

Specialized DNS Server based CDNs works through selecting the optimal edge server on the basis of users' local DNS IP addresses. In other words, in the authoritative DNS for CDN, the administrator in advance chooses the optimal edge server farm for local DNS IP address for domain names receiving CDNs service. When users resolve the name for corresponding domain names by local DNS, the authoritative DNS for CDNs informs IP addresses of the optimal edge server farm by the policy of the administrator and do 'Content Routing' [22][29][30]. A number of factors determine which optimal edge server is used in final resolution, such as availability of resources, network conditions etc. Load balancing can be implemented by specifying a low TTL field in a DNS reply. Please see the Figure 5.8

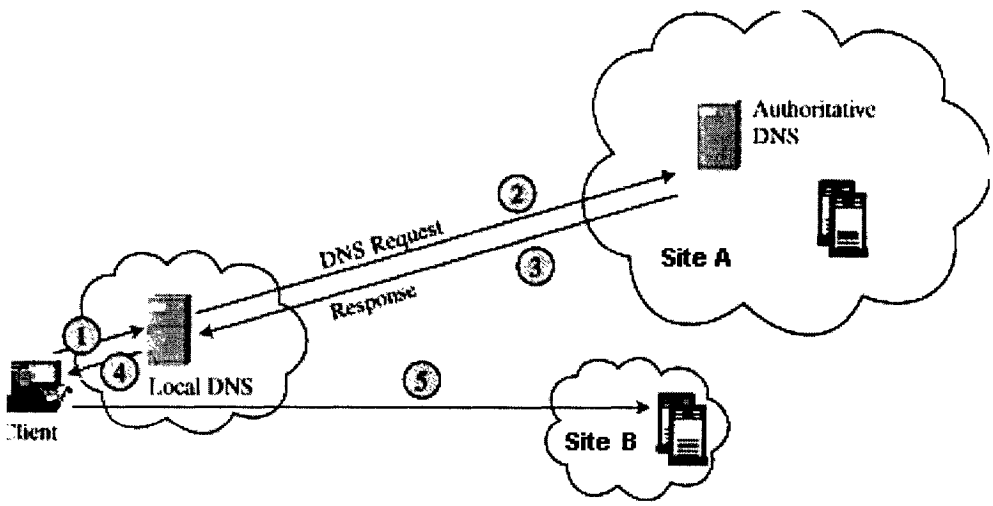

Figure 5.8: Specialized DNS Server Approach [18]

There are some strong points and shortcomings of DNS approach as follows:

Strong points:

DNS can provide a rudimentary form of load balancing if the authoritative DNS uses round robin across the IP addresses for a given domain name. It can be used to allow flexible and scaleable content deployment growth, avoid untraceable performance slowdown and outages.

Shortcomings:

At the same time DNS approach also has some main limits as follows:

- DNS-based Content Routing' technology chooses the optimal edge server farm on the basis of user local DNS IP address, so in case many users share only one local DNS, all users become crowded in one edge server farm. Therefore, this solution is only proper for the case when users are equally apportioned per each local DNS among several local DNSs.
- The update of authoritative DNS is perhaps delayed in some cases. For examples, sometimes when the server supporting the website is down, the authoritative DNS continues to return the server's IP address in response to local DNS queries until the authoritative DNS is manually updated. On the contrary, if the server is working, but the application that serves the website is down, customers receive an error code indicating that the requested website is temporarily unavailable.
- The local DNS runs the risk of providing the customer with a bad IP address for the requested domain because the local DNS caches the authoritative DNS's original response until the TTL of the bad IP address expires - even if the authoritative DNS has updated information containing a new IP address [18][21].

5.3.2 Global Server Load Balancer (GSLB)

Global Server Load Balancer (GSLB) is an attempt to provide load-balancing services to a set of sites within CDNs [3]. GSLB distributes services transparently across multiple web sites and server farm locations, and balances the traffic across those sites/servers on a global basis while monitoring web site/server and application health. By directing the client to the best site for the fastest content delivery, GSLB is deployed to support the multiple and geographical sites. The goal of GSLB is to improve the availability and performance of network. It modifies the response to direct the customer to the best site available. GSLB uses a sophisticated algorithm that consists of several policies and metrics in order to select the best site for a given customer [18]. There are some metrics to evaluate the site IP addresses as follows:

- Site health
- Geographic location of the client and sites proximity
- Response time required to retrieve specified content

When a client enters a URL into the browser for a particular hostname, the system sends a DNS query to their local DNS server, asking for the IP address representing that domain name and host. The local DNS server then examines its DNS cache to determine if it already knows this particular domain name and host. If it doesn't know the hostname, it hands off the request to GSLB. When GSLB receives the query, it will select a site based on the available metrics such as site health, geographic location of the client, site proximity, and response time required to retrieve specified content (Generally, GSLB consists Distributed Site Monitoring or Distributed Site State Protocol (DSSP) to check sites status and exchange health, load, response time as well as throughput information between sites in order to determine the health and response time of servers and applications at each site). GSLB first checks to see

if any healthy distributed sites are present in that region. If there are none, it looks for healthy distributed sites in other regions. Ultimately, GSLB is accomplished by the Authoritative DNS running returning the appropriate IP address to Local DNS servers. Please see the Figure 5.9 of GSLB as follows:

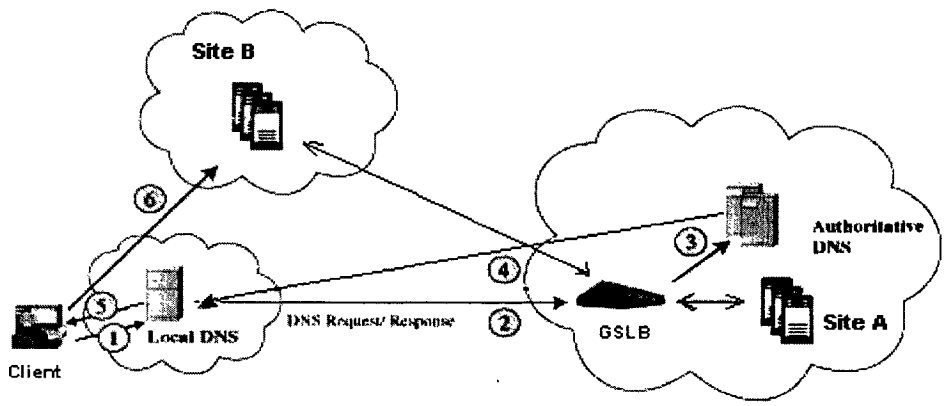

Figure 5.9: GSLB Approach [18]

GSLB also has some strong points and shortcomings.

Strong points:

- GSLB allows clients get content from sites located in their own geographic region automatically. This reduces response time and decreases the use of expensive international data connections.
- GSLB enhances the clients' experience by directing them away from congested networks and servers.
- GSLB Increases fault-tolerance and availability by allowing multi-site content and service deployment, guarding against failures in the event of local or regional network outages, power outages or natural disasters [19].
- GSLB can provide site redundancy and high application availability
- GSLB supports maximum scalability for ultimate flexibility within the network.

Shortcomings:

- In most cases, GSLB is not a cost-effective solution. The cost of purchasing the GSLB platform is significant and not dual-purposed.
- Generally, the content doesn't require a global reach. The performance gains might only be minimal outside of normal Internet operating, which is enough for content.

Specialized DNS Server or GSLB is the key component to redirect client to the selected site in CDNs. We use following Table 5.4 to show how those alternatives compare. It consists three points compared.

- Impact on congestion: How does this alternative reduce network and/or server congestion
- Cost: cost of equipment
- Reliability: The ability to ensure redirecting client to the right site and not overload in site

Implementation	Impact on congestion	Cost	Reliability
Specialized DNS Server	Reduces congestion by choosing a lightly loaded server	Cost of DNS	DNS-based redirection, it can ensure redirecting client to the right site. Without Load Balancer, sometimes, it will lead overload
GSLB	Reduces congestion by choosing a lightly loaded server	Cost of GSLB plus DNS	GSLB have good ability to ensure redirecting client to the right site and not overload in site

Table 5.4: Compare DNS and GSLB

If the original site can't satisfy the requirement of the client, the connection will transfer from original site to the most appropriate caching sites. The caching sites always connect the original site, or each other to cache content and receive the notification in order to serve the request of the client. So we find ensuring the good interconnecting with CDNs sites is very important. Next section, we will analyze different types of interconnection schemes which can be deployed within the CDNs sites.

5.4 Interconnecting Within CDNs

The CDNs consist the original site and many caching sites. The goal of CDNs is redirecting the client to the most appropriate site, which almost is using caching sites to replace of original site. However, when sometimes the caching sites cannot satisfy the client's request since it has no data which is required by the client. The caching site should communicate original site or other caching sites to get the data. So to ensure the interconnecting well between the caching site and the original site or caching site to caching site is very important to improve performance of network. In this section, we will compare and analyze different types of interconnection scheme within the CDNs. There are mainly two big schemes, which are creating private network connection and using public network connection. [1]

5.4.1 Private Network Connection

The simplest and the best performing way among sites is creating the private network connection. And the private network connection can be created by the wired

private network connection or by the wireless private network connection. Both types of network are discussed in the following subsections.

5.4.1.1 Wired Private Network Connection

The wired private network connection is an ideal way to create a physical private network connection between the original site and caching sites. The capacity of the private network connection is obviously faster than the public network. The delay happen on the private network connection is likely smaller than on the public network. The strong points and shortcomings of wired private network connection are as follows:

Strong points:

There is no competing traffic from extraneous sources on the wired private network. So it will ensure network performance and avoid bottleneck.

Shortcomings:

When to build special private line, the cost of approach will be increased. So the limit of wired private network connection is the budget. Operating a wired private network is expensive, and the performance obtained by this scheme may not be worth the price.

5.4.1.2 Wireless Private Network Connection

Another private network connection is wireless private network connection. The wireless private network connection could be a satellite network or a radio links depending on the physical separation between the different sites within the network.

A satellite network uses a geo-stationary satellite in the sky as an intermediate route, which receive the data from all sites and relay it toward to the desired target. Radio links are established along direct lines of sight between the sites. The wireless private network has strong points and shortcomings as follows:

Strong points:

- The wireless private network is a viable option when some of content distribution sites are in remote locations without good connectivity, such as mountain regions or desert where it is more difficult to built wired connectivity.
- The wireless private network can be multicast communication. It allows a site to send the same data to all of the receivers at the same time. The data could be received and processed by all the sites simultaneously.

Shortcomings:

- A wireless private network is not likely to perform better than the wired infrastructure. Even in some cases, it maybe has worse performance than the corresponding connection over the public network.
- This scheme also faces the problem of cost or budget as same as wired private network connection.

5.4.2 Public Network Connection

In many cases, the private network is not viable due to cost reasons. So public networks are always deployed within the CDNs. When it is no special requirement for quality of interconnection, Public Network Connection is suitable and able to observably reduce cost than creating Private Network Connection. Some coding and content adaptation, such that compressing all the data that is being sent over network can be applied to optimize Interconnection with public network.

Public network connection has following strong points and shortcomings:

Strong points:

Since Public Network Connection avoids extra budget to create a new private network. So this scheme will save cost for enterprise.

Shortcomings:

Public Network Connection typically would not have the same good performance as the Private Network Connection, so it is not suitable for some special requirement for distribution performance.

We use Table 5.5 to show how those alternatives compare. It consists three points.

- Methods: How does this scheme to work
- Cost: cost of scheme
- Reliability: The ability to ensure that interconnecting within CDNs is available

Scheme		Methods	Cost	Reliability
Private Network Connection	Wired	It is to create a physical private network connection between the different sites	Cost is very high to build a special private line	It is very reliability but cost is high.
	Wireless	It could be a satellite network or a radio links depending on the physical separation between the different sites within the network	Cost is high to build a wireless private line	It is a viable option when some of content distribution sites are in remote locations without good connectivity, However, sometime it is less reliable than public network
Public Network Connection		Use Public Network. Can use coding or content adaptation to optimize on the Public Network	Cost is low to use Public Network	It is less reliable than private network connection. However it can significantly reduce cost.

Table 5.5: Compare Different Scheme for Interconnection Within CDNs

5.5 Different Approach to Ensure Security of CDNs

More and more organizations are placing an ever-increasing amount of systems and sources on the Internet. At the same time, these increasing systems and sources are potentially vulnerable to malicious attacks. Virus, vicious attacks, network hackers, unconsciously misplays and even unauthorized usage. Each one of these attacks may result in systems broken. Why are attacks so hard to avoid? The reasons are complex. They consist various factors, such as software bugs, anonymity or necessary weakness, etc. Therefore, how to ensure security for CDNs is a very important issue.

The security of CDNs mainly includes three sides. There are:

- Ensuring the security of original site and caching sites of CDNs
- Ensuring the security of communication between different sites within the CDNs
- Ensuring the security of content which is in CDNs sites

We will discuss and analyze them separately as follows.

5.5.1 Ensuring the Security of CDN's Original and Caching Sites

Various attacks are happening in the Internet today, and are likely to continue in the foreseeable future. There are three basic forms of attacks which can bring down a site.

- **Poison Attacks:** A poison attack sends toxic information to the targets. This can take the form of a malformed packet, which the receiver doesn't understand or an oversized packet that exceeds the buffers of the target system.
- **State Attacks:** A state attacks consume resources on a destination site by forcing the site to expend much more effort than the receiver in processing and tracking state information.
- **Capacity Attacks:** A capacity attacks simply use up more capacity than the victim has.

Firewalls and filters are the good ways to prevent these attacks and ensure security of CDNs. The firewall is a security system intended to protect an organization's network against external threats, such as hackers coming from the network. A firewall prevents computers in the organization's network from communicating directly with computers external to the network and vice versa. The firewall allows only limited access to the various sites within CDNs from different location. In the design of CDNs, the firewall should be installed in front of the load balancer to permitted only traffic address to reach the load balancer. Another firewall should be between load balancer and actual severs to ensure only the load balancer is sending packets to the actual servers, and that packets are flowing externally only from selected ports. Each site may also has management consoles or special connection to other sites. So it would require its own set of firewalls. The rules that firewalls use to pass or deny packets are generally called filter, which is applied by anything other than a firewall.

The strong points and shortcomings of the firewall and filter are as follows [5]:

Strong points:

- Firewall and filter can massively restrict access, and then maintain logs of what data is accessed when and by whom. For example, a firewall can be configured to allow access only to and from certain IP address and special ports.
- Firewall and filter can also use log where packets originate, if they were allowed to enter the network or if they were denied. These logs can be used to show what systems might have been improperly accessed, as well as provide a basis for possible litigation.
- Firewall and filter can be used to maintain session state information to examine the traffic flow for all sessions between systems. If a given packet matches the rules database in the firewall and filter or is part of an exiting already-authenticated session, it can be passed through. This is a nice way of neutralizing certain types of IP address spoofing and inserting specialized packets into an existing traffic system.

Shortcomings:

- Firewall and filter also have limit. When the bottleneck or congestion happen in the network, the traffic or packets is likely to delay at these points. Firewall and filter make this worse, not better. They can exacerbate an already bad situation, and improperly restrictive rules will reduce the functionality of the network below required level.

5.5.2 Ensuring Security of Communication between Sites

In order to guarantee the caching sites always satisfy different geography client requirement., caching sites should always update information from the original site or exchange data with other caching sites. The key factor for it above is ensuring the

security of communication between different sites within CDNs, which consists mainly of three ways [1] as follows:

- A private network connection is the first way to ensure security of communication between the different sites. It can be established among the different CDNs sites and used for all inter site communication.
- Virtual private network (VPN) - (A technology that uses encryption and IP tunnelling to communicate on the public network so that their messages are as safe from being intercepted and understood by unauthorized users as the communication were connected by private line.) is the special security technology which can be established between the CDNs sites and run on the public network.
- Special secure transport protocol is also the way to ensure the security. The connections that are established between the CDNs sites should be authenticated and have to use secure transport to ensure the communication [9]. For example, SSL (Secure Sockets Layers: A proposed open standard developed by Netscape Communication for establishing a secure communications channel to prevent the interception of critical information, such as credit number [3]) is to enable secure electronic financial transactions on the World Wide Web as well as work with other Internet services through using public key encryption.

On one hand, private network, VPN or secure transport protocol will ensuring security of communication between different sites, however on the other hand, any type of the above could come at the cost of degraded performance. These will make access between the CDNs sites even more expensive.

5.5.3 Ensuring the Security of Content in CDNs Sites

The goal of CDNs is redirecting clients to the most appropriate sites. So in more cases, clients always access the caching sites and get required data and content. Ensuring the security of content in CDNs sites is the three type of security concern that we discuss in this section. Some content at the caching site or original site are sensitive and require control. For example, with the growing concerns for privacy over the Internet, some private information of client needs to be protected from unauthorized access. There are mainly three approaches to ensure the security of content at the CDNs sites. We will analyze them as follows:

- **Processing only at original site** It means all client requests that deal with accessing secure pages, or that provide sensitive information about the users, such as credit card, should only be handled by the original site. This approach is for the highest level safety required. It doesn't allow any caching site to deal with these sensitive content in order to guarantee enough safety.

- **Qualification authentication** If caching site and original site all belong to one organization, all the CDNs sites can share the credential information of clients and keep in their own sites. When the client access the CDNs sites to get the sensitive content, the sites will authenticate the qualification of that client and check if the client should be provided access to these content. When the client get qualification authentication, the site will provide the sensitive content to the client.

- **Central authorization** When caching site and original site belong to different organization, they may not wish to share their lists of subscribers with each other. So central authorization would be a solution. It means caching site provides

sensitive content to the client only based on the authorization by the original site. All the information of clients is stored in the original site database. The original site validates the credentials of client. When the clients access the caching site to get the sensitive content, the caching sites check the original site to authenticate the user. Once the original site validates the credentials of clients, it will issue a ticket to the clients. This ticket is an authorization to allow caching site to provide what content to the client.

Ensuring security is an important aspect of CDNs design. Depending on the environment in which a CDN operates, we can combine the above approaches to work together in order to ensure the security of original site and caching sites, communication between different sites and content that is at the caching sites or original site within CDNs.

5.6 Different Issues for CDNs Management

CDNs are made up of one original site and several caching sites, which could be distributed in different locations. Sometimes, it is a long distance between them. In order to guarantee hardware, software and content of CDNs operating well, CDNs management should be considered first. CDNs management consists three aspects:

- Managing machines which belong to different CDNs sites
- Ensuing the consistency of content distributed across multiple sites
- Accounting and billing management

5.6.1 Machines Belong to CDNs Sites Management

A CDNs site may consist of various machines, such as several routers, load balancers, DNS, firewall, and servers. The goal of machines management of CDNs is to guarantee CDNs sites to operate efficiently. There are two types machines management for CDNs site. One is local management; the other is remote management.

Local management is through local administrators to operate, configure and maintain various machines in CDNs sites. The benefit of local management is local manned administration, which is configuring in time. The limit is to increase cost of operation, because each CDNs site should have its own local administration.

Remote management is managed remotely. The administrator manages and configures machines remotely through a special private secret path or using password access through common path. The benefit of the approach will save budget because only few administrator can manage and control whole CDNs sites remotely. The limit of remote management is, sometimes, perhaps it cannot configure systems and modify error in time. So in this case, it' d better for CDNs sites to have some self-configuring and management ability in order to ensure system operating well in the case that the remote management cannot be immediate.

5.6.2 Content Management to Ensuring the Sites' Consistency

When content is replicated or cached from the original site to caching site, if the multiple copies of the same content maintained at the different sites are inconsistent, the clients perhaps will be confused and can't get correct information. Thus, content management to ensure consistency scheme needs to be considered for the CDNs.

There are different consistency management schemes that can be maintained the consistency of the content distributed across multiple CDNs sites depend on the different requirement for the content. They are as follows:

- Some content has no strong requirement for consistency. For example, some information has already been out of season or is unimportant for clients. In this case, we can use Periodically Update Scheme to complete this assignment.
- Other contents may not require absolute consistency at all time and can tolerate a little limited time of inconsistency. If the content represents some general information of an individual and can tolerate a little limited time like a couple of days, we can use Updated Notification Scheme to ensure content consistency.
- Finally, some content requires all caching be kept consistent at all time. For example, the balance of checking account of a bank, it should be ensured absolute consistency at all time, otherwise the inconsistency can result in serious financial penalties. We can use Consistent Update Scheme to realize this task.

We will look at them as following subsection.

5.6.2.1 Periodically Update Scheme

This scheme is suitable for the content that has no strong requirement for consistency. The multiple copies of content that are distributed in the network are updated at periodic intervals through pull model or push model. Pull mode is which the caching site pulls content requiring updated from the original site as needed. Push mode is which the original site initiates to content requiring updated distribution to the caching sites. Practical scheme can use either of the two modes or a mixed mode in which the content can be pushed to the caching sites as well as pulled from the original site. Periodically update scheme works well for the content that has no

strong requirement for consistency, is not suitable for the content whose copies must always be kept consistency with each other.

5.6.2.2 Update Notification Scheme

Some of contents have requirement for consistency, at the same time can tolerate a little limited time of inconsistency. Updating notification scheme is a good way to satisfy this requirement above. Update notification is that anyone noticing an updated sends a message to inform all the participants in CDNs concerning the fact that content has been changed. The notification can contain information where the most recent copy of the content can be obtained. Generally on the each update, the original site sends a notification to all of the caching sites informing them that the special content has been changed. Each caching site that receives the update can invalidate any local copies of the record that it may have, and retrieve the new copy from the original site. Sometimes if a special content is only cached at a few CDNs sites, the notification can inform only these sites instead of all the CDNs sites in order to save network resources and cost. Update notification scheme works well for the content record that can tolerate a little limited time of inconsistency. It can't guarantee consistency of the different copies of content; it only simply reduces the inconsistency.

5.6.2.3 Consistent Update Scheme

For some content that cannot tolerate any amount of inconsistency, it needs absolute synchronization and consistency. In this case, consistent update scheme can be used. In the preparation phase of update, the original site sends out a message to all of the caching sites within the CDNs. Each caching site will obtain the appropriate

locks for performing the update, and sends a positive acknowledgement back to the original site. If one of the caching sites is unable to obtain lock, it will send a negative acknowledgement to the original site, so update will be aborted. If all caching sites sends positive acknowledgement back to the original site, the original site will ask each caching site to perform the transaction of update. When update of each caching site complete, the caching site also send a positive acknowledgement to the original site. If the original site receives the positive acknowledgement from all of the caching sites, it will ask all of the caching sites to release the locks as a final step. Consistent update scheme uses Locks to ensure the content consistency.

5.6.3 Accounting and Billing Management

Accounting and billing management are for diagnosing performance problem in order to avoid errors and improve performance of network. It consists of collecting statistic data in the archive to analyze for a long-term capacity and resource planning, as well as process the error and urgency situation happened in the network. We will discuss and analyze them as follows:

5.6.3.1 Statistic Data of Archive Analysis

In the routine case, accounting and billing management collects various types of statistics data, such as user access frequency and pattern, the performance of the system, and the resource utilization of the different machines in the sites, etc. These data need to be kept in an archive so that it can be analyzed and processed subsequently. The archiving is always done at the original site, the caching sites

send statistics data to the original site for accounting and billing management. There are two approaches for the caching site to send statistic data to the original site.

- Caching sites maintain the statistic data locally and send it to the original site in a batch at regular intervals. The benefit of this approach is that it can avoid rush hour of network. It can send when the network is not loaded. However, it also has limit such as it maybe loss some data if the CDNs site fails.
- Caching sites send the statistic data to the original site as soon as archive is generated. This approach is more reliable than the first approach. But it maybe slower because it need write and send content in real time.

5.6.3.2 Error and Urgency Situation Process

When any errors or urgency situation happen in the CDNs, they need to be reported back to the administrator promptly and be processed as soon as possible rather than in an archival mode. They cannot wait for the future analysis and process. There are three approaches for the collection of errors and urgency situation in real-time [1]:

- The first approach is for a program at the original site to periodically query the caching site for any errors or urgency message. The advantage of this approach is that it can keep the balanced load on the original site to avoid the over laden congestion in the network. The drawback of this approach is that each caching site has to wait for diagnostics from original site, some errors or urgency situation may need immediate attention but will still have to wait until the original site connects again.
- The second approach is for an accessing and invoking program at the caching site to send the error or urgency message to the original site directly. In this case,

it has the benefit to allow for immediate handling of the error or urgent message, but the drawback is that it is dangerous. Because several caching sites can overwhelm the original site by generating too many messages at the same time.

- The third approach is a combination of the above two. In this case, the original site using polling collects most of the error and urgency message. However, in the case of a condition that needs to be handled urgently, the caching site sends a notification to the original site requesting that it be polled quickly. The original site will poll the caching site as quickly as it can. This third approach offers an intermediate compromise of two approaches above. It not only avoids caching sites to wait for so long but also it can't result in overwhelming the original site.

In this Chapter, we compared and analyzed various CDNs components and implementation. It included how to build an original site and caching sites, different methods for client redirection approaches, specialized DNS server and global sever load balancer, interconnection within CDNs, enduring security and CDNs management. We discussed and analyzed them with several sections. In the first section, there were front-end load balancer, broadcast and filter, smart directory server and client redirection implementations to build an original site or caching sites. In the second section, it was divided into two parts. One subsection was using active or passive scheme to determine the most appropriate sites for the client, the other subsection was using load balancer triangular routing, special directory servers, clients redirection protocols and wide area routing, etc, to redirect the client to the most appropriate sites. In the third section, we compared and analyzed and strong points and shortcomings of specialized DNS sever and global sever load balancer respectively. On the forth section, there were private network and public network connection to ensure the interconnection of CDNs well. In the fifth section, we looked at three sides of ensuring security of CDNs, which were ensuring the security of

original or caching sites, security of communication between different sites within CDNs and security of content in CDNs sites. In the Last section, there were mainly three issues for CDNs management. The first was machines management at the sites of CDNs, the second is content management to ensure the consistency, and the third was accounting and billing management. In Chapter 6, according to the analysis of components above, we will select some main current CDNs companies or providers as examples to compare and evaluate their products and service solutions.

Chapter 6

Compare and Evaluate Some CDNs Products & Services

In Chapter 5, we compared and analyzed various CDNs components and implementations. This Chapter is divided into two sections. In the first section, we will select some main current CDNs companies as examples to compare their products that are main and necessary for CDNs design, such as Caching Products, Global Server Load Balancer and SSL Accelerator. In the second section, we will compare and evaluate the services of some representative CDNs providers. We also list the more information of these CDNs vendors products and services on Appendices for readers reference.

6.1 Compare Different CDNs Products

Content Delivery Network (CDN) is essentially an overlay network of customer content, which redirects the clients to the most appropriate caching site by the server load balancer estimation and selection, at the same time, it also needs security products to ensure network operation safety and security. In general, the clients obtain content from these caches directly rather than from the origin, a server is considered "the most appropriate" for a client based on many possible criteria, such as network distance to the server, network conditions, and the load on the server, which will be determined by server load balancer. Currently, there are some

companies that produce relative CDNs products, such as Nortel Networks, Cisco, ClickArray Networks, etc. We will compare their products separately as follows.

6.1.1 Compare Caching Products

Many caching products exist in the market today. We can't cover all of them. We just select some representative caching products as the examples to analyze their strong points and shortcomings. These caching products include Array 1000 (ClickArry networks), SA-7000 (Cacheflow), Netcache C6100 (Network Appliance), Cache Engine 590 (Cisco), Inktomi Traffic Server (Inktomi), and Volera Excelerator (Volera). Appendix I lists more features descriptions of these caching products mentioned above.

Array 1000

The Array 1000 is the product of ClickArray Networks Inc., which integrates Web traffic acceleration, management and security components in a single device. It combines server load balancing, global SLB, reverse proxy caching, clustering and web server security (firewall functionality), SSL (Secure Sockets Layer) acceleration and content rewriter in order to improve web traffic performance (Please see Appendix I.1 to get more information). Array 1000 has the following strong points and shortcomings:

- **Strong points:** Array 1000 is an integrated web service device, so it can effectively reduce hardware costs, save space, easy to set up and simplify management. It is suitable for service providers or small to midsize networks.

- **Shortcomings:** Since Array 1000 integrates essential Web services into a single device, it can create a single point of failure if the appliance is not clustered. In addition, although Array 1000 integrates SLB, SSL acceleration function accompany caches, these additive functions can't be as strong as independent products. So Array 1000 might not be suitable for large enterprises with complex network architectures and which require high levels of security and redundancy.

SA-7000

SA-7000 is a product from Cacheflow, which uses a disk for holding content and integrating the optional SSL accelerator and content rewriter function. SA-7000 is the solution specifically built to offload content delivery and encryption tasks from Web servers (Please see Appendix I.2 to get more information). The strong points and shortcomings of SA-7000 are:

- **Strong points:** SA-7000 is an integrated device, it will save IT resource and easy to install. It reduces response time for Web site users and improve the scalability of site .
- **Shortcomings:** SA-7000 is deployed "in front of" any Web server and integrates lots of functions into a single device. This perhaps leads inaccessible if it is a single point of failure. Since the price of SA-7000 is high, it is not suitable for small enterprise application.

Netcache C6100

The NetCache C6100 is the product of Network Appliance Inc., which is a content delivery appliance intended to use as an edge accelerator for ISPs and Enterprise

corporations (Please see Appendix I.3 to get more information). NetCache C6100 has the following strong points and shortcomings:

- **Strong points:** NetCache C6100 can support major streaming media formats. It is suited for both video-on-demand and live stream splitting ("Live stream splitting refers to the ability of filters to obtain information from the Internet once, then makes it available locally on a Windows Media Technology server for access by other clients." [82]).
- **Shortcomings:** Since NetCache C6100 is not pre-load content, so its peak throughput is limited.

Cache Engine 590

Cache Engine 590 belongs to the Cisco Systems Inc., which uses a disk to store content. It is integrated into the network infrastructure for ISPs and larger enterprises (Please see Appendix I.4 to get more information). The Cache Engine 590 has the following strong points and shortcomings:

- **Strong points:** Cache Engine 590 minimizes redundant network traffic that traverses WAN links. It is good for bandwidth optimization. With Web Sense Enterprise software, Cache Engine 590 enables administrators to monitor, manage and report accessing content, which can increase employee productivity, lower bandwidth usage and reduce legal liability.
- **Shortcomings:** Cache Engine 590 can't support streaming media and HTTP 1.0. It also has low peak throughput, only 25-40Mbps. In addition, it needs additive equipment 7000 class Cisco router to deploy together. So it can increase cost.

Inktomi Traffic Server

Inktomi Traffic Server is the solution of Inktomi Inc., which accelerates delivery of web content and streaming media by storing frequently requested content closer to users (Please see Appendix I.5 to get more information). It has some strong points and shortcomings:

- **Strong points:** Inktomi Traffic Server can cache static and all major streaming. It can improve performance and reduce "upstream" bandwidth usage by storing frequently requested content at the edge of the network, physically closer to the end users.
- **Shortcomings:** Since Inktomi Traffic Server system is focus on software, its peak throughput and peak requests/sec are hardware-dependent. It will be easily affected by hardware equipment. In addition, it needs Inktomi representative to install, so the set up is not very easy.

Volera Excelerator

Volera Excelerator is a content networking platform that works in flexible configurations to meet clients content delivery requirements. It accommodates multi-directional demands for content through two configurations: Web browser acceleration and Web site acceleration (Please see Appendix I.6 to get more information). Volera Excelerator has the following strong points and shortcomings:

- **Strong points:** Volera Excelerator can accelerate content transforming between end users and the Internet as well as that between a Web site and its

connection to the Internet. It caches frequently used content and has filtering capabilities. These will give enterprise control over which content should be accessed and stored.

- **Shortcomings:** The operating systems of Volera Excelerator is proprietary, so it will reduce its compatibility. This perhaps limits its wide application.

We compare these Caching Products above more features through a Table 6.1 as follows. These items in the column are important factors to compare and evaluate Caching Products. We got the relative data from Web-Caching site (This site is dedicated to provide a comprehensive guide to the resources and in support of caching on the World Wide Web), Caching Products companies' web site and Products manual/white paper.

Product	Array 1000	SA-7000	Netcache C600	Cache Engine 590	Traffic Server	Excelerator
Company	ClicyArray Network	Cacheflow	Network Applicance	Cisco Systems	Inktomi	Volera
Website	www.clickarray.com	www.cacheflow.com	www.netapp.com	www.cisco.com	www.inktomi.com	www.volera.com
Target Market	Small or midsize networks	Big Enterprises & Mid, Large ISPs	Tier One Service Providers, Broadband Providers	National backbone, regional ISPs, large enterprises	Tier One Providers	Small, mid, carrier ISPs; Small business, WANs
Systems Focus	Hardware & Software	Hardware & Software	File System Software	Hardware	Software	Hardware & Software
Operating Systems	Proprietary	Proprietary	Proprietary, Network Appliance File System	Web Cache Control Protocol -- Proprietary	Solaris, Linux, Win2k, HP-UX, UNIX, IRIX	Proprietary
Ease of Installation	Plug-N-Play	Plug-N-Play	Needs policy routing or layer 4 switch configuration	Significant Router & Cache Engine Reconfiguration	Requires Inktomi Representative	Plug-N-Play
Browser Sys Admin	Yes	Yes	Yes	Yes	Yes	Yes
Content Pre-loading	Yes	Yes	No	No	Yes	Yes
Support HTTP1.1	Yes	Yes	Yes	No	Yes	Yes
Peak Throughput	500Mbps	300Mbps	155Mbps	25-40Mbps	Hardware-dependent	Not stated
Peak Requests/Sec	20,0000	Thousands	Thousands	Not stated	Hardware-dependent	12,300
Streaming Media	Support	Support	Support	No	Support	Support
Appliance	Yes	Yes	Yes	Yes	No	Yes
Scalability	High	High	High	High	High	High
Partnerships	Microsoft, RSA...	Akamai, Foundry,F5...	Alteon...	Mirror Image, Microsoft...	Sun, Intel, 3Com, HP...	Akamai, Digital Island...
Additive Equipment	None	None for policy-based routing; Layer 4 switch	None for policy-based routing; Layer 4 switch	7000 class Cisco router	Layer 4 switch, WCCP router for transparency	Layer 4 switch or WCCP router for transparency
Price Range (US$ 2002)	$25,000 based	$4,495-112,995	$16,550- 65,720	$15,000 base	$24,000 per CPU	$1,000-$99,000

Table 6.1: Compare Different Caching Products

According to Table 6.1, these items mainly reflect product target market, performance/ character, and administration.

Firstly, we find different company's product has different target market. For example: Array 1000 and Excelerator are designed for small or midsize networks, however SA-7000 and Cache Engine 590 are suitable for big or lager enterprise. There is different price range for consumer selection. Table 6.1 also shows the information of different partnerships.

Then, we can know the performance and character form table 6.1. Except Netcache C600 and Cache Engine 590, other cache products apply content pre-load, hence the peak throughput of Netcache C 600 and Cache Engine 590 are relatively lower than the others. Please notice, since Traffic Server system only focuses on software, its peak throughput and peak requests/sec are hardware-dependent; it will be significantly affected by the hardware on which the software is implemented. All of the cache products that we select in table 6.1 have high scalability. Almost all the cache product can support HTTP 1.0 and deliver streaming media, however, Cache Engine 590 is an exception.

Finally, we can get administration information from table 6.1. Almost all the cache products utilize proprietary operating systems, which perhaps limit their compatibility. All cache products in table 6.1 can support browser system administration. Almost all of them are easy to install, however Traffic Server needs Inktomi representative to install, so the set up is not very easy. Some of the cache products do not need additional accessory equipment, which can save cost for consumers; others need additional equipment to work together, such as Cache Engine 590 needs to work with 7000 class Cisco router.

Consumers can select suitable Cache Product depending on their requirement and products characteristics.

6.1.2 Compare Global Server Load Balancer

GSLB is an important component in CDNs to improve network performance and scalability through redirecting the client to the most appropriate caching site. Many choices are available for GSLB solutions in the marketplace. We are not able to cover all of them. The following is a brief analysis of some major vendors and their solutions. They include 3-DNS (F5), Arrowpoint & Distributed Director (Cisco), WSD (Radware), Alteon WebOS (Nortel) and Server Iron (Foundry Network). We compare and evaluate the products' strong points and shortcomings. In addition, we use a table to compare their performance based on four points: Client Testing, Server Testing, Load Balancing and Management. Appendix II lists more features descriptions of the GSLB mentioned above.

3-DNS

3-DNS is the product of F5 Networks, which is an intelligent load balancing solution to geographically distributed Internet sites and data centres. It manages and distributes Internet requests across multiple server sites. Clients' requests are distributed according to data centre and network conditions, such as round trip time or packet loss (Please see Appendix II.1 to get more information). 3-DNS has the following strong points and shortcomings:

- **Strong points:** 3-DNS can offer built-in e-mail or pager alerts when emergency happens. It has excellent security. 3-DNS can apply active scheme to do client and server testing, which is helpful to redirect the client to right site easily.
- **Shortcomings:** 3-DNS has high price because it needs special installation support.

Arrowpoint CSDNS

Arrowpoint CSDNS is the GSLB product of Cisco Systems, which has a series of products, including the CS-50, CS-100, CS-150 (stackable), and CS-800 (chassis) (Please see Appendix II.2 to get more information). The strong points and shortcomings of Arrowpoint CSDNS are as follows:

- **Strong points:** Arrowpoint CSDNS provides customers with a feature-rich, flexible content switching platform. It supports a full set of content-aware features and cookie-based switching as well as integrates with existing Cisco products to meet the needs of Cisco customers.
- **Shortcomings:** Arrowpoint CSDNS can't work with third party products or standalone hosts. It has only three load balancing algorithms, and its topology algorithm is limited to continental region.

Distributed Director

Distributed Director is an other load-balancing solution of Cisco Systems, which redirects end users to the closest responsive sever, which is determined by such factors as client-server proximity and client-server link latency (Please see Appendix II.3 to get more information). Distributed Director has the following strong points and shortcomings:

- **Strong points:** In DNS caching name server mode, DD works for all IP traffic. DD can return an IP address of the best server to the client's local DNS, independent of the particular Internet service requested at that IP address. In HTTP redirect mode, the Cisco DD can provide HTTP session redirection services. DD leverages off the fact that in such situations there is generally a border router which can collect metrics and transmit them back to the DD through DRP (Director Response Protocol).

- **Shortcomings:** DD can only test its clients using a simple TCP null socket request. It has a simple tiered solution with a discrete order of preference. In addition, DD only has six load balancing algorithms.

Web Server Director (WSD)

Radware Web Server Director (WSD) offers layer 4-7 server load balancing and GSLB. It includes several load balancing options, such as triangular routing approach. It also uses active test to redirect clients to right site(Please see Appendix II.4 to get more information). WSD has the following strong points and shortcomings:

- **Strong points:** WSD has more effective traffic control regulation it can guarantee optimal utilization of WEB servers or any intranet server applications. It can use active scheme to help to do flexible addressing and server distribution. It also provides servers with an additional layer of protection as well as support maximum user access with minimum number of IP addresses. Also, WSD can avoid single point of failure effectively.
- **Shortcomings:** WSD has relatively high price, although it allows growth in small increments. It also has the common shortcomings of active scheme (More information, please see 5.2.1.1), This might generate significant traffic and increase delay time for the client to get the result from WSD.

Alteon WebOS GSLB

The main product of Alteon Systems is WebOS GSLB, which provides Layer 2 through 7 Web application, management and traffic control services that seamlessly integrate with Nortel Network's Alteon stackable and modular Web Switch product portfolio (Please see Appendix II.5 to get more information). Alteon WebOS GSLB has the following strong points and shortcomings:

- **Strong points:** Alteon WebOS GSLB has lots of functionality, including various health-check options. It has good control capability for high bandwidth as well as a good capacity for IP Filtering.
- **Shortcomings:** Alteon WebOS GSLB has no client testing. It also can't test 3rd party SLB's and the hosts that aren't directly attached to an Alteon device. In addition, it has clumsy GUI management interface.

Server Iron

Foundry Networks' Server Iron product is a layer 4-7 load balancer with built-in GSLB functionality. The Server Iron can employ DNS-based Footraces to determine site selection, in addition to monitor load metrics (Please see Appendix II.6 to get more information). Server Iron has the following strong points and shortcomings:

- **Strong points:** Server Iron has powerful content switching capabilities, including URL, Cookie and SSL. It also includes high performance VPN/Firewall load balancing. In addition, it can support for all major streaming media protocols.
- **Shortcomings:** Server Iron can't support active client testing.

We compare these GSLB products features by Client Testing, Server Testing, Load Balancing and Management in Table 6.2 (The capability of these four points is showed by "High", "Middle" and "Low". "-" means we don't find the clear data to determine.) For a detailed description, please see "Appendix II GSLB Descriptions from Vendors Websites". These items in the column are important factors to compare and evaluate GSLB. We got the relative data from GSLB companies' web site and Products manual/white paper.

Product	3-DNS	Arrowpoint	DD	WDS	WebOS	ServerIron
Comany	F5	Cisco		Radware	Nortel	Foundry
Client Testing	High	Middle	Middle	High	-	Middle
Active	High	Middle	Middle	High	-	-
TCP Probing	High	Middle	Middle	-	-	-
UDP Probing	High	-	Low	High	-	-
Passive	-	-	Middle	-	-	Middle
Server Testing	High	Middle	Middle	Middle	Middle	Middle
Native SLB	High	Middle	-	Middle	Middle	High
Connections	High	-	-	High	-	High
Availability	High	High	-	High	High	High
Latency	High	-	-	High	High	Middle
Customization	High	-	-	-	-	Middle
SNMP	High	-	-	-	-	-
3rd Party SLB	High	-	Middle	Middle	-	-
Connections	High	-	-	-	-	-
Availability	High	-	Middle	Middle	-	-
Latency	-	-	Middle	-	-	-
SNMP	High	-	-	-	-	-
No SLB	Middle	-	Middle	Middle	-	-
Connections	Middle	-	-	-	-	-
Availability	Middle	-	Middle	Middle	-	-
Latency	Middle	-	Middle	-	-	-
SNMP	Middle	-	-	-	-	-
Load Balancing	High	Low	Middle	Middle	Low	Middle
Tiered System	High	Low	Middle	Low	Low	High
Static Algorithms	High	Middle	Low	-	Low	Low
Round Robin	High	High	Low	-	-	-
Ratio	High	-	Low	-	-	-
Random	High	-	High	-	-	-
Dynamic Algorithms	High	Middle	Middle	Middle	Low	Middle
Round Trip Time	High	High	High	High	Middle	Middle
Completion Rate	High	-	-	-	-	-
Hop Count	High	-	High	High	-	-
Least Connections	High	High	-	High	-	High
Packet Rate	High	-	-	-	-	-
Others	High	Low	Low	Middle	Low	Middle
SLB Capacity	Middle	-	-	High	-	High
Topology	High	-	-	Low	Middle	Middle
Quality of Service	High	-	-	Middle	-	-
Maximum Connections	Middle	-	-	High	High	High
Global Availability	High	High	High	-	-	High
Production Rules	High	-	-	-	-	-
Management	High	Low	Middle	Middle	Low	Middle
Interdevice Communication	High	Middle	Middle	Middle	Middle	Middle
IP Filtering	High	High	High	High	High	High
Hardware Redundancy	High	Middle	-	Middle	Middle	Middle
Statistics Collection	High	Middle	Middle	Middle	Middle	Middle
GUI	High	Low	-	High	Middle	Middle

Table 6.2: Compare Different GSLB

GSLB utilizes Client Testing and Server Testing to measure and reflect distance between client and cache site. Load Balancing shows the ability of GSLB to measure site load situation. Management reflects the GSLB administration ability.

In Table 6.2, we first can find 3-DNS and WDS have higher ability of client testing than others. For the server testing, 3-DNS has higher ability. It not only does tests on native SLB, but also can work for 3rd part SLB. Almost all GSLB in Table 6.3 can use static and dynamic algorithms to measure site load balancing. WebOS has relatively limited ability; other products have more ability. Management ability reflects whether it is easy to operate and administrate, which consists of statistics collection and GUI, etc. For example: Since Arrowpoint and WebOS don't have good GUI, their management aren't easy.

Consumers can select suitable GSLB Product depending on their requirement and products characteristics.

6.1.3 Compare SSL Accelerator

Ensuring the security of network distribution is other important part of CDNs. Here, we use SSL Accelerator as examples to discuss and analyze it. SSL Accelerator works as a point network device that resides in front of the Web server, typically used in conjunction with server load balancers. The SSL Accelerators accept connections from the client and use the special cryptographic accelerating hardware within them to speed up the process as well as transmitting data securely over TCP/IP networks. In this section, we select some main current SSL Accelerator as examples to compare and evaluate their features. They are Array 1000 (ClickArray Network), iSD-SSL 2.0 (Nortel Networks), SA-700 (CacheFlow), NetStructure 7115 e-Commerece Accelerator (Intel), e-Commerce Controller 540 (F5) and SSL-R (SonicWall). Appendix III lists the more features descriptions of these SSL Accelerator mentioned above.

Array 1000

Array 1000 is the product belonging to ClickArray Networks Inc. It is the Integrated N+1 Clustering, which consists Web security, server load balancing, global server load balancing, reverse proxy cache, SSL acceleration and CDN content rewriter.
(More information of Array 1000, please see 6.1.1 Array 1000).

iSD-SSL 2.0

iSD-SSL 2.0 is the product of Nortel Networks (Formally Alteon Web Systems), which is a fully-featured Secure Sockets Layer (SSL) appliance integrating SSL acceleration, SSL extranet management, and secure application services into a single device (Please see Appendix III.2 to get more information). iSD-SSL 2.0 Acceleration has the following strong points and shortcomings:

- **Strong points:** iSD-SSL 2.0 is an integrated application services including load balancing, session persistence, and Layer 7 filtering to optimize secure application performance. It improves application performance by relieving servers of complex public key operations and bulk encryption. In addition, iSD-SSL 2.0 can reduce cost and simplify operations by digital certificate and key management consolidation.
- **Shortcomings:** iSD-SSL 2.0 can't support SSL Initiation and SSL Aggregation. In addition, it also doesn't support HTTP Header Insertion, URL Logging and Secure URL Re-write. Since it is an integration, so this perhaps leads to a single point of failure.

SA-700

SA-700 is a product from Cacheflow, which is integrated 1+1 clustering, consists reverse proxy. The SA-700 hardware is optimized for Web server acceleration, featuring a high RAM-to-disk ratio and a built-in Secure Sockets Layer (SSL) encryption/decryption

processor (Please see Appendix III.3 to get more information). SA-700 has the following strong points and shortcomings:

- **Strong points:** SA-700 can increase response time for Web site users as well as lower IT space. It is an Integrated SSL Cryptographic Processor and Secure Content Acceleration, which can accelerate both public (HTTP) and private (HTTPS) content through integrated SSL functionality.
- **Shortcomings:** SA-700 can't support Transparent Operation, SSL Initiation and SSL Aggregation. In addition, it also doesn't support HTTP Header Insertion, URL Logging and Secure URL Re-write.

NetStructure 7115 e-Commerece Accelerator

NetStructure 7115 e-Commerece Accelerator is a product of Intel. It boosts SSL transaction performance with patent-pending technology that offloads cryptographic functions from the server (Please see Appendix III.4 to get more information). There are strong points and shortcomings of NetStructure 7115 e-Commerece Accelerator as follows:

- **Strong points:** NetStructure 7115 e-Commerece Accelerator can increase connection rates and decrease response times in e-Business data centres with its interoperable drop-in installation design. It also supports a wide variety of cryptographic algorithms, including Blowfish and IDEA. The device reduces the average connection time for midsize Web objects.
- **Shortcomings:** NetStructure 7115 e-Commerece Accelerator can't support SSL Initiation, SSL Aggregation and Secure URL Re-write. In addition, it also doesn't support HTTP Header Insertion, URL Logging and HTTP/HTTPS Web GUI. The price of NetStructure 7115 e-Commerece Accelerator is also a little higher.

e-Commerce Controller 540

e-Commerce Controller 540 is a product of F5, which manages Secure Socket Layer (SSL) encryption/decryption to increase site performance and ability to scale SSL transactions (Please see Appendix III.5 to get more information). e-Commerce Controller 540 has the following strong points and shortcomings:

- **Strong points:** e-Commerce Controller 540 can increase server performance, and reduce administrative costs by offloading the burden of SSL traffic from web servers. It supports centralized certificate management to a single source to greatly simplify administrative duties. In addition, e-Commerce Controller 540 has interoperability with BIG-IP load balancer or any 3rd party load balancing products as well as it can be deployed quickly within any network.
- **Shortcomings:** e-Commerce Controller 540 can't support SSL Initiation, SSL Aggregation and Secure URL Re-write. In addition, it also doesn't support HTTP Header Insertion, URL Logging, Key/CSR/Cert Generation and Friendly SSL Failure Messages.

SSL-R

SSL-R is produced by SonicWall, which integrates seamlessly with Layer 4 load balancers and Layer 5-7 content switches to enable secure and intelligent content networking to improve performance and content location flexibility (Please see Appendix III.6 to get more information). SSL-R has the following strong points and shortcomings:

- **Strong points:** SSL-R can increases performance and reliability for Web sites and commercial applications. It also eliminates the need for costly multiple-server deployment and maintenance as well as content switch friendly. In addition, SSL-R can support SSL Initiation, SSL Aggregation, HTTP Header Insertion, URL Logging, and Secure URL Re-write. It also has big MTBF, more than 350000 hours.

- **Shortcomings:** SSL-R has embedded Hardware Platform, it perhaps limits its compatibility.

We compare the features of the above SSL Accelerator products in detail in Table 6.3. These items in the column are important factors to compare and evaluate SSL Accelerator. The data in table 6.3, we got from SSi Service Strategies Inc. site (It is professional site to compare and evaluate SSL Accelerator performance and character), SSL Accelerator companies' web site and Products manual/white paper.

Because there are lots of acronyms and special terms in this table, it is necessary to definite and explain them. These acronyms and special terms are as follows:

- **RSA:** " Rivest-Shamir-Adleman: One of the fundamental encryption algorithms or series of mathematical actions developed in 1977 by Ron Rivest, Adi Shamir, and Leonard Adleman. The RSA algorithm is the most commonly used encryption and authentication algorithm and is included as part of the Web browsers from Netscape and Microsoft." [83]. It is an algorithm for asymmetric cryptography. [51]

- **RSA Operations:** "The Key Exchange and digital signing that occurs at the start of all new SSL sessions. The RSA operation is a portion of the SSL Handshake, the extremely processor intensive communications used to establish a secure sockets connection." [83]

- **RSA Operations Peak:** How much of the peak RSA operation.

- **Concurrent Flows:** It is used to measure the memory management and overall networking efficiency of the underlying operating system, and of the SSL proxy code itself. [50]

- **Sustained Session Ids:** It is the speed at which an SSL Accelerator can encrypt and decrypt or "move" data.[50]

- **Chained-Certificates:** Whether support Chained Certificates("Chained certificates are used in several circumstances such as when a known, accepted certificate authority (CA) provides a certificate to attest that certificates created by a non-recognized party can be trusted. For example, a company may create its own certificates for internal use only; however, clients will not accept the certificates because a known CA has not created them. By chaining the trusted CA's certificate with private certificates; clients accept the internal certificates during SSL negotiations" [52]).

- **Transparent Operation:** Whether support Internal certificate generation for intranet testing and configuration. It is for system administration.

- **SSL Initiation:** "SSL Initiation is the ability to behave not only as an SSL termination point (an SSL server) but also as the initiator of an SSL connection (an SSL client). Much like the standard web-browser is the SSL client when connecting to a secure site." [85]

- **SSL Aggregation:** "SSL Aggregation is an extension of SSL Initiation capabilities. When initiating outbound SSL traffic." [84]

- **Key/CSR/Cert Generation:** Whether support Key/CSR/Cert Generation.

  - **Key:** "In cryptography, a key is a variable value that is applied using an algorithm to a string or block of unencrypted text to produce encrypted text. The length of the key generally determines how difficult it will be to decrypt the text in a given message." [83]

- **CSR:** "Certificate Signing Request (CSR)A CSR is a file containing your certificate application information, including your Public Key. Generate your CSR and then copy and paste the CSR file into the web form in the enrolment process." [86]
- **Cert:** SSL Certificates designed for securing leading web sites, as well as intranets and extranets

- **Client Certificate Support:** Whether support client-side certificates.

- **HTTP Header Insertion**: Whether support Hypertext Transfer Protocol (HTTP) header insertion, which can extract client certificate information for presentation to back-end servers by using HTTP headers.

- **URL Logging:** The field specifies the URL of the login page.

- **Asynchronous RTOS:** Whether support Asynchronous Real-Time Operating System (RTOS)

- **Ephemeral RSA Support:** Whether support ephemeral RSA key. "When using a cipher with RSA authentication, an ephemeral RSA key exchange can take place. In this case the session data are negotiated using the ephemeral/temporary RSA key and the RSA key supplied and certified by the certificate chain is only used for signing. Use ephemeral RSA key can save a lot of computer time." [53]

- **Secure URL Re-write:** Whether support Secure URL Re-write. It can improve efficiency by removing two hops, including a redundant redirect, Secure URL Re-Write guarantees that no sensitive data is ever exposed. [54]

- **Encrypted CLI:** Whether support Encrypted CLI, which can set-up fast and easily via console port command line interface (CLI). It is used to show the ability to Configuration and Certificate Key Management. [55]

- **MTBF:** "It is acronym for Mean Time Between Failures. The average time interval, usually expressed in thousands or tens of thousands or tens of thousands of hours (sometimes called power-on hours or POH), that will elapse before a hardware component fails and requires service."[56]

roduct	Array 1000	iSD-SSL 2.0	SA-710	Intel 7115	e-Com 540	SSL-R
Company	Array	Nortel	Cacheflow	Intel	F5	SonicWALL
RSA Operations Peak	800	400	200	600	800	200
Concurrent Flows	--	16000	6000	6000	16000	5000
Sustained Session IDs	--	8000+	-	-	16000	75000
Hardware RSA	Yes	Yes	Yes	Yes	Yes	Yes
Chained-Certificates	Yes	Yes	Yes	Yes	Yes	Yes
Transparent Operation	No	Yes	No	Yes	No	Yes
SSL Initiation	No	No	No	No	No	Yes
SSL Aggregation	No	No	No	No	No	Yes
Ephemeral RSA Support	Yes	Yes	Yes	Yes	Yes	Yes
Key/CSR/Cert Generation	No	Yes	No	No	No	Yes
Client Certificate Support	No	Yes	No	Yes	Yes	Yes
HTTP Header Insertion	No	No	No	No	No	Yes
URL Logging	No	No	No	No	No	Yes
Secure URL Re-write	No	No	No	No	No	Yes
Friendly SSL Failure Messages	No	Yes	No	No	No	Yes
Microsoft OWA 2000 Support	No	No	No	No	Yes	Yes
Encrypted CLI	No	No	No	No	No	Yes
HTTP/HTTPS Web GUI	Yes	Yes	Yes	No	Yes	Yes
SNMP	Yes	Yes	Yes	Yes	Yes	Yes
Asynchronous RTOS	Unix-Like	Unix-Like	Unix-Like	Unix-Like	Unix-Like	Yes
Hardware Platform	PC	PC	PC	PC	PC	Embedded
MTBF	-	>50000 hrs	-	-	>50000 hrs	>350000 hr
Price ($US)	$24,995	$10,995	$9,995	$15,995	$17,990	$6,995

Table 6.3: Compare Different SSL Accelerator

According to Table 6.3, these items mainly reflect SSL Accelerator product performance/characteristics and management. Please see " 6.1.3 Compare SSL Accelerator" to get detail explanation of these items.

Firstly, we find different company's product has different performance and characteristics. Higher RSA operation peak and Concurrent Flows reflect that SSL Accelerator has stronger ability to ensure security and operation. For example: Array 1000 and e-Com 540 have higher RSA operation peak, while iSD-SSL 2.0 and SSL-R have relatively lower; iSD-SSL 2.0 and e-Com 540 have higher Concurrent Flows, which reflect they have effective operation performance. Sustained Session Ids show the speed at which an SSL Accelerator can encrypt and decrypt or "move" data; e-Com 540 has relative stronger ability than others. All the SSL Accelerator in table 6.3 can support hardware RSA, Chained-Certificates, ephemeral RSA and SNMP, which can satisfy consumers' different requirement.

ISD-SSL 2.0, Intel 7115 and SSL-R can support Transparent Operation, which means they can support Internal certificate generation for intranet testing and configuration, so they have good system administration ability. Except Intel 7115, others SSL Accelerator can support HTTP/HTTPS Web GUI, which mean it is easy to administrate.

Please notice: only SSL-R has SSL Initiation, SSL Aggregation, HTTP Header Insertion, URL Logging, Security URL Re-write and Encrypted CLI, which show SSL-R has stronger function. At the same time it has relative high MTBF and relative low price.

Different product has different target market. Consumers can select suitable product according to their requirement and products characteristics.

In this section, we selected some representative current CDNs companies to compare and evaluate their products features respectively. These products included Caching Product, Global Server Load Balancer and SSL Accelerator that are necessary for CDNs design. Next section, we will compare and analyze the service solutions of some representative CDNs providers.

6.2 Compare Service Solutions of Different CDNs Providers

In last section, we compared and evaluated some companies' CDN products' features. In this section, we will select several CDNs companies to compare their CDNs service solutions. These companies include Adero, Akamai, CacheWare, Cidera, ClearWay, Digital Island, epicRealm, iBeam, Mirror Image Internet and Pushcache. We chose these companies because they are in the top positions in the CDNs market according to the statistic of Information week. Since the CDNs market is rapidly changing, new CDNs providers emerge while others disappear or are acquired by other companies. There are a sketch map concerning CDNs services market of different CDNs providers in 2001.

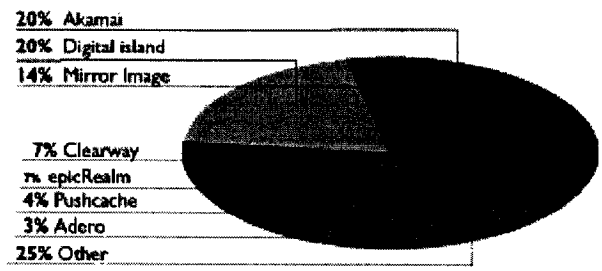

Figure 6.1: The Services Market of Different CDNs Providers [66]

We will describe and analyze these CDNs services solution of different providers separately as follows. Appendix IV lists the more solutions for service descriptions of these CDNs Providers mentioned above.

Adero

Adero(http://www.adero.com) was found in 1998, which was a venture-funded, global Internet services company and its headquarter is in Boston, MA, USA. Lauded by Computer world magazine as one of the 100 Emerging Companies to Watch in 2000, Adero has over

230 employees worldwide and its major strategic investors include: Intel, America Online, Inktomi, Reuters and Microsoft. Adero enables e-business with global content distribution services. It has servers in more than 30 countries and continues to expand that network.[108] Adero provides content delivery solutions to carriers and hosting providers through the established Adero GlobalWise SM Network and content delivery services (Please see Appendix IV.1 to get more information of GlobalWise). Adero applies DNS-based methods to redirect the clients to right site (More information includes strong points and shortcomings, please see 5.2.2.2 and 5.3.1).

Akamai

Akamai (http://www.akamai.com) began at the Massachusetts Institute of Technology (MIT) in early 1995. Now it is one of the most famous CDN providers. Akamai has edge servers deployed in Sri Lanka, Guatemala, and Qatar, in addition to 60 other countries on 6 continents and has 13000 servers. Its Technology Partner includes BEA Systems Inc., IBM Corporation, Microsoft Corp. and F5 Networks Inc., etc big IT companies. In April 2000, Akamai purchased InterVu. Akamai provides content delivery and streaming media services, along with global traffic management.[109] EdgeSuite is a main CDNs service solution of Akamai, which is an integrated suite of services for content and application delivery, content targeting, edge processing, business intelligence, and streaming media (Please see Appendix IV.2 to get more information of EdgeSuite). Akamai combines DNS-based method with URL rewriting ones to redirect the clients to the right site (More information includes strong points and shortcomings, please see 5.2.2.2, 5.2.2.3 and 5.3.1).

CacheWare

CacheWare (http://www.cacheware.com) is a rapidly growing, privately held software technology company focusing on delivering enterprise content control and acceleration solutions based on caching technology. The privately funded company started operations in

September 1999 and lists three primary investors - Adler & Co., Rogers Investment, and J.F. Shea & Co.- and one technology partner, Sun Microsystems. CacheWare's headquarter is in Silicon Valley. Now CacheWare has been acquired by Fort Hill Systems (http://www.forthillsystems.com).[110] EdgeSystem includes EdgeServer and EdgeManager, is a main CDN service solution of CacheWare, which is a low overhead server appliance that can reside anywhere in the corporate intranet or extranet and takes the load off an origin server by acting as the intermediary between origin and edge servers (Please see Appendix IV.3 to get more information of EdgeSystem). The redirection methods of CacheWare includes DNS-base and GSLB technology (More information includes strong points and shortcomings, please see 5.2.2.2, 5.3.1 and 5.3.2).

Cidera

Cidera (http://www.cidera.com) was found in 1997, now serves hundreds of POPs (points of presence) in North America. Its corporate headquarter is located in Laurel, Maryland. By its satellite-based distribution network for the Internet, Cidera offers a complementary means of transporting broadband content for content providers, aggregators, and distributors. Cidera's network is satellite-based (More information including strong points and shortcomings, please see 5.4.1.2) and specializes in transporting data streams. It has more than 300 points of presence in North America and presence in Europe, with expansion into Latin America and Asia later this year.[111] Cidera Streaming Media Service is a main CDNs service solution of Cidera, which allows content providers, aggregators, and distributors to seamlessly and simultaneously deliver live streaming video and audio content into Cidera's servers located at numerous access points by Cidera's satellite broadcast network (Please see Appendix IV.4 to get more information of Cidera Streaming Media Service). Cidera applies GSLB methods to redirect the clients to the right site (More information including strong points and shortcomings, please see 5.3.2).

Clearway

Clearway (http://www.clearway.com) is a provider of server-based content delivery solutions that provides Web performance services to e-businesses of all sizes. Clearway has been acquired by Xcelera/Mirror Image in January 2001.[112] FireSite is the main CDNs service solution of Clearway, which is completely transparent to the origin server and existing infrastructure, with which the hosting may use for high availability or load balancing (Please see Appendix IV.5 to get more information of FireSite). Clearway applies URL rewriting methods to redirect the clients to the right site (More information includes strong points and shortcomings, please see 5.2.2.3).

Digital Island

Digital Island (www.digitalisland.com) offered content delivery in 1996, which is headquartered in San Francisco. It has raised $800 million by private placements and an initial public offering. Three customers: Compaq, Intel, and Microsoft also kicked in $45 million to help Digital Island expand its network to handle 7.5 million users. Digital Island provides global application hosting and content distribution over a private network that bypasses oversubscribed public networks. Digital Island has Web-hosting facilities in New York, Santa Clara, Honolulu, Hong Kong, Tokyo, and London that provide network access to 27 countries. Streaming content delivery is also provided. Digital Island purchased Sandpiper Networks and is likely the second most popular CDNs provider. In 2001 Cable & Wireless purchased Digital Island and markets the services under the Exodus brand.[113] GlobePort is a main CDNs service solution of Digital Island, which is a high-speed connection service that extends e-business delivery to self-hosted customer servers, via two leased lines, frame relays, or other appropriate fast pipes (Please see Appendix IV.6 to get more information of GlobePort). Digital Island combine DNS-based method with URL rewriting ones to redirect the clients to the right site (More information including strong points and shortcomings, please see 5.2.2.2, 5.2.2.3 and 5.3.1).

epicRealm

epicRealm (http://www.epicrealm.com) was found in 1995, which specializes in a worldwide E-commerce network for the business-to-business market. Its network backbone covers North America, Europe, and Asia, and lets customers be served by local servers. epicRealm caches static and dynamic content, database-driven content, and even encrypted content. epicRealm provides content delivery solutions to carriers and hosting providers through eXT Technology, which is the ingeniously simple collaboration between an epicRealm infrastructure acceleration appliance and web application (Please see Appendix IV.7 to get more information of eXT Technology).[114] EpicRealm apply redirection software and protocol to redirect the clients to right site (More information includes strong points and shortcomings, please see 5.2.2.4).

iBeam

iBeam (www.ibeam.com) is a broadcasting company whose Intelligent Distribution Network and infrastructure delivers high-quality streams to audiences. Its strategic partners include Microsoft, Network Appliance, and RealNetworks. iBeam specializes in streams via satellite (More information including strong points and shortcomings, please see 5.4.1.2) rather than terrestrial lines. This method reduces the number of hops to transmit the stream, thereby reducing packet loss. In December, 2001, iBeam was acquired by Williams Communications Group.[115] Its CDNs service realizes by iBeam Intelligent Distribution Network (Please see Appendix IV.8 to get more information of iBeam Intelligent Distribution Network). iBeam applies GSLB method to redirect the clients to the right site (More information including strong points and shortcomings, please see 5.3.2).

Mirror Image Internet

Mirror Image Internet (http://www.mirror-image.com), a provider of Internet content delivery solutions, improves Web performance by delivering content regardless of location or demand. It is a subsidiary of Xcelera.com Inc., and boasts a list of partners, including Cisco Systems, Compaq, Hewlett-Packard, and Oracle [116]. instaDelivery Internet Service is the main CDNs service solution of Mirror Image Internet, which stores Web objects and manages content delivery to provide a completely outsourced object server solution (Please see Appendix IV.9 to get more information of instaDelivery Internet Services). Mirror Image Internet apply DNS-based combined with URL rewriting methods to redirect the clients to the right site (More information including strong points and shortcomings, please see 5.2.2.2, 5.2.2.3 and 5.3.1).

Pushcache

Pushcache.com (http://www.pushcache.com) was found in 1998, which is an Internet software company focused on the development and sale of software products based on the pushcache and "push done right" (Please see Appendix IV.10 to get more information of Pushcache Push Software).[117] Pushcache.com apply redirection software and protocol to redirect the clients to right site (More information includes strong points and shortcomings, please see 5.2.2.4).

We compare detail features of these CDNs providers and service solutions above in a Table 6.4 as follows. We got the relative data from Telezoo website (Telezoo is a professional company to provide performance measure, service research, market analysis to buyers), CDNs service providers' web site and Solution manual/white paper.

Company	Adero	Akamai	CacheWare	Cidera	Clearway	Digital Island	epicRealm	iBeam	Mirror Image	Pushcache
Service	GlobalWise	EdgeSuite	EdgeSystem	SM Service	FireSite	GlobePort	eXT Tech	iBeam CDN	instaDelivery	Push Software
CDN Type:	Traditional	Traditional	Software	Satellite-Based	Traditional	Traditional	Software	Satellite-Based	Traditional	Software
Supported Content Types	Streaming	Images, Streaming Digitized Downloadable Files, HTML	HTML, Static Content, Streaming	Flash, HTML, Static Content, Streaming	HTML, Static Content, Streaming	Authentication, Encrypted Content, Flash, HTML, Static, Streaming	Static, Dynamic Database-Driven content, Encrypted	Chat, Encrypted Content, Flash, HTML, Static Content, Streaming	HTML, Static Content, Streaming	HTML, Static Content, Streaming
Supported Streaming Format	Microsoft Windows Media, Real	MWM QuickTime, Real	MWM QuickTime, Real	Macromedia Flash, MWM QuickTime, Real	MWM QuickTime, Real	Macromedia Flash, MWM QuickTime, Real	MWM QuickTime, Real	Media Flash, MWM MPEG, QuickTime, Real	MWM QuickTime, Real	Microsoft Windows Media, Real
ISP/Provider relationship	Maintained by Customer	Maintained by Customer	Maintained by Vendor	Maintained by Vendor	Maintained by Vendor	Maintained by Vendor	Maintained by Vendor	Maintained by Vendor	Maintained by Vendor	Maintained by Vendor
Web & Media Servers	Provided by Customer	Provided by Customer	Provided by Vendor	Provided by Customer	Provided by Vendor	Provided by Vendor	Provided by Vendor	Provided by Vendor	Provided by Vendor	Provided by Customer
Redirection	DNS-based	DNS-based/ URL Writing	DNS-base and GSLB	GSLB	URL Writing	DNS-based/ URL Writing	software and protocol	GSLB	DNS-based/ URL Writing	software and protocol
Management/Reporting Features	Logging…	Billing, Site Usage, Traffic Analysis ..	Billing, Logging,	Billing	Billing, Traffic Analysis..	Billing, Site Usage, User Demographics..	Billing, Logging,	Billing, Traffic Analysis, User Demographics	Billing, Site Usage, Traffic Analysis…	Billing, Logging, Traffic Analysis…
Maintenance Support	24x7 Network Monitoring	24x7 Network Monitoring	24x7 Network Monitoring	24x7 Network Monitoring	24x7 Network Monitoring	24x7 Network Monitoring, Field Service	24x7 Network Monitoring	24x7 Network Monitoring, Field Service	24x7 Network Monitoring	24x7 Network Monitoring
Partners	AOL, Inktomi..	F5, IBM…	Sun..	Novell, Cisco..	Mirror Image.	Cisco, Inktomi.	Sun…	Microsoft…	Compaq, HP.	Bea Systems
Coverage Area	More than 30 countries	More than 60 countries	North America	Africa, Europe, Middle East, North America	North America	Africa, Asia-Pacific, Europe, Middle East, America..	North America, Europe, and Asia	Asia-Pacific, Europe, Latin America, North America	North America	North America

Table 6.4: Compare Different CDNs Providers and Services

According to Table 6.4, Firstly, we find different provider has different CDNs type. For example: Cidera and iBeam provide CDNs service based on Satellite, Puchcache only utilizes software to provide a caching service.

Then, we can know different solution can utilize different sever selecting and client redirection methods or a combination method. Almost all CDNs solutions can store and deliver the content in any form, such as pictures, digitized audio & video, and text documents. Some of CDNs solutions are maintained by vendors, however others need customers to perform maintenance themselves.

Finally, we can get information of partners of different CDNs providers. Table 6.4 also reflects different CDNs service provider's coverage area. Some companies are global companies, which cover lots of main countries in the world, such as Akamai, which covers more than 60 countries on 6 continents and has 13000 servers in the world. Some providers belong to local area, which mainly cover one or two regions, such as Clearway mainly provide service on North America.

In this Chapter, we selected some representative current CDNs companies to compare and evaluate their products and services. This Chapter was divided into two sections. In the first section, we selected some main current CDNs companies as examples to compare their products, which include Caching Product, Global Server Load Balancer and SSL Accelerator. Firstly, Caching Products included Array 1000 (ClickArry networks), SA-7000 (Cacheflow), Netcache C6100 (Network Appliance), Cache Engine 590 (Cisco), Inktomi Traffic Server (Inktomi), and Volera Excelerator (Volera). Secondly, GSLB included 3-DNS (F5), Arrowpoint & Distributed Director (Cisco), Web Server Director (Radware), Alteon WebOS (Nortel) and Server Iron (Foundry Network). Finally SSL Accelerators included Array 1000 (ClickArray Network), iSD-SSL 2.0 (Nortel Networks), SA-700 (CacheFlow), NetStructure 7115 e-Commerece Accelerator (Intel), e-Commerce Controller 540 (F5) and SSL-R

(SonicWall). In the second section, we selected top 10 CDNs services providers to compare their CDNs solutions. These companies included Adero, Akamai, CacheWare, Cidera, ClearWay, Digital Island, epicRealm, iBeam, Mirror Image Internet and Pushcache. According to requirement, we also used illustration and table to explain relative content.

Chapter 7

Application

In this Chapter, we will discuss the objects of CDNs application and generic application of CDNs. Finally, we will select Cisco ECDN and Volera's Velocity CDN for E-learning as the example to do case study.

7.1 Objects of CDNs Application

The objects of CDNs application means who need CDNs. The final beneficiary undoubtedly is final users. However, what we analyze here is the direct objects or customers of CDNs application. There are mainly two big objects of CDNs application: Enterprise and Services Providers[5]. Discussion and analysis are as follows:

7.1.1 Enterprise

Big Enterprise or large corporation such as IBM, Intel or General Electric, always have several branches located at multiple campuses. The typical enterprise application is hosted at one of the campuses and accessed by users all across the enterprise [1]. If a user's campus has good connectivity to the hosting campus, performance is typically good. However, if the connectivity of the user to the hosting campus is through a congested backbone link, the perceived performance can be

quite erratic. As an example, there is a company that hosts a directory of all employees at a site in Canada. Most of campuses within Canada might have good connectivity with the hosting campus, and performance of application is quite acceptable. However, users at sites in Europe and Asia may not be quite happy at the performance of the system.

CDNs application within the enterprise can solve this problem very well and bring tremendous benefit. By distributing many caching servers at each of different campuses, users are likely to have a relatively good connection to one of the caching sites. Applications that are used across multiple campuses can benefit from the performance improvement available due to the presence of these caching servers. Consider the example of the directory application mentioned above, caching server may be placed in Europe and Asia to improve the response time experienced by users at those locations. The caching servers in Europe or Asia will separately handle all requests from all the campuses in Europe or Asia instead of Canada.

In some cases, enterprises get the benefits of CDNs application from purchasing CDNs services from providers. In other cases, certain enterprise will implement CDNs technologies by themselves. These enterprises purchase load balancing for servers, gateways, firewalls, SSL accelerator, VPNs, and caches to install, test, operate, control and maintain CDNs by themselves [1].

More and more companies will be interested in Enterprise CDNs since more and more content is required, and multimedia becomes even more commonplace in the corporate setting.

7.1.2 Service Providers

With the development of Internet market, it becomes more and more competitive among the service providers. Service providers need to continue offering more valuable services to their customers to get more profitability. Basic services, such as data centre collocation and Internet access, are commodity services with low (or even negative) margins. CDNs application enable service providers to offer valuable new capabilities and functions to the services they sell today [5]. There are various service providers currently, Internet Data Centre (IDC) & Internet Content Providers (ICP), Application Service Providers (ASP), Bandwidth & Managed Security Service Providers, and Internet Service Providers (ISP). We will discuss and analyze them separately as follows.

Internet Data Centre (IDC) and Internet Content Providers (ICP)

The Internet Data Centre (IDC) is the heart of Content Networking, which provide required data within network. Internet Content Providers (ICP) are the companies to provide various content through Internet. The services of IDC or ICP can be very different. It could be very powerful and impressive or be rarely visited. Effective IDC or ICP can deliver rich, compelling content quickly and completely every time. Internet data or content range from basic text-and-graphics to multimedia showcases with complex graphics, valuable downloads and interactive elements. Yet no matter how impressive of the data or content, ultimately it is only as good as the process that brings it to the end users. Applying CDNs can furthest avoid congestion, reduce latency and improve network performance. CDNs application will reduce infrastructure costs, increase profits. All of them will significantly Improve IDC and ICP business competitiveness.

Application Service Providers (ASP)

Application Service Providers can offer services such as load balancing, SSL acceleration, reverse proxy caching, high-availability firewall and VPNs, and content

delivery services to their customers. These value-added services not only have the potential to bring in more revenue, but also create a much stronger tie with the customer. By deploying CDNs, Application Service Providers can provide more availability, security and performance service on the Internet.. The more complex and values the service is, the more locked in the customers are [5].

Bandwidth and Managed Security Service Providers

With the Bandwidth Service access networks, the demand for high quality and scalability of network delivery content is growing. However, there are still bottlenecks and slowdowns in accessing web content such as streaming media for end users, which can not be solved only by bandwidth alone. Applying CDNs in the bandwidth will significantly improve network performance and availability. At the same time, offering highly available and manageable security services is also essential when developing a network distribution. Firewall, VPN, and IDS load balancing are far superior to build in high availability solutions, especially for a provider focused on giving the best in security services [5].

Internet Service Providers (ISP)

Many ISPs provide more than simple network level connectivity to their customers. For many common types of applications, an ISP may provide caches. The application cache can help in reducing the application response time, as well as the bandwidth that needs to be carried across an ISP network. Some ISPs use CDNs to provide other functions, e.g., protecting against spammed mail or filtering out objectionable content [5].

For example, many ISPs provide proxies for web servers. These proxies can implement functions such as caching of static images and static content, transforming the pages, or filtering out objectionable content. Such a proxy may be implemented in

a transparent manner such that a client is unaware of its presence. Each proxy can be used as a caching server. An ISP may combine the collection of proxy servers it supports at its POPs in order to deliver a CDNs solution to its customers. Of course, ISPs can select implementation and manage CDNs by themselves or by third-party CDNs suppliers - Content Distribution Service Providers (CDSP).

The advantage of ISPs by using CDNs is obvious. It is much easier, faster, and more secure for users to access the content they desire. That will bring more revenue and profit margins to the ISPs provider.

7.2 Generic Application of CDNs

In this section, we will look at several generic application of CDNs that are suitable for the CDNs environment and discuss how they can take benefits of the CDNs.

Static Web Pages

The first application of CDNs we examine is the delivery of static web pages over the Internet or a corporate intranet. Any object accessible via a web browser has an URL associated with it. A static web page is an object with HTML text and links to other objects. These linked objects are referred to as embedded objects, and are typically graphics and images that are used to create a visually appealing impression for the users. A large component of the user-perceived delay in accessing a page over a network is due to the time it takes to load the embedded objects. The application of CDNs can significantly solve this problem and dramatically improve the performance since these objects can be retrieved from the caching sites in a CDNs, rather than

from original site itself. The appropriate caching site is different for different clients, depending on the location of the client in the network [1].

Streaming Multimedia Delivery

Streaming multimedia delivery is the second application of CDNs. Multimedia content (audio, video, and animations) on websites can be accessed in one of two ways: either by downloading the entire content to a client and then playing it locally, or by using a streaming server. The streaming sever enables a client to start viewing the content much faster, and is the preferred method for long streams or for watching real-time events in the network. If the caching sites implement the function of a multimedia streaming server, such a service can be provided at the caching site instead of requiring a client to access the original site. Network performance can play a critical role in the quality of the stream seen by the client [1]. With application of CDNs, since the caching sites are much closer to the clients, accessing streams from the cached copy is likely to provide much better performance.

Advertisement Generation

Among the components that are often seen on web pages are the banner advertisements that are usually at the header or the footer of a page. These web advertisements are typically generated by a script running at the origin web or by a third-party advertisement company. Appling CDNs can redirect the target clients to access the page on which the advertisement placed. If the CDNs routing mechanism maintains some rough degree of geographical proximity, it is likely that most clients accessing a caching site are from the same geographical region. Thus, a caching site can generate advertisement that are targeted for a local audience. The caching site refreshes periodically from the original site [1]. For example, a caching server located in Toronto may generate advertisements for the local NBA basketball team - Raptors whenever the sports page of newspaper is accessed, whereas the caching site

located in San Francisco, California, US, would generate advertisements for a different team.

Content Adaptation

The individual preference of clients may be significantly different. For an example, a client in China may want to see a page in Chinese, whereas a client in the United States would prefer to see all pages in English. If a server knows the preferences of a client, it can render the content in the format that is desired and accepted by the client. However, in traditional network delivery approach, the server may not be able to know the preferences. A server located in the Eastern United States may be accessed by clients across the world. It is hard for the server to know the native language of a client. CDNs application can deal with this problem well. Multiple caching sites of CDNs are present and located in geographical places, a caching site has a high probability of getting clients with default language preference for the local language. Although the original site in the Eastern United States with English, a caching site located in China can translate all document into Chinese by default, and a caching site located in France can translate all document into French by default. At the same time, CDNs can offload and reduce the need for processing cycles at the original site.

In some cases, clients need use hand-held devices or palmtops to access information in Internet or Intranet instead of desktop computers. As a result of variety of access devices, the simple HTML formatted page can no longer fit the needs of all clients. WAP (Wireless Access Protocol) protocol rather than HTTP protocol should be used for these clients who use palmtop devices. In traditional network delivery approach, in order to support WAP, it needs every site has a translator to translate content from HTML to WML ("Wireless Markup Language is a markup language based on XML, and is intended for use in specifying content and user interface for narrowband devices, including cellular phones and pagers" [92]). With application of

CDNs, it is not very necessary for all the site to support the WAP protocol as traditional approach. You only let some caching sites have the translator to support WAP rather than original site. Therefore none of the original site needs to support WAP [1].

Mail Server

Mail server is another application of CDNs. A mail server is a machine that stores electronic mail for users and allow them to read and send mail as needed. A variety of protocols can be used between the user and mail server. Many Internet Server Providers as well as many other sites, such as Hotmail (http://www.hotmail.com) or Yahoo (http://www.yahoo.com) provide mail server to their users. If these mail provider sites locate somewhere only one place in the Internet to response all accessing of their users, when a user is in a remote location, the perceived performance of the mail server may become rather poor. In this case, the performance can be improved significantly by moving the web server closer to the client by exploiting a CDNs. The mail server would provide better mail access to the clients by caching their content closer to them. In order for the scheme to work well, clients should typically be accessing the same caching site for extended periods. When a client is first directed to a caching site, the caching site obtains the mail file belonging to the client and cache it locally. It may use a protocol like POP ("Post Office Protocol: POP is a protocol that allows a client to download a mail database from a remote server to the local file system "[3]) to implement this function. The client manipulates the mail file on the caching server by using the cache to read, send, or delete mail message. At the same time, CDNs will also pay attention to the registration and authentication of a client for caching sites and original site. At periodic intervals, the mail database maintained at the caching sites is synchronized back to the mail servers at the original site [1].

Shopping Sites

One of the most common types of sites on Internet are the on-line stores that offer electronic sales of item ranging from books, CDs, and software to airline tickets and even everyday grocery items. The performance of many of these shopping sites can be improved significantly by applying CDNs. In most on-line stores, most clients typically browsed through the catalogues rather than make actual purchase. In the traditional approach, all the clients access only original site. If too many clients are concurrently access, everyone will get a poor performance. Applying CDNs will be able to avoid this situation. The browsing function is read-only operation and can be moved to the caching sites within a CDNs. A client can browse through the entire catalogue or search for a special page. The caching site can satisfy these requests by either maintaining a replicated copy of the entire catalogue or by using query caching techniques (Query caching will work well provided some types of queries that are performed more frequently than others). The purchasers at store need to provide some private or secure information such as clients' credit card number, mailing address, etc. So these purchasers' information is better to keep and deal with in the original site instead of caching sites [1]. In a real on-line store, a necessary function associated with browsing is the maintenance of a shopping cart. Within the CDNs application, the shopping cart could be kept and maintain in caching sites. When the client is ready to purchase the items in the shopping cart, the request will be sent to the original site.

Besides these application of CDNs mentioned above, there are also many other web-based applications, such as on-line auctions, search engines, real estate brokerage sites, etc or non web application. In general, all these application have similar features to be suitable for CDNs: either it has a large percentage of read-only operation, or being accessed more frequently, or can operate without strong consistency in its data.

7.3 Case Study – CDN Application for E-learning

In the last section, we discussed and analyzed several generic applications that are suitable for the CDNs environment. Actual case of CDNs, such as E-learning, always combine several application of CDNs. In this section, we select Cisco Enterprise Content Delivery Network (ECDN) and Volera Velocity CDN for E-learning as an example to do a case study to show the application of CDNs in practice.

With the development of broadband networks and powerful computer systems, online education or E-learning begins to emerge and attract more and more companies' attention for many reasons. According to the analysis of IDC, the E-learning market increased very quickly from $4.05 billion in 2001 to $11.41 billion in 2003 [89]. E-learning enables these companies to reach audiences anytime and anywhere, centralize information, reduce costs, and improve productivity and competitiveness. To overcome the online learning issues such as surrounding quality, and network congestion, CDNs application is a good solution to handle these challenges.

Cisco Enterprise Content Delivery Network (ECDN)

Cisco ECDN can transparently redirect learner requests for online learning content to the most appropriate source for delivery by global network of intelligent edge nodes. This intelligent routing and delivery works for static and streaming multimedia web content, email or transaction processing, as well as learning or communications content. Cisco ECDN also provide the capability for live broadcasts. ECDN mainly consists three parts: Cisco Content Distribution Manager, many Cisco Content Engines, and an IP/TV Broadcast Server [88]. Cisco Content Distribution Manager works as a decision maker, which applies DNS-based method combined with URL rewriting ones to redirect the clients to the right site (More information includes strong points and shortcomings, please see 5.2.2.2, 5.2.2.3 and 5.3.1). Cisco Content

Distribution Manager also takes responsibility for avoiding overload on one caching sever. Cisco Content Engines are installed on various locations and work as edge servers for caching content to response clients. In addition, Cisco also has a special device - IP/TV Broadcast Server, which works for caching and delivery media streams and live MPEG video. Deploying Cisco ECND can distribute any live and on-demand content, and eliminate bandwidth bottlenecks.

Volera Velocity CDN

Volera Velocity CDN can make rich online learning, such as live or on demand to employees, partners, suppliers, and customers anywhere in the world. It is mainly composed of four elements: Volera Excelerator, Volera System Controller, Volera Content Controller and Volera Content Accountant. In addition, Volera Velocity CDN also can include a Volera Media Excelerator for caching and delivery streaming media content.[91][93] Volera Excelerator is cache instead of original site to delivery content to clients. Volera System Controller works as a monitor for Excelerator caches. Volera Content Controller works for caching content update and management. Volera Content Accountant is an accounting and reporting system that enables network operators to report on the delivery of content to end-users or corporations to charge-back departments using the CDN (More information includes strong points and shortcomings, please see 5.6.2 and 5.6.3). Volera Velocity CDN solution does not have its own Redirection Equipment, it requires third-party technologies to deliver request routing solutions. Nortel Networks Alteon is always used as GSLB in Volera Velocity CDN solution for redirecting client to the most appropriate site and avoid overload in the network. Deploying Volera Velocity CDN can provide faster, richer quality of content for education online.

We compare some main features of Cisco ECDN and Volera Velocity CDN in Table 7.1

Case	Cisco ECDN	Volera Velocity CDN
Company	Cisco	Volera
Website	http://www.cisco.com	http://www.volera.com
Main Components	Cisco Content Distribution Manager, Cisco Content Engines, IP/TV Broadcast Server	Volera Excelerator, Volera System Controller, Volera Content Controller, Volera Content Accountant, Volera Media Excelerator
Decision Maker	Cisco Content Distribution Manager work by DNS-based and URL rewriting	Requires third-party to deliver request, such as Nortel Networks Alteon work by GSLB to redirect clients
Support Live Video	Yes	Yes
Media Devices	IP/TV Broadcast Server	Volera Media Excelerator
Support Content Type	Static, streaming Multimedia, Email or Transaction processing, Live broadcasts	Flash, HTML, Static Content, Medial Streaming, Live broadcasts
Application Type on E-learning	Live broadcasts, Video on demand, Virtual classroom, Interactive multimedia, Virtual labs, Management and assessment	Live broadcasts, Video on demand, Virtual classroom and lab, Interactive training.
Benefits	It can eliminate bandwidth bottlenecks, distribute any live and on-demand content and keep central management, improve e-learning content distribution and save cost.	It can provide convenient live and on-demand training options to employees and partners. This will save bandwidth costs by caching training modules and splitting streams to employees. It ensures a high-quality e-learning experience.

Table 7.1: Compare Cisco ECDN and Volera Velocity CDN

Chapter 8

Conclusions

In this thesis, we compared and evaluated CDNs with other congestion control & network improving methods, various CDNs components & implement and different CDNs products & services. In addition, we also analyzed CDNs applications. Now, we summarize CDNs approaches and make the conclusions. They consist of conclusions drawn from Chapter 4, Chapter 5 and Chapter 6, overall evaluation of CDNs (strength & weakness of current CDNs approaches), CDNs applications environment (suitable & unsuitable application for CDNs), and future issues (hot Points & threats to future CDNs).

8.1 Conclusions drawn from Chapter 4, Chapter 5 & Chapter 6

8.1.1 Conclusions for Chapter 4 --- Selecting Suitable Performance Improvement and Congestion Avoidance Methods

Chapter 4 mainly described and analyzed different congestion avoidance and network improving methods, which included Congestion Control, Capacity Planning, Quality of Service and Content Delivery networks.

We always apply Congestion Control and Capacity Planning in the network design in order to avoid network congest and overload. According to consumers' different

requirement, some choose QoS, some select CDNs. Through analysis, we know CDNs improves distribution and avoids congestion by replicated content from original site to caching site. It needn't drop packets or differentiate packets into different priority level artificially, which can avoid loss of packets as well as avoid the daunting task to decide which set of traffic flows should get preferred performance than others. However it is necessary to manage lots of geographically distributed CDNs sites or servers, which should be a complex and costly work. Different methods can be applied according to different requirement. In some case, more than one technology can be combined. For example, when we apply CDNs to improve network performance, we still need to forecast traffic in order to calculate the number of cache servers to deploy, which is used similarly as in Capacity Planning.

8.1.2 Conclusions for Chapter 5 --- Selecting Suitable Implementation to Build CDNs

Chapter 5 compared and analyzed various CDNs components and implementations. It included how to build an original site and caching sites, decision maker (specialized DNS server and global sever load balancer), sever selecting for client redirection approaches, interconnection within CDNs, enduring security and CDNs management.

Building original or caching site involves selecting a server in the site to respond to client requests. No matter what methods are utilized, the key is to select a lightly loaded sever, at a low cost, and to ensure availability and avoid a point of failure.

Decision Maker is one of key components to redirect client to most appropriate CDNs site. DNS and GSLB are mainly decision maker to redirect client to right site

according to site health, geographic location of the client, site proximity, and response time. In practical application, DNS and GSLB always integrate together. Decision maker can utilize active or passive schemes to determine the right site. Active scheme gets up to date information on network performance compare to passive scheme; passive scheme generates less overload traffic.

Private network and public network are two approaches for interconnection within CDNs. Private network can ensure better interconnection than public network, but it also has higher cost. Wireless is applied when some of content distribution sites are in remote locations without good connectivity, such as mountain regions or deserts where it is more difficult to built wired connectivity.

Ensuring security and effective management of CDNs sites, communications and content are also important points in CDNs implementation. Selecting different implement will according to different requirement. Sometime we should use a combination of approaches.

8.1.3 Conclusions for Chapter 6 --- Selecting Suitable CDNs Products and Service Solutions

Chapter 6 mainly evaluated and analyzed Caching Products, GSLB, SSL Accelerator and CDNs service solutions.

The main factors for selecting suitable product or service solutions are product performance & characteristics, management & operation, and price range & target market. Some products are integrated devices, which can effectively reduce hardware costs, save space, be easy to set up and simplify management. It is suitable in small to midsize networks. Some independent products have stronger

functionality than integrated devices, which is suitable for lager enterprises and special requirements. Some products are easier to install and management, however some others maybe more difficult and need additive equipment to work together. Different CDNs service solution has its own coverage area, type and redirection methodology. Consumers can select suitable product or service solution according to their requirements and product characteristics.

For examples: If I have a small company, I don't have strong and special requirement for CDNs product, however I hope to save hardware cost and IP space. Array 1000 is a good choose. It is an integrated web service device (Cache, SLB, SSL acceleration), easy to set up and simplify management. It also have reasonable price to be suitable for small networks.

8.2 Overall Evaluation of CDNs

8.2.1 Strength of Current CDNs Approaches

The strengths of CDNs lie in the fact that it adds intelligence to network infrastructure. This intelligence can be leveraged as a platform to host value-added services within the network infrastructure. Such value-added services include the proper distribution and storage of content. As a result, the consumer network edge can be leveraged for strategically placed value-added services in the data plane. We describe these main strength of CDNs as follows:

- CDNs bring content closer to its receivers, which results in faster download times. Then it avoids congestion in the network, and also results in better network performance.

- CDNs can handle more clients than general network services, it significantly improves the scalability of network.

- CDNs can store and deliver the content in any form, such as pictures, digitized audio and video, software, and text documents. At same time, it can prevent unauthorized users from accessing network and filter out inappropriate or offensive content access, such as virus.

- CDNs can reduce the sever load and protect servers from overload and flash traffic. Through configuring, managing, monitoring and upgrading edge servers, it can improve the availability and reliability of networks.

- CDNs preserve the existing customer relationship, generates a higher margin revenue stream. These will improve the enterprise business competitiveness.

8.2.2 Weakness of Current CDNs Approaches

Although CDNs have lots of strength, on the other hand, there is also drawback by using CDNs. The main drawback is its manageability and high cost, since CDNs increase the number of devices required, such as servers which also need geographically distributed. This perhaps leads to the expense for exploitation of a CDNs relatively high and reduce application performance. The main weaknesses of using CDNs are as follows:

- CDNs consist of many sites that are geographically distributed. Deploying CDNs within network also needs increase the number of devices required, such as many servers. These geographical sites or servers need to be managed and administered. Such administration is much more complex and higher cost than managing one server in a single location.

8.3 CDNs Applications Environment

8.3.1 Suitable Application for CDNs

The basic goal of CDNs is to improve the response time to users and to increase the scalability of systems. Many types of application have characteristics that make them a good match for CDNs. Most commercial offerings of CDNs are based on the web. The web browsing includes a large component of static data (Images, static HTML files, etc.), some types of dynamic data (Data which need update periodically. e.g., Product catalogue of online store), media file (Music clips or video images, etc.) and web-based shopping stores (surely these contain a combination of the first three: static/dynamic/media). CDNs are not only benefit for Web-based application, but also can improve performance to traditional Internet-based services, such as file transfer (ftp) servers or mail servers. Summarizing an application may be a good candidate for CDNs if it satisfy the following criteria.

- **It has a high ratio of reads compared to writes** Bring the content closer to the clients, CDNs can improve the performance of read request when the fraction of writes is a small part of the overall request. The overall systems will also see improved performance. For example, an online store at which 90% of requests are simply browsing through store and only 10% result in purchase. Browsing is

for reading product catalogue, which has large fraction of overall request. It can be handled by caching sites. Purchase needs writes that means real transaction. Because of issues of security and authorization, these purchase transaction should only be dealt with by original site, however it has less fraction than reads.

- **Client access patterns tend to access some set of objects more frequently** Different clients perhaps have different requirements for the content replicated in caching sites. If the total content presented at the original site is replicated at the caching site, of course it can satisfy all the clients' requests, but it also needs more resource and response time. If some set of objects to which the clients access more frequently, by using CDNs to replicate these fraction of content, it can significantly improve systems performance.

- **Limited windows of inconsistent data are acceptable** CDNs always use caching sites to cache data and service to clients instead of the original site. So some inconsistencies in the data ought to be acceptable.

- **Data update occur relatively infrequently** Caching site inevitably updates data from the original site according to the requirement within period of time. However if the rate of data update is too frequent, it will significantly increase operation cost of CDNs. Therefore, suitable application for CDNs should be data update occur relatively slow. For example, an online store always keep the fixed product catalogue for customer reading. The changes only would be done when the store needs to update the information in its product catalogue periodically.

Some detail description and analysis about applications for CDNs were discussed at length in Chapter 7.

8.3.2 Unsuitable Application for CDNs

Not all applications are good candidates for a CDNs. There are some main candidates following that are not a good match for CDNs.

- **Applications modifies shared data and requires concurrency control** For example, a single chat room where information from multiple sources needs to be combined in sequence is not very suitable for performance improvement using CDNs.

- **Applications requires strong consistency of data** Thus, if the department database are frequently updated and needs to keep absolute consistency, CDNs may not be a proper solution. In these cases, CDNs need to acquire locks and their response time will be degraded. For example, share-list of Wall Street and online stock jobbing, which requires absolute consistency to ensure benefits of investors. Although deploying special CDNs solution can improve their distribution performance, it will increase high extra fee to ensure their consistency and degrade response time. (More information, please see 5.6.2.3).

- **Applications have strong security** In this case, it is maybe not a technical issue. Generally, it is easier to build a single secure site than to maintain good security at multiple, geographically distributed caching sites. If application needs strong security, it is unsuitable to use CDNs, since security always increases expense of operation. For example, when it always needs customers to submit their private information, such as credit card number, personal telephone number, etc, it'd better to be handled by original site instead of caching sites. So the function of CDNs can't bring into play.

- **Applications have licensing needs** In other case, licensing agreements may prohibit storing copies of data (such as music or video files) or applications (such as some software licenses) at the caching sites. Here, legal and policy issues are the big obstacle for CDNs using [1].

8.4 Future Issues

8.4.1 Hot Points of Future CDNs

Although CDNs technology emerged in the world just less than 10 years, it has got big success. Now there are currently lots of companies focusing on this field and more and more companies pay attention to it and will enter CDNs area. For the future development of CDNs, the hot points are maybe as follows:

- **Stream Media Delivery** Combined with replication technologies CDNs have potential to offer efficient multicast delivery of especially rich content. A major future CDNs market will be streaming media. As streaming gains widespread adoption, CDNs market growth will accelerate. As a result, the cost of CDNs products and services will decrease over time, which will drive adoption rates up. New value-added services for content distribution, adaptation, or negotiation can easily be implemented and offer large opportunities for a successful future of CDNs.

- **CDN companies integration** Traditionally, telecom companies and content creators focus different area. Now more and more satellite and broadcasting companies deliver their TV program via Internet as well (video on demand, e.g.,

www.omroep.nl, www.bbc.co.uk, or www.cnn.com). Noting that telecom companies and content creators have been integrated (e.g., AOL and Time Warner, www.aoltimewarner.com; or Telefonica and Endemol Entertainment, www.endemol.com. This is an interesting area to create win-wins. On one hand, since some big content providers either want to reduce their rental cost of content delivery service from other CDSP (Content Distribution Service Provider), or they have special requirement which general content delivery service can't satisfy. On the other hand, as CDSP, which also hope to focus a special area and bind its customers, this will reduce its cost of technology research & development, and bring fixed profit. In this case, both of them will have demand for integration. Maybe, in the future, we might find a CDSP such as Akamai merge with a content provider such as CNN or HMV.

- **CDNs and ASP synergy** The distribution of content can integrate CDNs and the ASP of databases (Storage Service Provision). With Internet becoming a large archive, indexing and tagging data becomes important. For CDNs, an important point is content management for searching and retrieval. CDN and ASP synergy can effectively work for data management issues. Similar issues include data warehousing, and the persistence of digital storage formats in time.

- **CDNs and GRID synergy** GRID ("Allowing access to dispersed information, and knowledge discovery and extraction from spread knowledge resources. They make use of cognitive techniques and tools such as data mining, machine learning, content semantics, ontology engineering, information visualization and intelligent agent "[94]) is to be interpreted in a broad sense including technologies for Peer-to-Peer Computing (P2P). It uses commodity hardware to enable the coordinated use of geographically distributed resources without central control, which can reduce the workload of DNS or GSLB. Middleware platforms of CDNs combine with distributed operating systems and parallel computing with GRID may increase distribution performance.

- **Personalization and diversity** Mobility is a trend, which includes personalization and diversity. Traditional CDNs mainly focus on the content: organizing and delivering it. With mobility showing up, not only the content, but also the context becomes important. Content adaptation for mobile and wireless devices, adapting content to personal preferences and location-based services show up. For example, the clients accessing information in Internet or enterprise Intranet can either use desktop computer as well as palmtops. These need CDNs have mobility to be able to support multiple protocols, such as HTTP protocol for browsers with desktop computer, and WAP for wireless devices. (More information, please see 7.2)

- **Globalization** From a business point of view, future CDNs mean globalization. With the development of technology and increase of client requirement, CDNs should develop and grow from local service to globalization, the factors like authentication and authorization of users, accounting, etc will increase. For example, now Akamai has more than 13000 servers in 60 countries within 6 big continents. It has become a global CDNs provider, although it has been a small company from its beginning in the early 1995.

8.4.2 Threats to Future CDNs

Although it seems a beautiful prospect, there are also several threats which could spoil this prosperous CDNs future. We briefly describe them as follows:

- **Legal Issues** Content caching and replication is an important functionality in CDNs. However if content owners don't agree with such distribution of their content legally, CDNs can not show the function.

- **Security Aspects** Security aspects are always the topic in network. It is hard to express clearly, such as authentication, authorization or denial of service, which perhaps become threats to CDNs sometimes.

- **High Expense** The current CDNs technology and business models are changing rapidly. Enterprises maybe need to invest huge amount of money in the CDNs research and development. How to make a balance between cost and profit of CDNs also is a big problem worthy of thinking.

These threats to future CDNs either come from technology or non-technology area. However, we believe if we pay more attention to these problems and try to avoid them consciously, we will get a successful future of CDNs.

Bibliography

[1] Dinesh C. Verma, "Content Distribution Networks An Engineering Approach", A Wiley-Interscience Publication, 2002

[2] Dougles E. Comer, "Computer Networks and Internet"; Second Edition, Prentice-Hall Inc., 2000

[3] "Microsoft Press Computer Dictionary", Third edition; Qinghua University Publication, 1999.

[4] Wouter B. Teeuw, "Content Distribution Networks", Telematica Instituut, 31/01/2002
http://www.telin.nl/Middleware/cdn/ENindex.htm

[5] Scot Hull, "Content Delivery Networks: Web Switching for Security, Availability, and Speed"; McGraw-Hill/Osborne, 2000

[6] Jian Lu, Enjoy Web, Inc. "Reactive and Proactive Approaches to Media Streaming From Scalable Coding to Content Delivery Networks", IEEE, 2001,
http://www.enjoyweb.com/company/pdf/itcc2001.pdf

[7] Akamai White Paper, "Delivering the Profits: How the right content delivery provider can drive traffic, sales and profits through your Web site", Akamai Technologies, Inc. 2000
http://developer.akamai.com/pdf/Delivring_The_Profits.pdf

[8] Keynote White Paper, "Using Keynote Measurements to Evaluate Content Delivery Networks", Keynote Systems, Inc., 2000

http://www.avoka.com/resources/keynote/wp_cdn.pdf

[9] HTRC White Paper, "The Content Delivery Network Market", The HTRC Group LLC., 2000,
http://www.htrcgroup.com/pdffiles/cdn.pdf

[10] "CDN Sota",
http://www.telin.nl/Middleware/cdn/state-of-art/cdnsota_toc.html

[11] Jan Ozer, "The Technology Basics and Keys to Provider Selection", June 12, 2001, http://www.extremetech.com/article2/0,3973,12761,00.asp

[12] "Glossary - SSL & Network Security Terms ", SSi Service Strategies Inc.,
http://www.ssl-technology.com/glossary.htm#R

[13] K. Nichols, S. Blake, F. Baker and D. Black, IETF RFC 2474, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", 1998

[14] Internetworking Technologies Handbook, "Network Technology Fundamental"; Chapter 49, "Quality of Service Networking"
http://www.cisco.com/univercd/cc/td/doc/cisintwk/ito_doc/qos.htm

[15] Sally Floyd and Van Jacobson, "Random Early Detection Gateways for Congestion Avoidance", IEEE/ACM Transactions on Networking, August 1993,
http://www.icir.org/floyd/papers/red/red.html

[16] "Distributed Weighted Random Early Detection",
http://www.cisco.com/univercd/cc/td/doc/product/software/ios111/cc111/wred.pdf

[17] Silvano Gai, "Content Delivery Networks", Cisco Systems, USA.

http://staff.polito.it/~baldi/RetiCalcSPEO/lucidi/CDN.PDF

[18] "Application Note: Global Server Load Balancing with ServerIron"; http://www.foundrynet.com/solutions/appNotes/GSLB.html#fig1

[19] White Paper, "Enhancing Web User Experience with Global Server Load Balancing", Alteon WebSystems, Inc.
http://www.nortelnetworks.com/products/library/collateral/intel_int/gslb_wp.pdf

[20] Matthew Liste Thrupoint: "Content Delivery Networks (CDNs) – A Reference Guide";
http://www.ciscoworldmagazine.com/webpapers/2001/03_thrupoint.shtml

[21] A. Barbir, B. Cain, F. Douglis et el, "Known CDN Request-Routing Mechanisms", Internet-Draft, Expires: Dec. 1. 2001;
http://www.content-peering.org/docs/draft-cain-cdnp-known-request-routing-00.html

[22] Cintel FAQ, "CDN Solution",
http://www.cintel.co.kr/eng/supot/faqg/supot_faqg_pro.asp

[23] Craig Wills, Yin Zhang, " On the Use and Performance of Content Distribution Networks", Balachander Krishnamurthy, 2001
http://www.acm.org/sigcomm/imw2001/imw2001-papers/10.pdf

[24] A. Shaikh, R. Tewari, and M. Agrawal, "On the Effectiveness of DNS-based Server Selection", Proceedings of IEEE INFOCOM, 2001 April 2001,
http://www.research.ibm.com/people/a/aashaikh/research/papers/infocom01.pdf

[25] J. Challenger, et al., "A Publishing System for Efficiently Creating Dynamic Web Content ", Proceedings of IEEE INFOCOM 2000, http://www.cs.duke.edu/education/courses/spring02/cps296.1/papers/CIWFR-INFOCOMM2000.pdf

[26] Dinesh C. Verma, Seraphin Calo, and Khalil Amiri, IBM Thomas J Watson Research Center "Policy-Based Management of Content Distribution Networks"; http://www.research.ibm.com/people/d/dverma/papers/CDNPolicy.pdf

[27] Irwin Lazar and William Terrill "Exploring Content Delivery Networking"; http://www2.dc.net/ilazar/content.pdf

[28] John Dilley, Bruce Maggs, Jay Parikh, Harald Prokop, Ramesh Sitaraman, and Bill Weihl, " Globally Distributed Content Delivery", Akamai Technologies, Proceedings of IEEE INTERNET COMPUTING, September - October 2002, http://www-2.cs.cmu.edu/afs/cs.cmu.edu/project/phrensy/pub/papers/DilleyMPPSW02.pdf

[29] " Content Distribution Networks and Quality of Service", file:///C:/WINDOWS/Temporary%20Internet%20Files/Content.IE5/Q7CR8J2B/preg10t10%5B1%5D.ppt#283,1

[30] Kirk Johnson, John Carr, Mark Day, Frans Kaashoek, " The Measured Performance of Content Distribution Networks", file:///C:/WINDOWS/Temporary%20Internet%20Files/Content.IE5/BYRLCIG5/S4-1%5B1%5D.ppt#256,1

[31] "ClickArray Networks, Inc. Array 1000 Web Server Offload Performance Evaluation", Publisher: The Tolly Group Date Published:01/08/01, http://researchandmarkets.com/reportinfo.asp?cat_id=129&report_id=1486

[32] "Clickarray Networks, Inc. announces Array 1000, the industry's first true integrated web traffic management appliance for websites, MSPs, HSPs, and ISPs", May 7, 2001, http://www.arraynetworks.net/pr_05082001.html

[33] "ClickArray's Array 1000 Integrated Web Device Falls Short of Corporate-Class", Lori MacVittie, October 29, 2001, http://www.networkcomputing.com/1222/1222sp3.html

[34] "A One-Stop Web Appliance", Francis Chu, October 15, 2001, http://www.eweek.com/article2/0,3959,34514,00.asp

[35] " Server Accelerator 7000 Series, Content Intelligent Networking Products for Content Providers", http://www.teneo.co.uk/site/products/sa7000.htm

[36] "Cacheflow Products information", http://www.asiasoft.com.hk/cacheflow_info.asp

[37] "Network Appliance NetCache C6100", http://www.b2net.co.uk/netapp/network_appliance_netcache_c6100.htm

[38] "C6100 Products information", http://www.ankor.co.il/products/products_content_kfir2.asp?categ=10&subcateg=90

[39] "Cisco Cache Engine 500 Series", http://www.netfastusa.com/Technology/Content.htm

[40] "smart plug-in for inktomi traffic server" http://www.openview.hp.com/products/smartplugins/spis/Documents/Product_HTML-465.asp

[41] "Volera Products Information"
http://www.volera.com/products/acceleration/excelerator.html#how

[42] "Volera Excelerator",
http://www.volera.com/products/whitepapers/documents/excelerator.pdf

[43]. "Performance Comparison of ArrayNetworks Array 1000 Layer 4-7 Web switch vs. Cisco CSS 11800 and Alteon180e".
http://www.etestinglabs.com/clients/reports/arraynet/arraynet.pdf

[44] "GSLB Comparison: F5 3-DNS, Foundry ServerIron, Nortel Alteon WebOS, Cisco Arrowpoint", http://www.f5.com.sg/f5apac/default.asp

[45]. "Web Server Directors (WSD)",
http://www.internet2000.com/TransitPages/RadWare/WebText-RadWare_WSD.htm#P0_0

[46]. "GSLB Comparison: F5 3-DNS, Radware RND-NP, Cisco Distributed Director",
http://www.f5.com.sg/f5apac/default.asp

[47]. "Radware Web Server Director",
http://www.radware.com/content/products/wsd/default.asp

[48] "Distributed Director Product Overview",
http://www.alliancedatacom.com/manufacturers/cisco-systems/content_delivery/distributed_director.asp

[49] "Alteon SSL Accelerator",
http://www.nortelnetworks.com/products/01/alteon/isdssl/index.html

[50] White Paper, "SSL Performance and Capacity Planning", SonicWALL Inc., http://www.firewallsdirect.com/white_papers/sonicwall/SSL_planning.pdf

[51] San Jose, "Cavium Networks Introduces Industry's Highest Performance Network Security Processor Family" , Cavium, October 15, 2001 http://www.cavium.com/newsevents_products.html

[52] TECH notes, "Importing a Global Server Certificate from Verisign and other PKCS#7 certificates into the SonicWALL SSL Offloader", Prepared by SonicWALL, Inc., 07/30/2002 http://www.verisign.com/support/install/sonicwall/sonicGlobal.pdf

[53] http://www.openssl.org/docs/ssl/SSL_CTX_set_tmp_rsa_callback.html

[54] White Paper, "The Importance of Secure URL Re-Write", SonicWALL Inc., http://www.firewallsdirect.com/white_papers/sonicwall/SecureUrlRewrite.pdf

[55] White Paper, "Avaya SSL 100", Avaya Communication, http://www1.avaya.com/enterprise/brochures/vpn1395.pdf

[56] " Intel NetStructure 7110/7115 e-Commerce Accelerator", http://www.intel.co.jp/support/netstructure/commerce/31044.htm

[57] "Smashing the SSL Speed Trap", Lori MacVittie, June 11, 2001 http://www.networkcomputing.com/1212/1212f49.html

[58] "new Intel NetStructure 7115 e-Commerce Accelerator", http://www.gotocol.com/intel7115.html

[59] " Server Accelerator 700 Series, Content Intelligent Networking Products for Content Providers", http://www.teneo.co.uk/site/products/sa700.htm

[60] "CacheFlow 700 Product Information"
http://www.cacheflow.com/products/700/specs.cfm

[61] "Enhanced speed and superior scalability for secure on-line processing",
http://www.f5.com/f5products/bigip/Ecom540/

[62] "Cisco/ArrowPoint Content Smart Web Switches"
http://isp-lists.isp-planet.com/isp-equipment/0203/msg08981.html

[63] "Alteon Products Information",
http://vegan.net/MRTG/alteon.php

[64] "SSL-R", http://www.sonicwall.com/products/sslr.html

[65] http://www.ssl-technology.com/ssl_comparison.htm

[66] "CDN service providers",
http://www.telin.nl/Middleware/cdn/state-of-art/cdnsota_CDN_service_providers.html#IX_3

[67] Craig Wills, Yin Zhang, " On the Use and Performance of Content Distribution Networks", Balachander Krishnamurthy, 2001
http://www.acm.org/sigcomm/imw2001/imw2001-papers/10.pdf

[68] "Adero Expands Its GlobalWise Services to Enable Hosting Providers, Carriers, and ISPs to Enter Content Distribution Market", WALTHAM, Mass, Jan. 5, 2001
http://www.webhostnews.com/press/january2001/5/0105adero.shtm

[69] "Content Delivery Network Services"
http://www.webreference.com/internet/software/site_management/cdns.html

[70] "Backbone Content Delivery",
http://www.ecomlink.org/E_incubator/Best_Sites.asp?CategoryID=019

[71]" Content Delivery Systems",
http://www.lexisone.com/legalresearch/legalguide/internet_business_solutions/content_delivery.htm

[72] "Content Delivery Networks: An Introduction", White paper, May 2002
http://cdn.hcltech.com/WhitePapers/CDNIntroductionWP.pdf

[73] "Cidera Services Portfolio"
http://www.cidera.com/services/index.php#

[74] White Paper, "GlobalWise Content", By Webvisions,
http://www.webvisions.com/adero/globalwise_content.htm

[75] "Content Delivery From The Source"
http://www.isp-planet.com/technology/clearway_cdn.html

[76] "Globeport Functional Specifications"
http://www.chriscuster.net/functional_specs.html

[77] "iBeam Product Description"
http://www.internetacceleration.com/vendor_profiles/iBeam%20Williams.htm

[78] "Overview of CDN organisations"

http://www.telin.nl/Middleware/cdn/state-of-art/cdnsota_Overview_of_CDN_organisations.html

[79] "Proxy Cache Comparison" http://www.web-caching.com/D:/Thesis/new/Proxy%20Cache%20Comparison.htm

[80] http://www.telezoo.com/asp/sc/sc.asp?idcats=788&history=^709^721

[81] http://www.web-caching.com/cdns.html

[82] "Streaming media filter", Microsoft TechNet, http://www.microsoft.com/technet/treeview/default.asp?url=/technet/prodtechnol/isa/proddocs/isadocs/cmt_streamfilter.asp

[83] "Glossary - SSL & Network Security Terms ", SSi Service Strategies Inc., http://www.ssl-technology.com/glossary.htm#R

[84] "SSL Aggregation and Session Re-use", SSi Service Strategies Inc., http://www.ssl-technology.com/ssl_aggregation.htm

[85] "SSL Initiation and SSL VPN's", SSi Service Strategies Inc., http://www.ssl-technology.com/ssl_vpns.htm - SSL initiation

[86] "128-bit SSL Digital Web Certificates - Generating a Certificate Signing Request (CSR) using Plesk Server Administrator 2.5", Century Web Design Ltd, http://certs.centurywebdesign.co.uk/support/csr_generation/plesk.html

[87] White Paper, "E-Learning Content Management vs. Content Delivery", Cisco Systems Inc., 2002 http://business.cisco.com/servletwl3/FileDownloader/iqprd/88758/kbns.pdf

[88] White Paper, "Tapping Into Cisco Enterprise Content Delivery Networks for High-Impact, Cost-Effective Learning", Cisco Systems Inc., 2001
http://www.cisco.com/en/US/netsol/ns110/ns49/ns50/ns264/networking_solutions_white_paper09186a00800a85a9.shtml

[89] Karen Brown, "Web Educated: Cisco initiative aims for promising e-learning market", Issue of Broadband Week, June 18, 2001
http://www.broadbandweek.com/news/010618/010618_apps_edu.htm

[90] White Paper, "The Cisco Internal Implementation of an Enterprise Content Delivery Network to Support E-Learning", Cisco Systems Inc., June 2001

[91] White Paper, "Volera Solutions for eLearning", by Volera, Inc.
http://www.volera.com/products/whitepapers/documents/elearning.pdf

[92] Technology Reports, "WAP Wireless Markup Language Specification (WML)", The Open Mobile Alliance Ltd., November 19, 2002
http://www.oasis-open.org/cover/wap-wml.html

[93] White Paper, "Building Content Distribution Networks with Volera Velocity CDN", by Volera, Inc.
http://www.volera.com/products/whitepapers/documents/cdn_whitepaper.pdf

[94] "CDN Mechanisms",
http://docencia.ac.upc.es/EPSC/Documentos/DistribucioContinguts.ppt

[95] http://www.clickarray.com

[96] http://www.cacheflow.com

[97] http://www.netapp.com

[98] http://www.cisco.com

[99] http://www.inktomi.com

[100] http://www.volera.com

[101] http://www.f5.com

[102] http://www.radware.com

[103] http://www.alteonwebsystems.com

[104] http://www.nortelnetworks.com

[105] http://www.foundrynet.com

[106] http://www.intel.com

[107] http://www.sonicwall.com

[108] http://www.adero.com

[109] http://www.akamai.com

[110] http://www.cacheware.com

[111] http://www.cidera.com

[112] http://www.clearway.com

[113] www.digitalisland.com

[114] http://www.epicrealm.com

[115] www.ibeam.com

[116] http://www.mirror-image.com

[117] http://www.pushcache.com

Appendices

Appendix I

Caching Products Descriptions from Vendors Websites

1. Array 1000

The Array 1000 from ClickArray Networks Inc. (http://www.clickarray.com) stands out among the caches because it integrates Web traffic acceleration, management and security components in a single device. It is purpose-built device that combines server load balancing, global SLB, reverse proxy caching, clustering and web server security (firewall functionality), SSL (Secure Sockets Layer) acceleration and content rewriter in order to optimize Web traffic performance [33][34][95]. According to the press release of ClickArray Networks Inc [32]: " with the Array 1000, hardware costs can be reduced by as much as 64%, operations costs by as much as 78%, rack space and power requirements by as much as 90% and maintenance costs by as much as 59% saving hundreds of thousands of dollars ". Array 1000 has some features as follows:

- " Array 1000 is an integrated Web service device that can save money and IT resources by consolidating hardware into a single, easy-to-set-up and managed device. It is best suited for service providers or small to midsize networks where an integrated Web appliance will be able to effectively reduce hardware costs, save space and simplify management. However, since Array 1000 integrates essential Web services into a single device, it can create a single point of failure if the appliance is not clustered. And clustering can quickly become expensive. Array 1000 is also not designed to replace enterprise-class firewall, caches and

load balancing hardware. Array 1000 might not be a good fit for large enterprises with complex Web architectures that are optimized for running custom applications and that require high levels of security and redundancy.

- " Array 1000's reverse proxy cache enhances the Server Load Balancer performance by caching static contents such as images and text files to RAM. After a client made a successful request, the cache fulfills subsequent requests, freeing up server CPU cycles to handle other tasks. It provides standard server load balancing with standard rules such as weighted round robin and least connections. It also easily sets up content-aware load balancing capabilities to map content requests to different server groups.
- " Array 1000 appliance provides a simple state inspection packet-filtering firewall. However, it doesn't offer VPN capabilities, since most firewall appliances provide this service. " [31]

Figure Appendix.1 shows Array 1000:

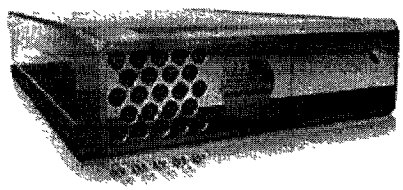

Figure Appendix.1: Array 1000 [31]

2. SA-7000

SA-7000 is a product optimized for server-side reverse proxy caching that are from Cacheflow (http://www.cacheflow.com). " SA-7000 is specifically designed to improve the performance, scalability, security and manageability of high-traffic Web sites. Deployed in minutes "in front of" any Web server, the SA-7000 dramatically

accelerates the delivery of Web content to users, and typically serves 5 to 10 times the content of a single Web server. This platform is ideally suited to organisations and applications such as E-Commerce, E-business, Content and Portal Sites, Enterprises and Web Hosting Providers. " [35][36][96]. The benefits of SA-7000 are:

- " Faster response times for Web site users (50 to 80%)
- " Higher site scalability (up to 10 times) to handle growing traffic volumes
- " Up to 90% lower space and power requirements
- " Significant capital and operational cost savings
- " Dramatically improved customer satisfaction and loyalty " [36]

Figure Appendix.2 shows SA-7000:

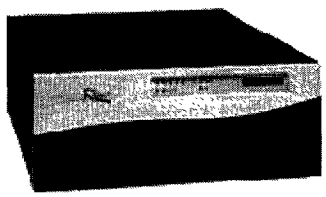

Figure Appendix.2: SA-7000 [35]

3. Netcache C6100

The NetCache C6100 is from Network Appliance Inc. (http://www.netapp.com). " NetCache C6100 can deliver the ultimate performance and reliability for the data centre and other high-bandwidth locations. The C6100 solutions support 180(+) Mb per second for HTTP environments and 1000(+) Mb per second for streaming applications. Large content libraries—up to 2TB of storage—can be reliably stored and protected with RAID support. Enterprises and ISPs use the NetCache C6100 to improve end-user response times, manage quality of service, reduce bandwidth

costs, and provide security and content filtering controls. With more than 99.99% uptime, the NetCache C6100 delivers reliable access to mission-critical data. "[38][97] NetCache C6100 has lots of features:

- " The NetCache C6100 maximize flexibility by consolidating support for all major streaming media formats, as well as HTTP, FTP, and NNTP, on one content delivery platform.
- " The NetCache C6100 has unlimited scalability, which is easily scale caching storage from 18GB to multiple terabytes. It also can reduce redundancy and bottlenecks to improve the performance.
- " Complete Streaming Media Support. Support for RealNetworks, RealSystem, Microsoft Windows Media, and Apple, QuickTime. For both video-on-demand and live stream splitting.
- " Easy Administration. NetCache C6100 ensures simple deployment, reliability, and ease of management. " [37][38]

Figure Appendix.3 shows NetCache C6100:

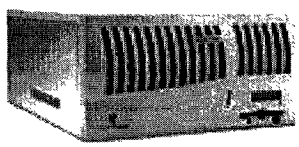

Figure Appendix.3: NetCache C6100 [37]

4. Cache Engine 590

Cache Engine 590 is from the Cisco Systems Inc.(http://www.cisco.com). " It is a high-end Cache Engine that services ISPs and large enterprise to reduce WAN bandwidth usage, accelerate network performance, and increase network scalability.

Utilizing the intelligence of the network, the Cisco Cache Engine 590 localizes traffic patterns by transparently caching frequently accessed content and then locally fulfilling successive requests for the same content. Because it is integrated into the network infrastructure, it is a low cost of ownership, enabling ISPs and enterprises to cost-effectively deploy the Cisco network caching solution on a wide-scale basis and gain the benefits of caching throughout their network. " [39][98] The Cache Engine 590 has the following key benefits:

- " Accelerated content delivery---The Cache Engine 590 accelerates content delivery by locally fulfilling content requests rather than traversing the Internet/intranet to a distant server farm. This solution helps to protect clients from uncontrollable bottlenecks, delivering more consistent network service quality and content availability.
- " Optimized WAN bandwidth usage---The Cache Engine 590 minimizes redundant network traffic that traverses WAN links. As a result, WAN bandwidth costs either decrease or grow less quickly. This bandwidth optimization increases network capacity for additional users and traffic and for new services such as voice.
- " Greater content access control---The Cache Engine 590, in conjunction with Web sense Enterprise software, enables administrators to monitor, manage and report on employee access to non-business and objectionable content. The result is increased employee productivity, lower bandwidth usage and reduced legal liability. " [39]

Figure Appendix.4 shows Cache Engine 590:

Figure Appendix.4: Cache Engine 590 [39]

5. Inktomi Traffic Server

Inktomi (http://www.inktomi.com) has been in the caching business for a long time and is now entering the CDNs market. " Inktomi Traffic Server is a scalable, reliable and high performance solution that accelerates delivery of web content and streaming media by storing frequently requested content closer to users. It is designed to help enterprises and service providers manage resources efficiently, enable rich media applications without expensive upgrades, and ensure content availability and quality of services. " [40][99] Traffic Sever has some benefit features:

- " Simultaneously distribute advanced services, such as delivering streaming media, filtering, content transformation and authentication services.
- " Improve performance and reduce "upstream" bandwidth usage by storing frequently requested content at the edge of the network, physically closer to the end users. Improves the quality of service for end users. Whether the end users are customers of an ISP or employees of a company, they will spend less time waiting for Internet content.
- " Optimizes bandwidth utilization and reduces infrastructure costs by serving requests from the local cache rather than always returning content from an origin server. Reduce hardware expenses by caching static and all major streaming formats in one integrated edge device." [40]

Figure Appendix.5 shows Inktomi Traffic Server:

Figure Appendix.5: Inktomi Traffic Server [40]

6. Volera Excelerator

Volera (http://www.volera.com) is the Novell spin-off dedicated toward Internet infrastructure, specifically regarding caching. " As a part of Velocity CDN solution, Volera Excelerator is an innovative content networking platform that works in flexible configurations to best meet clients content delivery requirements. Volera Excelerator accommodates multi-directional demands for content through two powerful configurations: Web browser acceleration (Web browser acceleration: In forward proxy mode, Volera Excelerator sits between Web browsers and the Internet to place content closer to end users and serve popular content from local cache) and Web site acceleration (Web site acceleration: In reverse proxy mode, Volera Excelerator intercepts requests on their way to client's Web server and then supplies frequently requested content from cache). " [41][100] Volera Excelerator has the following features:

- " Increase Web site visitor satisfaction and retention - Volera Excelerator improve enterprise's ability to attract and retain customers by creating, highly responsive Web experience for all site visitors. It can improve response times for customers, eliminate delays and refuse connections and provide a faster, richer quality of experience for visitors to enterprise's Web site.
- " Improve employee productivity - Volera Excelerator allows enterprise's employees to access frequently used content from the local network. As a result, enterprise's staff members can concentrate on doing their jobs and not on waiting for content from the Internet or corporate Intranet. It can enhance access to corporate data and Web content, improve employee productivity by eliminating long Internet waits and filter unproductive content and prioritize business-critical traffic. In addition, Volera Excelerator's filtering capabilities give enterprise control over which content may be accessed and stored, focusing network resources on content with business value.

- " Reduce IT expenditures--Optimize bandwidth use for reduced costs and leverage enterprises' existing Internet infrastructure to scale without adding costly Web servers and IT resources. With Volera Excelerator, it is an immediate drop in the amount of bandwidth necessary to deliver the same level of web content to users. In addition, Volera Excelerator economically increases Web site scalability. " [41][42]

Figure Appendix.6 shows Volera Excelerator:

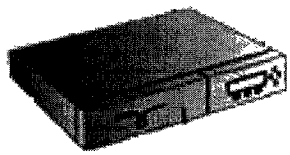

Figure Appendix.6: Volera Excelerator [41]

Appendix II

GSLB Descriptions from Vendors Websites

1. 3-DNS

3-DNS is the product of F5 Networks (http://www.f5.com). " The 3-DNS is a high availability and intelligent load balancing solution for geographically distributed Internet sites and data centres. It manages and distributes Internet requests across multiple, redundant server sites - regardless of the platform type or combination, and without requiring additional software on the client's servers. End user requests are distributed according to data centre and network conditions such as round trip time, packet loss and other QoS metrics, ensuring the highest possible availability for website ." [101] 3-DNS has the following features:

- " **Client Testing:** 3-DNS has several ways to actively check clients. All of the tests involve the 3DNS using big3d agents (The big3d agent collects performance information on behalf of the 3-DNS and runs on 3-DNS) at various sites test between each geographically distributed site and the clients LDNS server. This testing can take way of one of the three possibilities.

  - " UDP (User Datagram Protocol: The TCP/IP protocol that provides application programs with connectionless communication service. UDP is an alternative to TCP. You could call it an IP protocol, but not a TCP/IP protocol.): The big3d agent can do a NS lookup of the LDNS and measure response time, or completion rate.
  - " TCP (Transmission Control Protocol: The TCP/IP protocol that provides application programs with access to a connection-oriented communication

services): The big3d agent can do a TCP NS lookup of the LDNS and measure response time, or completion rate

- " Traceroute: The big3d can traceroute the path between the agent site and LDNS site. It can also just send a traceroute (UDP) packet to the destination site and measure response time or completion rate of it.

" Currently, 3-DNS does no dynamic passive probing of the network.

- " **Server Testing:** This is split into three major categories, using BIG-IP (BIG-IP is the local area application traffic management solution, ensuring high availability, reliability, security, and scalability for Web applications), using other vendors SLBs, or using No SLB.

  - " Using BIG-IP: 3DNS uses the big3d agent to query numerous data points from each site that has a BIG-IP. It can find out information such as number of connections per site, the availability of a VIP per BIG-IP health checking, the latency between that site and a particular LDNS (Long Distance Network Services), the number of packets/sec that particular site is passing.
  - " Using Other Vendors SLB: 3-DNS can make decisions based on information gathered via SNMP (Simple Network Management Protocol: The protocol that specifies how a network management station communicates with agent software in remote devices such as routes). Therefore it can still gather information like packets/sec, connections/sec, total connections, availability of a VIP. It cannot gather LDNS latency.
  - " No SLB: 3-DNS approaches this similar to how BIG-IP does, It can do a layer 3, 4 and 7 test to check the availability of the server. If the server has a MIB to read, 3DNS can also read it for performance information.

- " **Load Balancing:** The important thing to note about 3-DNS load balancing is that it is a three tiered system. Each LB algorithm can go in any tier, and the

request will be load balanced by the first tier capable of responding. The load balancing algorithms can be divided into three major categories.

- " Static, or those that add no intelligence to load balancing beyond server state. 3-DNS offers three static: Round Robin, Random and Ratio.
- " Dynamic, or those categories that take network information into account for load balancing. 3-DNS offers five different dynamic algorithms: RTT, Completion Rate, Hop Count, Least Connections, and Packet Rate.
- " Special, or those that offer static intelligence beyond what static can offer. They include those like Topology, QoS, and GA.

- " **GSLB Management:** This is the functionality that helps glue all of the other functionality together. " [44][101]

Figure Appendix.7 shows 3-DNS:

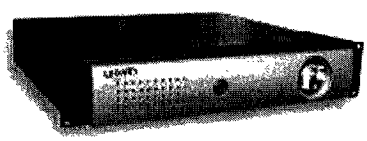

Figure Appendix.7: 3-DNS [101]

2. Arrowpoint CSDNS

ArrowPoint Communications, founded in 1997 on Acton, Massachusetts, was purchased by Cisco Systems (http://www.cisco.com) in the summer of 2000. It has series of products, including the CS-50, CS-100, CS-150 (stackables), and CS-800 (chassis) which were merged into the Cisco CSS line of load balancing switches [43][98]. Arrowpoint CSDNS has the following features:

- " **Client Testing:** Arrowpoint v3 does no client testing. Arrowpoint v4 uses a separate CS150 to maintain a database that does RTT (Round-trip time: the time required for a network communication to travel from the source to the destination and back. RTT is used by some routing algorithms to aid in calculating optimal routes.) testing and keeps information for the CSDNS. This testing uses TCP.

- " **Server Testing:** Arrowpoint can test servers through their SLB functionality and then share that information with any CSDNS service on geographically distributed Arrowpoint devices. Arrowpoint cannot test 3rd party SLB's, nor can they test hosts that aren't directly attached to an Arrowpoint device.

- " **Load Balancing:** Arrowpoint has three LB algorithms: Topology, Round Robin, and Least Connections. Their topology algorithm is limited to continental region.

- " **Management** : Arrowpoint products has a good capacity for IP Filtering. " [44][98]

Figure Appendix.8 shows Arrowpoint CSDNS:

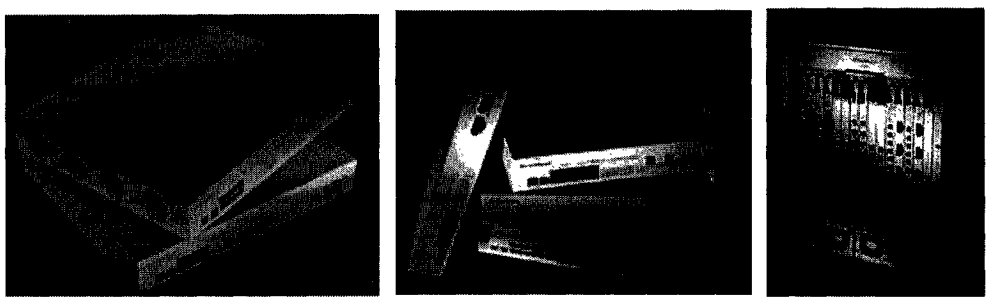

Figure Appendix.8: Arrowpoint C-50, C-150, C-800 [55]

3. Distributed Director

Distributed Director is the multisided, load-balancing solution of Cisco Systems (http://www.cisco.com). It transparently redirects end users to the closest responsive sever, which is determined by such factors as client-server proximity and client-server link latency [55]. Distributed Director has the following features:

- " **Client Testing:** Distributed Director has a schema that tests clients in one of two ways. The first client testing is a traditional TCP based RTT test. The Distributed Director (DD) signals to the agent DRP router that is sitting at each geographically distributed site. The router then does the TCP RTT request and returns the information to the DD. The second test that the DD can utilize is a passive hop count metric that has the router looking at BGP (Border Gateway Protocol) AS tables. This approach then compares the number of AS's away each geographical site is away from a client.

- " **Server Testing:** Distributed Director will test each server with a TCP null socket connection. This test is incremented in minutes.

- " **Load Balancing:** Distributed Director honours both tiers and weighting. They have several LB algorithms including RTT, AS hop count, administrator preference, and random. RTT and AS hop count make use of a DRP (Director Response Protocol. This is the protocol DD uses to query DRP server agents (routers) for BGP and IGP (Internal Gateway Protocol: A protocol that governs the transmission of routing information) routing metrics between the clients and the distributed servers) enabled router at the same geographic site as each server farm.

- " **Management:** DRP is a UDP based protocol that is not encrypted, but can be somewhat secured with ACL's (Access Control List: A list associated with a file that contains information about which users or groups have permission to access or modify the file). " [46][48]

Figure Appendix.9 shows Distributed Director:

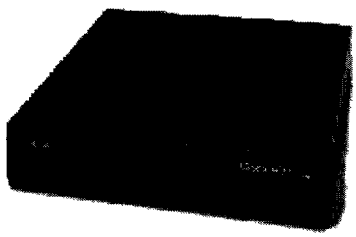

Figure Appendix.9: Distributed Director [48]

4. Web Server Director (WSD)

Radware(http://www.radware.com) Web Server Director (WSD) offers layer 4-7 server load balancing and GSLB. " It ensures the full availability, optimizes operation and completes security of server farms, as well as guarantees the reliability and highest performance of mission critical applications across the network. " [45][47] [101] WSD has the following features:

- " **Client Testing:** WSD has two ways to determine client location.

  - " ICMP: This method is used to find a RTT measurement between the client and the WSD devices.
  - " Traceroute: This method uses high port UDP (User Datagram Protocol) as defined by the RFC (Request for Comments) that includes traceroute.

- " **Server Testing:** The WSD tests clients using the NP as a Pro+ using the typical SLB methods. These tests do a fair job ensuring availability of a service between layers 4 and 7. For Non Radware endpoints, the WSD uses ICMP and TCP to test the availability of the server and service.

- " **Load Balancing:** WSD includes several load balancing options. The first option is triangulation that involves sending a request to one of the WSD devices. This device then performs its load balancing logic and either returns a HTTP redirect, or does an IP proxy to another WSD device. Radware allows for several algorithms to be implemented. These implementations are functionally straight-forward. They use UDP or ICMP for the measurement calculations. Multi-level redundancy and load balancing between servers, WSD units and distributed sites for complete business continuity across global networks.

- " **Management:** Radware products all have a nice SNMP based management console. " [46][47]

Figure Appendix.10 shows Web Server Director (WSD):

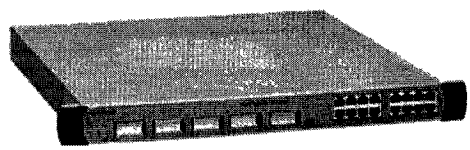

Figure Appendix.10: Web Server Director (WSD) [47]

5. Alteon WebOS GSLB

Alteon Systems (http://www.alteonwebsystems.com) was bought by Nortel Networks (http://www.nortelnetworks.com) in 2000. The main product of Alteon Systems is

WebOS GSLB, which provides load balancing, application redirection, high availability, bandwidth management and security services.[103] Alteon WebOS GSLB has the following features:

- " **Client Testing:** Alteon does no client testing. They do however do performance and latency testing between sites.

- " **Server Testing:** Alteon can test servers through their SLB functionality and then share that information with any CSDNS service on geographically distributed Alteon devices. Alteon cannot test 3rd party SLB's, nor can they test hosts that aren't directly attached to an Alteon device.

- " **Load Balancing:** Alteon has two basic load balancing algorithms, Topology and Server Performance. Alteon's topology allows user to put in networks and network masks and associate those networks with sites. The server performance test involves each geographically distributed Alteon device testing the VIP on its peers and sharing the response time information it gathers.

- " **Management:** The whole capacity of management is below Middle for Alteon. However it has a good capacity for IP Filtering. " [44][103]

Figure Appendix.11 shows Alteon Web OS GSLB:

Figure Appendix.11: Alteon Web OS GSLB [103]

6. Server Iron

Foundry Networks' (http://www.foundrynet.com) Server Iron product is a layer 4-7 load balancer with built-in GSLB functionality. " The Server Iron can employ DNS-based Footraces to determine site selection, in addition to monitoring RTT and load metrics. Foundry's Server Iron family of Web switches provides high performance content and application aware server load balancing. Server Iron extends the functionality of a traditional Layer 2/3 switch into higher layers by examining the content beyond the packet header to provide intelligent content switching." [105] Server Iron has the following features:

- " **Client Testing:** Foundry does no active client testing. Foundry tests for client space passively. They try and correlate client IP to client LDNS by net mask. The GSLB makes a load balancing decision based on which SYN-ACK combination was faster to each of the sites.

- " **Server Testing:** This category is split into three subcategories, utilizing ServerIron SLB, utilizing third party vendor SLB, and utilizing no SLB.

  - " Using Server Iron SLB: Foundry allows for L4 and L7 testing of each site from each GSLB. They have several canned L7 tests that span the common utilities, such as http, ftp etc
  - " Using Other SLB: Foundry has no solution at this time
  - " Using No SLB: Foundry has no solution at this time

- " **Load Balancing:** Foundry uses a seven tier ordered list. This list is as follows:

  - " The server's health
  - " The site Server Iron's session capacity threshold

- " The round-trip time between the remote Server Iron and the DNS client's subnet
- " The geographic location of the server and the site available session capacity
- " The site Server Iron's FlashBack speed (how quickly the GSLB receives the health check results)
- " The site Server Irons administrative preference
- " The Least Response selection

" This list can be reordered at will, but not weighted.

- " **Management:** Server Iron has a good capacity for IP Filtering. " [44][105]

Figure Appendix.12 shows Server Iron:

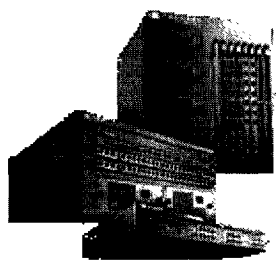

Figure Appendix.12: Server Iron [105]

Appendix III

SSL Accelerator Descriptions from Vendors Websites

1. Array 1000

Array 1000 is the product of ClickArray Networks Inc.(http://www.clickarray.com). Array 1000 is the Integrated N+1 Clustering, which consists Web security, server load balancing, global server load balancing, reverse proxy cache, SSL acceleration and CDNs content rewriter. The more information of Array 1000, please see Appendix I.1 Array 1000.

2. iSD-SSL 2.0

iSD-SSL 2.0 is the product of Nortel Networks (Formally Alteon Web Systems). " The Alteon (http://www.alteonwebsystems.com) iSD-SSL 2.0 Accelerator is a fully-featured Secure Sockets Layer (SSL) appliance integrating SSL acceleration, SSL extranet management, and secure application services into a single device. It is a powerful solution for seamless deployment into any network with the ability to manage high SSL traffic volumes, secure remote access, optimize back-end server infrastructure, and lower security costs. " [49] iSD-SSL 2.0 Acceleration has the following features:

- " SSL acceleration improves application performance by relieving servers of complex public key operations and bulk encryption
- " Secure application proxy deployments create instant SSL extranets for secure remote access to corporate applications

- " Integrated application services such as load balancing, session persistence, and Layer 7 filtering optimize secure application performance
- " SSL offloading improves server and application utility by recapturing lost performance
- " Digital certificate and key management consolidation lowers costs and simplifies operations; end-to-end encryption maintains data confidentiality and integrity in an accelerated environment. " [49][103]

Figure Appendix.13 shows iSD-SSL 2.0:

Figure Appendix.13: iSD-SSL 2.0 [49]

3. SA-700

SA-700 is a product from Cacheflow (http://www.cacheflow.com). It is specifically designed to improve the performance, scalability, security and manageability of high-traffic Web sites. " The platform is packaged in a compact 1U form factor, a major advantage in space-constrained data centres. The SA-700 hardware is optimized for Web server acceleration, featuring a high RAM-to-disk ratio and a built-in Secure Sockets Layer (SSL) encryption/decryption processor. This processor can manage 10-40 times more secure sessions than a standard Web server, allowing the SA-700 to accelerate the delivery of both public (HTTP) and private (HTTPS) content. clQ CacheOSTM Server Edition is the SA-700 Series system software which is expressly tuned for the workload of a high-traffic Web site. The software includes advanced features like an intelligent "Akamaizer", which automatically prepares

content for the Akamai FreeFlowSM network, and protection against Denial-of-Service attacks, which can crash a Web site. " [49][96] SA-700 has the following features:

- " Faster response times for Web site users (50 to 80%)
- " Higher site scalability (up to 10 times) to handle growing traffic volumes
- " Up to 90% lower space and power requirements
- " Significant capital and operational cost savings
- " Integrated SSL Cryptographic Processor. Processes 400 key negotiations per second, ~20 times the power of a standard Web server
- " Secure Content Acceleration. It can accelerate both public (HTTP) and private (HTTPS) content through integrated SSL functionality
- " Denial of Service (DoS) Protection. Prevents sites from crashing during hacker-initiated DoS attacks.
- " Robust security. Certified by a leading security audit firm for safe placement outside the firewall
- " Simple to Manage Appliance
- " Dramatically improved customer satisfaction and loyalty " [49][60][96]

Figure Appendix.14 shows SA-700:

Figure Appendix.14: SA-700 [60]

4. NetStructure 7115 e-Commerece Accelerator

NetStructure 7115 e-Commerece Accelerator is a product of Intel (http://www.intel.com). " It boosts SSL transaction performance with patent-pending technology that offloads cryptographic functions from the server. With its interoperable drop-in installation design, the Intel NetStructure 7115 e-Commerce Accelerator increases connection rates and decreases response times in e-Business data centres, enhancing the web experience for the clients. It supports a wide variety of cryptographic algorithms, including Blowfish and IDEA. The device reduced the average connection time for midsize Web objects. "[46][47][48][106] It has the following features:

- " Handles up to 600 connections per second and 6,000 simultaneous SSL sessions
- " Remote management and event monitoring capabilities
- " SNMP MIB (Management Information Base) & MIB II support
- " Client certificate authentication
- " Provides RC2, RC4, DES and 3DES level of encryption
- " Offloads SSL handling from e-Commerce servers
- " Interoperable with most servers and operating systems
- " Dynamically scalable architecture
- " Easy, drop-in installation between router and servers
- " Optional support packages through Intel NetStructure Support Services " [48]

Figure Appendix.15 shows NetStructure 7115 e-Commerce Accelerator:

Figure Appendix.15: NetStructure 7115 e-Commerce Accelerator [48]

5. e-Commerce Controller 540

e-Commerce Controller 540 is a product form F5 (http://www.f5.com). " It manages Secure Socket Layer (SSL) encryption/decryption to increase site performance and ability to scale SSL transactions. It offloads the burden of SSL traffic from web servers, increases server performance, and reduces administrative costs. " [61]e-Commerce Controller 540 has the following features:

- " Offloads CPU-draining SSL encryption/decryption from web servers - freeing them to serve more content
- " Support for Client Certificate authorization
- " Linearly scales to meet any SSL processing requirement by adding additional BIG-IP e-Commerce Controllers
- " Centralizes certificate management to a single source, greatly simplifying administrative duties
- " Eliminates the need to buy and install SSL-capable server software on every server within the network
- " Secure web management interface for simplified administration
- " Interoperability with BIG-IP load balancer or any 3rd party load balancing product
- " Deployed quickly within any network " [61][101]

Figure Appendix.16 show e-Commerce Controller 540:

Figure Appendix.16 e-Commerce Controller 540 [61]

6. SSL-R

SSL-R is produced by SonicWall (http://www.sonicwall.com). " It is a 1U rack mount, 2-port SSL Offloader that supports up to 200 peak RSA operations per second and up to 5,000 concurrent connections. The SonicWALL SSL-R integrates seamlessly with Layer 4 load balancers and Layer 5-7 content switches to enable secure and intelligent content networking, guaranteeing persistence for stateful applications while maximizing performance and content location flexibility. Designed for high-performance production networks, the SSL-R is a solid-state appliance with no moving parts and redundant power supplies for maximum reliability. Multiple SSL-R can be combined to create high availability for SSL transaction processing, ensuring access to mission-critical secure servers and data. Storing up to 255 keys and certificates directly in the SonicWALL SSL-R ensures complete independence from Web servers and minimizes the complexity and hassle of setting up SSL on Web servers. SSL-R guarantee end-to-end security and access to sensitive data and mission-critical applications with advanced features such as back-end encryption, SSL initiation, and SSL aggregation. SSL-R is an integral part of any application infrastructure that needs to accelerate performance and scale traffic of secure Web sites and applications, without the expense and headache of additional servers. "[64][107] SS-R has the following features:

- " High Performance. Support for up to 200 peak RSA operations per second and up to 5,000 concurrent connections offers increased performance and reliability for Web sites and commercial applications
- " Cost Effective. Dedicated high-performance SSL offloading eliminates the need for costly multiple-server deployment and maintenance, dramatically boosting the performance of the Web sites and applications at a fraction of the cost of deploying additional HTTPS servers.

- " Content Switch Friendly. Seamless integration with Layer 4 load balancers and Layer 5-7 content switches guarantees persistence for stateful applications, enabling secure content networking and maximizing content flexibility and hardware investment while increasing Web site efficiency.
- " Reliable. A solid-state architecture with no critical moving parts eliminates potential failures common to PC-based SSL appliances, while redundant power supplies ensure maximum reliability to keep the secure Web site and applications up and running. Multiple SonicWALL SSL-R can be combined to create highly available SSL transaction processing, reducing administration time, overhead, and costs.
- " Advanced Features. SonicWALL SSL-R offers the robust, comprehensive SSL solutions with Back-End Encryption, SSL Aggregation, and Secure URL Rewrite, that dynamically eliminates potential data exposure and risks common to most Web-enabled applications, with no additional coding or changes to the application. " [64]

Figure Appendix.17 shows SSL-R:

Figure Appendix.17: SSL-R [64]

Appendix Ⅳ

CDNs Service Solutions Descriptions from Providers Websites

1. Adero

" Adero provides high performance, quality enhanced content delivery solutions to carriers and hosting providers through the established Adero GlobalWise SM Network and content delivery services. The GlobalWise Network is comprised of strategically placed servers around the world, which redirect content closer to the audience for on-net enhanced services. GlobalWise Network is a worldwide network of intelligent, multi-server nodes that include carrier-grade Sun Enterprise servers and best-of-breed software that can make the site easily accessable from around the world or just around the corner. " [68][69] GlobalWise Network has the following feature:

- " High Performance: GlobalWise transparently directs each user to the closest Adero Intelligent Node SM and delivers the Web site content to users from that node.
- " Easy Global Web Deployment: GlobalWise eliminates the cost and trouble of installing and maintaining mirrored Web sites around the world.
- " Protection Against Overload: GlobalWise reduces the load on enterprise's own Web server preventing the performance problems that occur when crowds are drawn to the Web site.
- " Simple and Transparent Configuration: Adero GlobalWise is a service with no hardware or software to install and no changes to the branding or URL. So the configuration is simple.

- " Content Pre-population: Adero Content Pre-Population ("push") technology ensures that every user sees the latest content. Content Pre- Population uses the patented multi-cast technology to simultaneously deliver the fresh content to all of our Intelligent Nodes.
- " Superior Reliability: GlobalWise provides around-the-clock network monitoring and support.
- " Complete Data on User Traffic: GlobalWise provides complete log files, collated from the worldwide servers. " [74]

2. Akamai

" EdgeSuite is the a main CDN service solution of Akamai, which is an integrated suite of services for content and application delivery, content targeting, edge processing, business intelligence, and streaming media. " [109] EdgeSuite has the following features:

- " EdgeSuite enables optimize connection between the Enterprise's content server and the Akamai platform of edge server to improve the network performance.
- " EdgeSuite offload assembly to the edge and devote the Enterprise's content server to generating update content. Boost reliability and scalability.
- " EdgeSuite offers reliable and secure delivery and storage of all types of site content, including graphic objects, images, streaming media, digitized downloadable files, and HTML.
- " EdgeSuite optimizes the delivery of all types of site content through the tiered nature of the Akamai Platform. Tiered distribution hubs connected with edge nodes close to requesting users enable efficient distribution of content based on

unique characteristics and patterns as well as reduce the load on the Web infrastructure and avoid flash crowds.

- " EdgeSuite offers support for distribution of secure Web content. Using Secure Sockets Layer (SSL) transport.
- " EdgeSuite provides an economical way to mirror the site and provides a backup if any disaster causes the source site to crash. It also reduces infrastructure costs. " [109]

3. CacheWare

" EdgeSystem includes EdgeServer and EdgeManager, is a main CDN service solution of CacheWare, which is a low overhead server appliance that can reside anywhere in the corporate intranet or extranet and takes the load off an origin server by acting as the intermediary between origin and edge servers. EdgeServer pre-positions requested data to the edges of the network, allowing remote users quick access to relevant information, thus reducing demand to the WAN and conserving bandwidth. EdgeManager is a policy server that intelligently identifies and selects appropriate information for replication to the edge. An EdgeManager includes a Web-based administrator interface to search, select, distribute, monitor and manage content throughout the network, including remote locations. " [110] EdgeSystem has the following features:

- " Enables enterprises to better manage their existing network infrastructure
- " Automatically pre-populate business-critical information to the edge, automatically balancing it with traditional caching to ensure the optimum user experience
- " Identify hot content for prioritized deployment while enabling true enterprise-wide information mobility

- " Provides intelligent edge management and sophisticated bandwidth management
- " Uses a comprehensive set of solution methodologies to identify priority content for edge replication
- " Integrates with leading third-party content management platforms
- " Employs a sophisticated rules-based policy engine to develop detailed content-specific parameters that determines optimal content distribution
- " Automatically manage content freshness and coherency
- " Offers an open 'edge services' platform that supports emerging edge services such as advanced content filtering, virus scanning, and etc.
- " Forward and reverse proxy and complete transparency support
- " Robust management console " [110]

4. Cidera

" Cidera Streaming Media Service is a main CDN service solution of Cidera, which allows content providers, aggregators, and distributors to seamlessly and simultaneously deliver live streaming video and audio content into Cidera's servers located at numerous access points by Cidera's satellite broadcast network. Cidera Streaming Media Service delivers content directly to the edge, which eliminates router hops and bottlenecks on the Internet, and ensures that end users will receive the best possible performance. "[111] Streaming Media Service has the following features:

- " Avoid congestion on the Internet.
- " Provides turnkey solutions compatible with industry formats, including Microsoft Windows Media, Apple QuickTime, and Real Networks.
- " Allows massive scalability and international coverage.

- " Offers rapid service activation.
- " Expands the presence with consolidated access to the edge of the Internet.
- " Increases the bandwidth efficiency.
- " Provides access to other Cidera transport and distribution services.
- " Protects brand equity by delivering consistently high quality. " [111]

5. Clearway

" FireSite is the main CDN service solution of Clearway, which is completely transparent to the origin server and to existing infrastructure the hosting may use for high availability or load balancing. "[112] FireSite has the following features:

- " FireSite's turnkey installation and transparent operation make it easy for Web hosts, Web integrators, ISPs and other service providers to increase the scalability, reliability and performance of their customers' sites.
- " FireSite needn't increase burdens of adding CDN hardware and cost, as well as easy installations.
- " FireSite can see content change, query the operating system, and assess server load. " [112]

6. Digital Island

" GlobePort is a main CDN service solution of Digital Island. GlobePort is a high-speed connection service that extends e-business delivery to self-hosted customer servers, via two leased lines, frame relays, or other appropriate fast pipes. "[113] GlobePort has the following features:

- " GlobePort provides customers with self-hosted content delivery and DI-hosted administrative access.
- " GlobePort increase the data transport revenues by reaching a larger addressable market and providing flexible connectivity to DI's network. " [113]

7. epicRealm

" epicRealm caches static and dynamic content, database-driven content, and even encrypted content, which provides high performance, quality enhanced content delivery solutions to carriers and hosting providers through eXT Technology. eXT technology is the ingeniously simple collaboration between an epicRealm infrastructure acceleration appliance and web application. " [114] eXT technology has the following features:

- " eXT technology can automatically synchronize the outdated stored copy in caches at the central site and/or at the network edge. This ensures that users always receive accurate data.
- " eXT technology can accelerate of all content, either clear or secured by SSL, that is generated by infrastructure servers and delivered via HTTP and HTTPS.
- " eXT technology can cache of dynamic, database-driven and application-generated content and compression of HTML, XML, and other text file formats. It also improve performance.
- " eXT technology can ensure Availability and Security, avoid service (DoS) attack.
- " eXT technology can save cost. It has lower capital expenditures and radically reduced cost per transaction. It has appliance packaging, minimal maintenance, web-based administration, centralized monitoring. " [114]

8. iBeam

iBeam Intelligent Distribution Network is a main CDN service solution of iBeam, which sends the stream from the source straight to the satellite, then back to edge servers at ISPs and major data centres. The data travels over only the last mile to the user on land lines. The iBeam Network enables streaming content to be served from a point close to the end user.[77] It has the following features:

- " iBeam's Intelligent Distribution Network has two elements that ensure reliable and smooth streaming experiences for end users. The first are proprietary media serving systems placed in the facilities of Internet Service Providers (ISPs). The second element is that iBeam has established direct peering relationships with the leading backbone providers for immediate, or "on-net" connectivity for their users.
- " iBeam's Intelligent Distribution Network is connected by satellite and fibre-optic cable, creating a very efficient, scalable system for serving streaming media. " [115]

9. Mirror Image Internet

" instaDelivery Internet Services is a main CDN service solution of Mirror Image Internet, which reliably stores Web objects and manages content delivery to provide a completely outsourced object server solution with high-availability, guaranteed capacity and optimized global performance. With secure, replicated object storage provided by Mirror Image's Content Access Point (CAP) network, instaDelivery Internet Services automatically distributes fresh Web content without costly origin server access and processing. The end result is guaranteed capacity, availability and

superior Internet performance, regardless of user location, traffic growth or peaks in demand. "[116] It has the following features:

- " Reduced Infrastructure Costs
- " Reliable Global Content Delivery
- " On-Demand Capacity and Storage
- " Comprehensive Control
- " Fresh, Highly Available Content
- " Greater Management Intelligence
- " Rapid, Low-Cost Implementation
- " Specialized Package Options
- " World-Class Technical Support " [116]

10. Pushcache

Pushcache.com provides CDNs services based on software - pushcache and "push done right". " The pushcache is a new communication and middleware architecture, based on web caching, capability of scalable and flexible data delivery to an unprecedented degree. "push done right" is the overall technology that is easy to use, efficient Internet data and multimedia distribution, usable by a wide variety of applications. Pushcache.com can deliver content to media and Web servers near clients by these Push Software. With Pushcache Push Software help, Unusual requests that are not cached locally are handled via sibling push caches and through archive fetching. Server logs containing user-usage information are pushed to the sites. " [117] The Pushcache Push Software has the following features:

- " Efficient content delivery
- " Localized content caching

- " Dynamic adjustment to load
- " Utilization of existing hardware
- " Scalability
- " Easy incremental growth
- " Real-time streaming support
- " Online active updates
- " User authentication
- " Automated algorithms for improved manageability of updates, fetching from sibling push caches and distributed storage based on interest. " [117]