

Content Distributed Networks

Mahoudreza Shirinsokhan

Betreuer: Ali Fessi

Seminar Innovative Internet-Technologien und Mobilkommunikation WS09/10

Lehrstuhl Netzarchitekturen und Netzdienste

Fakultät für Informatik, Technische Universität München

E-Mail: shirinso@in.tum.de

KURZFASSUNG

Um die Netzwerkstaus auf einem zentralen Server zu vermeiden, benutzt man Content Delivery Network-Systeme. Dabei können Probleme wie z.B. lange Wartezeiten oder Abstanden zwischen Server und Benutzer verringert werden.

Bei Content Delivery Network oder Content Distribution Network (CDN) handelt es sich um ein System von Rechnern, die eine Kopie von Dateien des Original-Servers beinhaltet und in verschiedenen Orten eines Netzwerks in die Welt platziert sind, um Bandbreite für den Zugriff auf die Daten von Clients im gesamten Netzwerk zu maximieren. Ein Client greift auf eine Kopie der Daten in der Nähe von einem Client an und nicht an gleichen zentralen Server, damit kann man die Bottleneck an einem Server vermeiden.

Die Dateien beinhalten Web-Objekte, Media-Dateien, Streams, Herunterladbare Objekte und andere Objekte der Internet-Lieferung (DNS, Routen und Datenbankabfragen).

Schlüsselworte

Content Distributed Networks, Content Delivery Network, Akamai, CDN

1. EINLEITUNG

Wenn man im Internet surft, Musik oder Software herunterlädt oder irgendwie im Internet nach Informationen sucht, hat man wahrscheinlich ein CDN dienst benutzt, ohne es zu wissen.

Dieser Dienst dient dazu, dass in einem Netzwerk durch CDN Netzwerk Traffic und Routing optimiert wird, in dem man den Fluss von Dateien im Netzwerk minimiert.

Dabei kann die Kapazität von Bandbreite der strategisch platzierten Server höher als die Netzwerk-Backbone-Kapazität sein. Dies hat zur Folge, dass eine größere Anzahl von gleichzeitigen Benutzern auf die Dateien zugreifen. Beispielsweise, wenn einer Netzwerk 10 Gbit/s-

Netzwerk-Backbone und 100 Gbit/s-zentrale Server-Kapazität zugeordnet ist, können nur 10 Gbit/s geliefert werden. Aber wenn 10 Server in 10 Verschiedenen Standorten platziert werden, kann die Gesamtkapazität $10 * 10$ Gbit/s sein. [1] Der Grund dafür liegt daran, dass z.B. beim Aufruf einer Seite, werden die Streaming Datei von einem Server in China und die Bilde von einem Server in Mexico geliefert, wobei der Original Server nur bei der Lieferung von HTML-Dateien belastet wurde.

Strategisch platzierten Server verringern die Anzahl der Verbindungen zum Server, öffentliche Peers, private Peers und Backbones, und führen zur Senkung der Kosten.

Es werden gleichen Daten geliefert, aber anstelle eines Zugriffs auf einem Server, werden die Dateien durch eine Umleitung vom Datenverkehr ausgelagert

2. Einblick in Content Delivery Networks

CDN-Knoten werden in der Regel an mehreren Standorten, häufig über mehrere Backbones bereitgestellt. Diese Knoten kooperieren miteinander, um Anforderungen für Content für Endbenutzer durch transparente Verlagerung von Content zur Optimierung des Übermittlungsprozesses zu erfüllen. Optimierung kann in Form der Reduzierung der Kosten für Bandbreite, Verbesserung der Endbenutzerleistung (Ladezeiten und Benutzer Erfahrungen) oder Erhöhung der globalen Verfügbarkeit von Inhalten erfolgen.

Die Anzahl der Knoten und Server, aus denen ein CDN besteht, ist von Architektur abhängig, einige erreichen Tausende von Knoten mit Zehntausenden von Servern. Andere bauen ein globales Netzwerk auf und haben eine kleine Anzahl von geographischen PoPs (Point of presence).

Anforderungen werden in der Regel algorithmisch auf Knoten umgeleitet um den Vorgang zu optimieren. Falls es sich um eine Leistungs-Optimierung handelt, können die Speicherorte, die für Serving-Inhalte am besten geeignet sind, gewählt werden. Dies kann durch Auswählen von Standorten, die die wenigsten Hops, wenige Netzwerk Sekunden entfernt vom Client oder die höchste

Verfügbarkeit ermöglicht werden und andererseits werden bei der Kostenoptimierung Speicherorte, die kostengünstigsten sind, stattdessen ausgewählt. [2]

2.1 Was ist CDN?

Content Delivery Networks (CDN), entwickelten sich erst in 1998 durch mehrere gespiegelte Webserver, die strategisch an verschiedenen Standorten platziert waren, um damit die Flash Crowds zu beheben. Eine Methode, die häufig verwendet wird, um Leistung und Skalierbarkeit zu verbessern, ist die geografisch verteilte Webserver-Einrichtung. Ein CDN ist eine Kombination aus einer Content-Delivery-Infrastruktur, einer Request-routing-Infrastruktur, einer Distribution-Infrastruktur und einer Accounting-Infrastruktur. CDNs verbessern die Netzwerkleistung durch Maximierung der Bandbreite, Verbesserung der Zugänglichkeit und Verwaltung der Richtigkeit durch Content-Replikation und bieten damit schnelle und zuverlässige Anwendungen und Dienste durch die Verteilung von Content auf Cache-Server.

können auch ein Netzwerk sein, welche alle Dateien oder nur eine bestimmte Dateityp gelagert haben



Abb. 2. Content Delivery Network (CDN) [3]

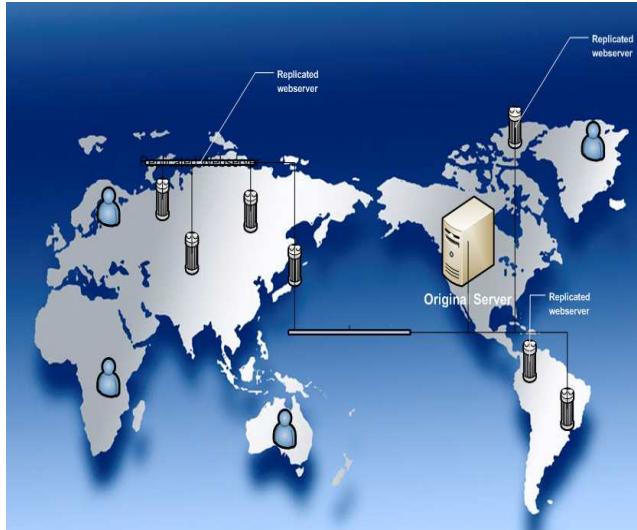


Abb. 1:Abstrakte Architektur von einem Content Delivery Network (CDN)

Abbildung 1 zeigt eine typische Content Delivery-Umgebung, in dem die replizierten Web-Server-Cluster sich am Rand des Netzwerks befinden und mit dem Endbenutzer verbunden sind. In solchen CDN-Umgebungen sind Webanforderungen von der Ursprungsservern abgerufen und an einen Benutzer geliefert. Die Dateien werden aber vom CDN-Server aufgerufen, was am nähsten liegt. Diese Randserver

CDN lagert Informationen für schnelle Bereitstellung von digitalen Inhalten, einschließlich statischen Inhalten (z. B. statische HTML-Seiten, Bilder, Dokumente, Software-Patches usw.), Stream-Medien (z. B. audio, Echtzeit-video-usw.) und unterschiedliche Content-Services (z.B. Directory Service, e-Commerce-Service, Datei-Transfer-Service usw.). Die Quellen des Inhalts sind große Unternehmen, Web-Service-Anbieter, Medienunternehmen, News Rundfunkveranstalter. Abbildung 2 zeigt die verschiedenen Inhalte/Services, die ein CDN für verschiedene Clients bedient.

Das Internet wurde nach dem End-to-End-Prinzip entwickelt. Dieses Prinzip hält das Core-Netzwerk relativ einfach und verschiebt die Intelligenz, so weit wie möglich an die Netzwerk-Endpunkte: die Hosts und Clients. Infolgedessen ist das Kern-Netzwerk spezialisiert, vereinfacht und optimiert, um nur Datenpakete zu weiterleiten.

Content Delivery Networks ergänzen das Netz durch die Verteilung des Verkehrs auf eine Vielzahl von intelligenten Anwendungen, die mit Techniken zur Optimierung der Bereitstellung von Inhalten konzipiert ist.

Web-caches speichern beliebte Inhalte auf den Servern, die die größte Nachfrage für die Inhalte angefordert haben. Diese gemeinsam genutzten Netzwerk-Appliances reduzieren der Bandbreitenanforderungen, die Belastung der Server und verbessern die Client-Reaktionszeiten für Inhalte, die im Cache gespeichert.

2.2 Grundlegende Interaktionen in einer CDN

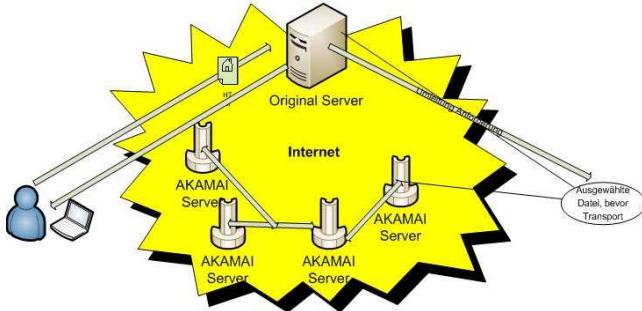


Abb. 3: Grundlegende Interaktionen in einer CDN-Umgebung

Abbildung 3 stellt die Ansicht der grundlegenden Bewegungen zwischen den Komponenten in einer Umgebung mit Content Delivery Network (CDN) dar. Hier XYZ ist der Content-Anbieter und Akamai ist die CDN, die den Inhalt der XYZ hostet.

- 1) der Client fordert durch Angabe der URL im Web-Browser Inhalte aus XYZ an. Anforderung wird direkt zum XYZ Server gesendet.
- 2) Wenn XYZ eine Anforderung empfängt, trifft es eine Entscheidung, die nur die grundlegenden Inhalte (z. B. Index-Seite der Website) bereitstellt, die aus seinen Ursprungsservern geliefert werden können.
- 3) die hohe Bandbreite dient für anspruchsvollen und häufig gestellten Inhalt (z. B. eingebettete Objekte – frische Content, Navigation Bar, Banner-Werbung).
- 4) mit dem proprietären Auswahl-Algorithmus wählt der CDN-Anbieter (in diesem Fall Akamai) den Replikatserver, der am nächsten zu dem Client liegt.
- 5) ausgewählter Replika-Server ruft die eingebetteten Objekte aus dem Ursprungsserver, liefert die Dateien an den Client und für das nächste Mal wird das Objekt gespeichert.

2.3 Wie CDNs funktionieren?

Welche Datei auf welchen Server gespeichert werden soll, wird durch das Algorithmen von CDN bestimmt. Dies kann von Client-Ort, Client-Typ(Handy, PC, Mac...) oder Server-Performance abhängig sein. Z.B. wenn jemand die Seite von Symantec von NY aufruft bekommt man einen anderen Server als wenn man Symantec-Seite von Boston oder München aufruft. Dabei bekommt man bei DNS-Aufruf verschiedene IPs und verschiedene A-Records zurück(Abb.4). Dies heißt Mapping

[NYC] host www.symantec.com

www.symantec.com	CNAME	a568.d.akamai.net
a568.d.akamai.net	A	207.40.194.46
a568.d.akamai.net	A	207.40.194.49

[Boston] host www.symantec.com

www.symantec.com	CNAME	a568.d.akamai.net
a568.d.akamai.net	A	81.23.243.152
a568.d.akamai.net	A	81.23.243.145

München Host: www.symantec.com

Additional records

n8d.akamai.net	A	72.247.38.37	3600s
n1d.akamai.net	A	72.246.55.131	2700s
n4d.akamai.net	A	72.246.55.132	2700s
n2d.akamai.net	A	74.203.241.23	3600s
n6d.akamai.net	A	72.246.55.135	1800s
n7d.akamai.net	A	72.246.55.137	2700s
n0d.akamai.net	A	72.247.38.37	1800s
n3d.akamai.net	A	72.246.55.134	1800s
n5d.akamai.net	A	72.246.55.133	3600s

Abb. 4: Akamai [4]

Es existiert zwei Typen von CDN-Network

- Network Owned CDN

Bei Network Owned CDN werden die Mapping durch anycasting bearbeitet

- Non Network CDN

Bei den CDNs, die kein Netzwerk besitzen, wird die Mapping durch DNS-Server realisiert. Hier werden die CDN-Server in Netz von anderen Anbietern platziert

Als Beispiel kann man folgenden Firmen für CDN nennen.

Akamai Technologies, Amazon CloudFront, AT&T BitGravity, CacheFly, CDNetworks (PantherExpress) Cotendo, EdgeCast Networks, GoGrid, Highwinds Network

2.4 Preise und Kosten

Da die Zahl von CDN-Anbieter seit 1998 zugenommen hat, kommt immer und wieder einen Preiskampf zwischen den Anbietern.

Da ein bestimmter Preis nur dann möglich wäre, wenn die Anforderungen genau analysiert sind, kann man hier ein paar Beispiele nennen. Bei Amazon wird für erste 50 TB

pro Monat 0.150 US Dollar pro 1 GB gerechnet, Die Preise können sich auch durch die Anzahl von Aufrufe oder Dateitypen variieren.[5]

Bei Akamai befindet sich die Preise nicht auf die Webseite, sondern soll man direkt mit einem Verkäufer Kontakt aufnehmen. Für meine Arbeit habe ich Akamai kontaktiert, damit ich einen Einblick von Kosten bei Akamai bekomme. Bei einer Webseite mit 500GB MRC kostet Ca. 400 \$ pro Monat und bei einer Webseite mit weniger als 500 GB MRC konnte ich keinen Angebot bekommen und bin ich an AKAMAI's Partner weitergeleitet. Das Angebot entspricht 200\$ pro Monat für 50 GB(4\$/GB). Bei Limelight entspricht das Angebot \$0.22/GB, wobei keine Streaming möglich ist. CacheFly dagegen bietet für \$0.50/GB auch Streaming. Aber beim Auswählen eines CDN-Anbieters soll auch auf die Geschwindigkeit, Verfügbarkeit und auch die Anzahl von Servern geachtet werden.

3. AKAMAI Technologie

Akamai ist eine non Network CDN, wer in 1998 in den USA gegründet worden ist. Der hat mehr als 6 Filialen in der Welt mit mehr als 15,500 Angestellten. Laut Akamai, besitzt die Firma mehr als 40,000 Server in mehr als 660 Städte in die Welt.

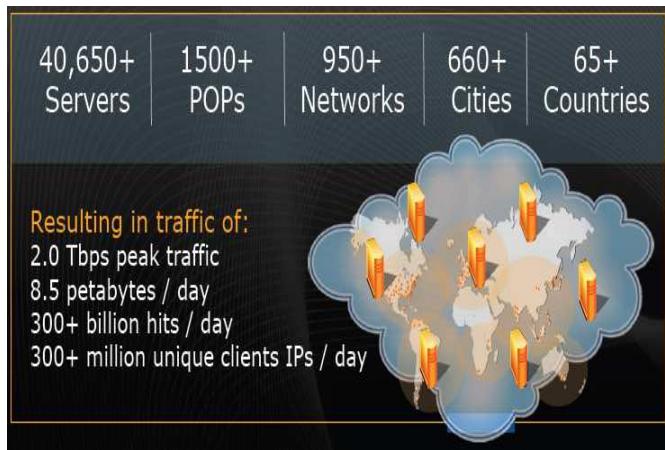


Abb. 5: Akamai [5]

Akamai Technologies entwickelte sich aus einem MIT-Forschungsbestreben, um eine Lösung für Flash Crowd zu finden. Es ist Marktführer bei der Bereitstellung von Content-Delivery-Diensten. Es besitzt mehr als 18.000 Server und mehr als 1000-Netze in 70 Ländern. [5] Das System wurde entwickelt, um Anfragen aus einer variablen Anzahl von Ursprungs-Servern am Netzwerk-Edge zu liefern. Akamai-Server können statische und dynamische Inhalte und Audio- und Video- Streaming an die Clients senden. Akamai-Infrastruktur übernimmt Flash- Crowd durch die Inanspruchnahme von mehreren Servern anstatt

der Server mit hohem Datenaufkommen. Das System leitet Client-Anfragen an den nächsten verfügbaren Server, der die angeforderten Informationen gespeichert hatte. Akamai bietet automatische Netzwerk-Kontrolle durch die Mapping-Technik, welches ein dynamisches, fehlertolerantes DNS-System verwendet. Das Mapping-System löst einen Hostnamen, basierend auf den Dienst, Standort des Nutzers und Netzwerk-Status auf. Darüber hinaus verwendet DNS für Netzwerk-Load-Balancing. Akamai Nameserver löst Hostnamen in IP-Adressen auf. Akamai Agenten kommunizieren mit bestimmten Board Routern als Peers, das Mapping-System verwendet BGP Informationen, um Netzwerk-Topologie zu bestimmen. Das Mapping-System verbindet das Akamai-Netzwerk Topologie-Informationen mit Live Netzwerk-Statistiken - wie traceroute Daten - um eine detaillierte und dynamische Sicht der Netz-Struktur und Qualitätsmaßnahmen für verschiedene Zuordnungen zu liefern.

Akamai DNS-basiertes Load-Balancing-System überwacht kontinuierlich den Stand der Leistungen und ihren Servern und Netzwerken. Um die gesamte Gesundheit des Systems zu beobachten, verwendet Akamai Agenten, die Endnutzer-Verhalten simulieren durch Abruf von Web-Objekten, Messung ihrer Ausfallraten und Download-Zeiten. Akamai verwendet diese Informationen, um Probleme zu finden und die Server-Performance zu messen. Jeder Content-Server sendet Berichte seines Ladung zu einer Überwachungs Software, die die Berichte an die lokalen DNS-Server weiterliefert. Der DNS-Server bestimmt dann, welche IP-Adressen beim Auflösen von DNS-Namen zurück gegeben werden.

3.1 Server Auswahl [5]

CDN verwendet viele Faktoren um einen Server auszuwählen

- Latency
- Throughput
- Packetloss
- CPU Load
- HD space
- Geographical place

3.2 Peer und ISP

Der erste und wichtigste Grund für Peer Verbindungen ist, dass CDN immer versucht, den am nährsten Server zu wählen, wird ein Peer Verbindung hilfreich. Ein anderer Grund dafür ist, Redundancy, dabei hat man mehr Ressourcen. Man soll auch nicht vergessen, dass durch Peering die Kosten von der Verbindung weniger wird. Dazu soll auch die Backup-Möglichkeiten, bessere Throughput, Weniger Autonomous system inzwischen, günstige Latency,

wenige Packet loss, Redundancy und Burstability auch genannt werden.

3.3 Akamai Server Type [5]

Es wurde drei Verschiedene Server-Typen durch Backtracing gefunden. Jeder Server ist für bestimmte Arbeit definiert.

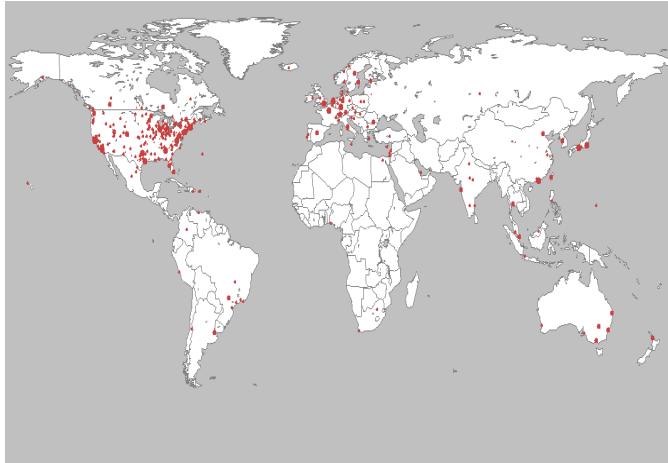


Abb. 6: Server Orts von Akamai[5]

3.4 Type A akamai.net (1964 insgesamt).

Diese Art entspricht dem konventionellen Verständnis von Akamai, also in Content Distribution Network: 1) jede DNS Resolution gibt mehr als 2 IPs zurück. Sollte die Anforderung von verschiedenen Plätzen gesendet werden, werden auch verschiedene IPs zurück gegeben worden (Abb. 4). Für Jeden Akamai Hostname gibt es mehr als 1000 IPs.

3.5 Type b Akamai.net (757 in Insgesamt)

Dieser wird nur zwecks Globale Load Balancing benutzt. Und nicht für Content Distribution. Dies wurde dadurch definiert, weil jeder Akamai-Cname von dieser Typ wird mit wenigen IPs verbunden.

3.6 Type C (539 insgesamt)

) Diese Type hat mehr als 36000 einzigartige IP Adresse und mehr als 25000 Server. Die Arbeit ist eine Mischung von Type A und Type B

3.7 Server-Ort

Durch DNS backtracing kann man die oben genannten Servers und dazugehörigen IPs bestimmen. Damit kann man festlegen dass mehr als 50% von Akamai Server sich in den USA befinden. Wobei auch eine große Anzahl von Servern sich in Europa befindet.

Country	# of IP	Percent (%)
United States	6,661	57.7
Japan	865	7.5
United Kingdom	704	6.1
Germany	545	4.7
Netherlands	384	3.3
France	364	3.2
Canada	284	2.5
Australia	164	1.4
Hong Kong SAR	158	1.4
South Korea	124	1.1
Others	1285	11.1

Abb. 7: Server Orts von Akamai[5]

ISP	# of IP	Percent (%)
Qwest	408	6.13
AT&T	157	2.36
Verizon	126	1.89
Road Runner	53	0.80
America Online	15	0.23
Charter	3	0.05
Cablevision	2	0.03
Others	5897	88.53

Abb. 8: Server Netwerks von Akamai[5]

Wie man auch in Abb.8 sieht, ist Akamai ein Non Network CDN-Anbieter, wer ihre Server im Netzwerk anderen ISPs platziert, wobei eine große Zahl von Servern in Owest Netzwerk platziert sind

4. Zusammenfassung

Auch wenn viele andere CDN diesen Dienst anbieten, wird immer die Welt der Technologie sich vergrößern. Und bei den neuen Technologien in Web und die verbesserten Geschwindigkeiten in der Webinfrastruktur vergrößern sich auch die Akamai.

Da die Anzahl von CDN Anbieter zugenommen hat, kommt oft mal zu einem Preiskampf zwischen den Anbieter. Da es in der Zukunft mehr CDN-Anbieter geben wird und auch die Hardware-Kosten durch die neue Technologie sich verringert, bin ich der Meinung dass die Preise von einem CDN-Dienst Stark abnimmt und damit besteht auch den kleinen Unternehmen die Möglichkeit, dass Sie auch CDN-Dienste benutzen

5. Literatur

[1] Internet Content Adaptation Protocol (ICAP) RFC 3507

[2] Known Content Network (CN) Request-Routing Mechanism, A. Barbir, B. Cain, R. Nair

[3] END-TO-END argument in System Design, J.H. Saltzer, D.P. Reed and D.D. Clark

[4] Peering with Content Distribution Networks von Christian Kaufmann Netnod IX Meeting 16th Feb 2009

[5] Understanding Hybrid CDN-P2P Chenq Huang, Angela Wang, Jin Li, Keith W. Ross

[6] Cloud Harmony, Bandwidth Disparity in the Cloud. January 2010

[7] Amazon Simple Storage Service
,<http://aws.amazon.com/s3/>

[8] Content Delivery Networks, Buyya, M. Pathan and A. Vakali (eds ISBN 978-3-540-77886-8) , Springer, Germany, 2008