CDN and Traffic-structure
Outline

- Basics CDN
- Traffic Analysis
Outline

- Basics CDN
  - Building Blocks
  - Services
  - Evolution
- Traffic Analysis
A Centralized Web!

- Slow
  - content must traverse multiple backbones and long distances

- Unreliable
  - delivery may be prevented by congestion or backbone peering problems

- Not scalable
  - usage limited by bandwidth available at master site

- Inferior streaming quality
  - packet loss, congestion, and narrow pipes degrade stream quality

The old centralized server oriented web did not scale.

Source: Bruce Maggs, CCGrid 2001 Keynote
Content Delivery Network (CDN)

Content Delivery Networks (CDNs) emerged as a solution to Internet service degradation
- Moving content to the “edge” of the Internet, close to end-users

Alternatives
- Increased bandwidth, Web caching, Web pre-fetching

CDN advantages
- Reduced server loads
- Distributed network traffic
- Reduced latency

First CDN emerged in 2001

Source: Content Delivery Networks: Overlay Networks for Scaling and Enhancing the Web [www.gridbus.org](http://www.gridbus.org)
Content Delivery Network (CDN)
Basic Building Blocks of a CDN.

- Infrastructure: servers that replicate the content of the server that publishes content (infrastructure, replication)
- Redirection mechanism to forward requests to servers, typically done through DNS
- Content synchronization and consistency: mechanism that guarantees that the content is fresh
- Operation support

Simple building blocks, Elementary interaction with DNS needed
Content Delivery Network (CDN)
Details of building blocks.

**CDN functional components**

Content delivery component
- Origin server and a set of edge servers (surrogates) to replicate content

Request-routing component
- Direct user requests to edge servers
- Interact with the distribution component to keep an up-to-date view of content

Content distribution component
- Moves content from the origin to edge servers and ensures consistency

Accounting component
- Maintains logs of client accesses and records usage of the servers
- Assists in traffic reporting and usage-based billing

CDNs can be realized on top of carriers IP networks: Easy traffic management by third parties

Source: Content Delivery Networks: Overlay Networks for Scaling and Enhancing the Web www.gridbus.org
Content Delivery Network (CDN)
CDN Supported Content and Services.

Static content
- Static HTML pages, images, documents, software patches

Streaming media
- Audio, real-time video

User Generated Video (UGV)

Content services
- Directory, e-commerce, file transfer services

Sources of content
- Large enterprises, Web service providers, media companies, and news broadcasters

Customers
- Media and Internet advertisement companies, data centers, ISPs, online music retailers, mobile operators, consumer electronics manufacturers, and other carrier companies

User interaction
- Cell phone, smart phone/PDA, laptop, and desktop

CDN

Contents/services

Clients

Music (MP3) / Audio

Cell Phone

Web Pages

E-docs

Desktop

Smart phone / PDA

Laptop

Streaming media
Content Delivery Network (CDN).
Content replication and synchronization.

Content outsourcing mode

- **Push**: mostly for real-time, dynamic
- **Pull**: static content

  - Non-cooperative pull-based
    - Upon cache miss, surrogate servers pull content from the origin server
    - Used by most CDN providers

  - Cooperative pull-based
    - Surrogate servers cooperate with each other to get the requested content in case of a cache miss
    - Only used by Coral CDN, making use of DHTs

CDN is highly decoupled from standard IP transport business
Content Delivery Network (CDN).
Deployment and redirection.

<table>
<thead>
<tr>
<th>CDN Infrastructure Deployment Schemes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Location of servers:</strong></td>
</tr>
<tr>
<td>▪ close to customers (akamai model): put hardware inside the ISP (leased or not), advanced virtualization, acts as a cache/proxy for the ISP so reduces ISP backbone network utilization</td>
</tr>
<tr>
<td>▪ well-connected data-centers (Google, Limelight): directly peer with large ISPs and at IXPs at strategic locations, not so ISP backbone friendly</td>
</tr>
<tr>
<td><strong>Relation to peerings:</strong></td>
</tr>
<tr>
<td>▪ close to customers does not require large peerings, only to populate the caches</td>
</tr>
<tr>
<td>▪ large peerings that require timely renegotiations and SLAs</td>
</tr>
<tr>
<td><strong>Sharing of infrastructure:</strong></td>
</tr>
<tr>
<td>▪ CDN dedicated to a single service/customer base</td>
</tr>
<tr>
<td>▪ general-purpose</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CDN Redirection Schemes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Static:</strong></td>
</tr>
<tr>
<td>▪ pre-allocation of user to a (set of) server (IPTV)</td>
</tr>
<tr>
<td><strong>Dynamic:</strong></td>
</tr>
<tr>
<td>▪ load-based: load-balancing across the CDN servers (irrespective of the network) (CDN optimization)</td>
</tr>
<tr>
<td>▪ network-based: aware of network properties such as latency and proximity-based (to the user) (service-oriented optimization)</td>
</tr>
</tbody>
</table>

Highly decoupled from standard IP transport business
Easy change of traffic structure by third parties. CDN Evolutions.

- Usually CDN provider do not have detailed information about the entry point of the customer.
- Most of the incumbent are not active in the CDN market.
- Traffic of ISP strongly depend on activities of CDN provides and OTT.

CDN: Is there an opportunity for incumbent operator?
Outline

- Basics CDN
- Traffic Analysis
  - Application MIX
  - Diversity of Traffic Sources Application – ISP collaboration
Application Mix.
Which applications – DT network.

Traffic dominated by HTTP ~ 63%

<table>
<thead>
<tr>
<th>Application</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTTP</td>
<td>63%</td>
</tr>
<tr>
<td>BitTorrent</td>
<td>9%</td>
</tr>
<tr>
<td>RTMP</td>
<td>3%</td>
</tr>
<tr>
<td>eDonkey</td>
<td>3%</td>
</tr>
<tr>
<td>NNTP</td>
<td>2%</td>
</tr>
<tr>
<td>SSL</td>
<td>2%</td>
</tr>
<tr>
<td>Shoutcast</td>
<td>1%</td>
</tr>
</tbody>
</table>
Measurement results and analysis: Application mix. Application mix & HTTP by content types and Top domains.

- HTTP dominates (57% of Bytes)
- Content types used:
  - 25% of HTTP is Flash-Video types, e.g., YouTube
  - 15% of HTTP is RAR-Archive, e.g. RapidShare
- Top Domains include:
  - One-Click Hoster
  - Video streaming
  - Software Downloads/Updates
- No significant hiding/tunneling via HTTP
- HTTP dominance due to popular high-volume content

HTTP Content-Types

- Flash-video 25%
- RAR-archive 15%
- image 11%
- video 8%
- other 23%
- unclassified 17%

TOP Domains (Aug’09)

1. 12.6%
2. 10.8%
3. 2.8%
4. 2.2%
5. 1.8%
6. 1.5%
7. 1.4%
8. 1.4%
9. 1.3%
10. 1.1%
11. 1.0%
Application mix:
Which applications – comparison to other networks?

*Based on limited dataset using payload inspection
Application mix: Residential Traces
Details (Maier 2009).

Source: Maier et al, IMC'09
Outline

- Basics CDN
- Traffic Analysis
  - Application MIX
  - Diversity of Traffic Sources Application – ISP collaboration
Internet Traffic Characteristics.
Location diversity.

Proportion of P2P traffic in Internet decreasing.
HTTP is the dominant traffic. High location diversity of important content.
Internet Traffic Characteristics.

- Top-10 Applications or Content Providers are responsible for around 50% the HTTP traffic.

- More than 60% of the HTTP traffic can be download from at least 3 different locations.
Diversity of content
Public peering points: Google.

General trend for huge OTT provider: Peering close to the customer, bypassing of tier 1 level.
Diversity of content
Public peering points: Akamai

Bandwidth in Gb/s
- 100 - 200
- 40 - 99
- 10 - 39
- 1 - 9
Traffic Engineering Opportunities.
Opportunities for redirection.

80% of HTTP traffic can be re-directed.

DNS queries of the top 10,000 hosts by volume as monitored in residential network of 20,000 DSL users.

Opportunity from for DNS usage: 80% of HTTP traffic can be re-directed.
Traffic redirection via DNS

1. Client queries DNS
2. DNS query sent to Provider DNS
3. Provider DNS queries External DNS
4. External DNS responds
5. DNS response sent to Client
Download times – Example: CDN

Zum Teil suboptimale Downloadzeiten!
Traffic engineering via server selection
Traffic engineering via server selection

Host A

Host B

Host C

Client
Traffic engineering via server selection
Server selection: How?

External DNS

Provider DNS

Internet Service Provider (ISP)

PaDIS

Client

Host

Provider-aided Distance Information System