Building next-gen CDN with extremely low power and high IO architecture

Hong Cai, PhD
CTO of Cloud Computing, ZTE TX
Agenda

• CDN Market Trends
• CDN Storage Architecture
• CDN Storage Optimization and Acceleration
## CDN Market Trends

<table>
<thead>
<tr>
<th>Rank</th>
<th>Upstream Application</th>
<th>Share</th>
<th>Downstream Application</th>
<th>Share</th>
<th>Aggregate Application</th>
<th>Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BitTorrent</td>
<td>24.53%</td>
<td>Netflix</td>
<td>34.21%</td>
<td>Netflix</td>
<td>31.09%</td>
</tr>
<tr>
<td>2</td>
<td>HTTP</td>
<td>14.27%</td>
<td>YouTube</td>
<td>13.19%</td>
<td>YouTube</td>
<td>12.28%</td>
</tr>
<tr>
<td>3</td>
<td>SSL</td>
<td>6.54%</td>
<td>HTTP</td>
<td>11.65%</td>
<td>HTTP</td>
<td>11.84%</td>
</tr>
<tr>
<td>4</td>
<td>Netflix</td>
<td>6.44%</td>
<td>iTunes</td>
<td>3.64%</td>
<td>BitTorrent</td>
<td>5.96%</td>
</tr>
<tr>
<td>5</td>
<td>YouTube</td>
<td>5.52%</td>
<td>SSL</td>
<td>3.42%</td>
<td>SSL</td>
<td>3.80%</td>
</tr>
<tr>
<td>6</td>
<td>Skype</td>
<td>2.23%</td>
<td>BitTorrent</td>
<td>3.40%</td>
<td>iTunes</td>
<td>3.33%</td>
</tr>
<tr>
<td>7</td>
<td>Facebook</td>
<td>2.17%</td>
<td>MPEG</td>
<td>2.85%</td>
<td>MPEG</td>
<td>2.62%</td>
</tr>
<tr>
<td>8</td>
<td>FaceTime</td>
<td>1.50%</td>
<td>Facebook</td>
<td>1.99%</td>
<td>Facebook</td>
<td>1.83%</td>
</tr>
<tr>
<td>9</td>
<td>Dropbox</td>
<td>1.20%</td>
<td>Amazon Video</td>
<td>1.90%</td>
<td>Amazon Video</td>
<td>1.82%</td>
</tr>
<tr>
<td>10</td>
<td>iTunes</td>
<td>1.15%</td>
<td>Hulu</td>
<td>1.74%</td>
<td>Hulu</td>
<td>1.58%</td>
</tr>
<tr>
<td></td>
<td>64.40%</td>
<td></td>
<td>76.24%</td>
<td></td>
<td></td>
<td>74.58%</td>
</tr>
</tbody>
</table>

Source: Sanvine, 2014
Challenges and Opportunities

• Challenges
  – CDN Price/Profit down

<table>
<thead>
<tr>
<th></th>
<th>Dan Rayburn (May ‘13)</th>
<th>CVN Prediction (Aug ‘13)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>- 45%</td>
<td>- 5% - 10%</td>
</tr>
<tr>
<td>2010</td>
<td>- 25%</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>- 20%</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>- 15%</td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>- 25%</td>
<td></td>
</tr>
</tbody>
</table>

• Opportunities
  – Value added process (such as injection/e2e process automation)
  – Multi-screen
IPTV CDN Architecture

IPTV CDN: it is a content distributing network, which consists of CDN Manager, one central node, Several regional central nodes and edge nodes.

- **CDN Manager**:
  - Function: CDN network manager, it will finish whole content scheduling, global load balance function.
  - Hardware statues: normal server.

- **CDN node**:
  - Function: finish VOD content storage, transmission and streaming media service;
  - Finish Live TV distributing, Live TV transcoding, FEC codec, Live TV storage and transmission functions and TSTV, TVOD, NPVR streaming media service
  - Hardware statue: Central node VS8000C/VS8000H, edge node VS3000.
Multi-S Converge CDN Architecture (IPTV node + OTT node)

Architecture:
- IPTV node is STB/RTSP independent deploying node, network architecture could be IPTV multiple levels.
- OTT node (PC/Mobile terminal Wi-Fi access) and OTT node (Mobile terminal 3G access) mainly provide HLS PC and mobile terminal service, according to project network situation, it could be set that one node is deployed in the same network plane, one node separately deployed in different network plane.)
IPTV Multi-Screen Converge CDN External Interface

- Content source
- CDN_M interface
- FTP

- IPTV /OTT service system
- CDN_S interface
- EPG
- CDN_A interface
- STB

- CDN network
High Performance, Large Storage, High Reliability Equipment

- **High Performance, Large Storage**
  - High performance: Single blade could process 4Gb I/O request.
  - Large storage: Providing PB level storage capacity.
  - High performance, low power consumption, more low carbon, green innocuity, environment protection without lead.

- **High Reliability Equipment**
  - **Key single board redundancy:**
    - Key single board main-slave back up configuration: Control single board, Live TV distributing single board main-slave back up configuration; switch board 1+1 redundancy configuration;
    - VOD single board sharing storage: storage I/O has no hot point. any one blade has fault, the other blade in node could visit whole storage content in node, VOD service has no affect.
  - **Key module redundancy:** power supply and fan are redundancy.
  - **Networking redundancy:** internal media interactive, control interactive, external service dual networking redundancy.
  - Adopting RAID1+0 dual copies storage technology, storage content is more reliable.
Sharing Storage Architecture

ZTE CDN node adopts self-made distributed file system to manage node storages resource, provides file read/write and storage content sharing function inside CDN node. Distributed file system has three parts such as FAC, FAS, FLR.

The main features of distributed file system:

- Sharing storage: each server in node could visit whole storage content in node, which evades single point of failure and bottleneck of hot films visiting.
- High reliability: Adopting excellent RAID1+0 strip storage technology in industry, guarantying storage content high reliability.
- Fully use the hard disk IO: Adopting data segment and multiple copies technology will distribute data segment in different FAS managing hare disk, and provide the external service simultaneously.
TSTV/TVOD Unified Storage Technology

- TSTV and TVOD adopt segment mode for storage, and only save for one copy, which effectively saves storage IO and storage capacity.
- Live TV transcoding node increases segment ID to channel code stream, according to present system time periodically, and add it in RTP expanding head, then, sends it to whole network. Adopting the mechanism, it could guarantee that whole networking nodes could receive Live TV code streaming with same ID.
ZTE unified CDN supports unified nodes deployment and multi-nodes distributed deployment, fully supports open internet protocols, including HTTP, HTTP Live Streaming and HTTP Progressive Download.

Note: The above CDN map deployment based on different place. It could be unified in one.
Multiple Ways to Realize CDN Storage Acceleration
Hot Content Scheduling within Node

- Inter-Node Hot Spot Schedule
  - Slide Schedule: improve responsiveness and storage utilization
  - Dynamic Priority Schedule: instead of fixing the priority at the beginning of task, this method will tune priority on the fly
  - Prediction Schedule: shorten users’ wait time by using correlation between TV series to predict next possible series

- Hot Spot Schedule within Node
  - Place hot contents in high speed media (memory and SSD) to improve overall system performance
  - Replicate hot contents crossing servers to remove bottleneck of one server

Diagram:
- HDD(SATA/SAS) Cache
- SSD Cache
- Ram Disk
- File IO Thread
- Content Schedule Policy Control
- Calculate “Hotness”

CDN Manager

Backbone CDN

PUSH

Edge CDN

PULL

Edge CDN
Small File Aggregation

Main Index

Sub-index0
- Start Index Hash
- End Index Hash
- Start Index Position
- End Index Position

Sub-index1
- ......

Sub-index0,1
- 1st URI Hash
- 1st URI Pos
- 2nd URI Hash
- 2nd URI Pos
- ......
- Last URI Hash
- Last URI Pos

Big File
- 1st URI Data
- 2nd URI Data
- 3rd URI Data
- Last URI Data

- **Improve storage utilization**
  - Merge small files to large files to improve storage space efficiency (defragment)

- **Lower file system pressure**
  - Decrease overall size and file number

- **Increase disk access speed**
  - Quickly locate and cache high frequency files to lower delay and increase hit rate

- **Improve distribution efficiency**
  - Increase speed of content delivery among nodes

- **Matching scenario**
  - Suitable for few change CDN system
Intelligent Scheduling

Provide intelligent schedule control. Improve utilization, save bandwidth and power by balancing busy/idle nodes, CPU/disk hibernation.

- Increase hit rate with intelligent schedule
- Low rate high resolution
- Compression to save bandwidth
- Hibernate idle node to save 50% power

Example: 500k concurrent user, 200 manage server, 20 CDN node:
- Save bandwidth: increase hit rate by 20%, save 500k*20%*2.5M=25G bandwidth
- Cost saving: 50% hibernate at idle, (300W*200 + 10000W*20)*50%*10 hours = 1300kWh, conservative estimation is 475k kWh saving annually
Enabling Whole Sale CDN Business Model

- **Application system injection mode**
- **CDN Web**
- **CP/SP Back-to-source mode**

**CDN Manager/GSLB**

- **Storage and bandwidth mgnt**
  - Inject and cache according to CP/content priority. Cap limit for a specific CP.
  - CDN Manager notify GSLB the priority and available bandwidth of CP/service

**Content and service schedule**

- Collect statistics of hot spot of CP and services. Control the content center to back-to-source pull down and distribute contents to service node.
- Nodes report to GSLB with performance according to CP/service
- GSLB control pre-allocated bandwidth of each CP/service

**Emergency offline**

- GSLB reject content according to content URL
- CDN delete specified content according to content URL