Netflix: Streaming Entertainment to 200 Million Members Around the World
Open Connect
Open Connect is Netflix’s content delivery network. It is global, widely distributed, efficient, and purpose-built for distributing Netflix’s content.
Open Connect delivers streaming video to over 200 million members, delivering over 100 Tb/s at peak.
The Open Connect Appliance

The OCA is the workhorse of the Open Connect network.

The OCA almost exclusively runs open-source software, including its OS (FreeBSD).
Typical Netflix OCA Workload

RAM

Data

Encrypted Data

Disks

CPU

Network Ports
Using commodity parts, we achieve 180 Gb/s serving TLS-encrypted connections with less than 50% CPU on a single 32-core 2.5-GHz CPU in 2 RU.
OCA Storage Details
Storage Technologies

- Flash
  - Moved to NVME attachment
- Spinning drives
  - Planning to move to dual actuator
Three Different OCA Flavors for Different Workloads

- All-flash appliance (180+ Gb/s)
- Large spinning disk/flash combination (up to 80 Gb/s)
- Small spinning disk/flash combination (up to 7 Gb/s)
Filesystem Technologies

- **UFS**
  - Used for content drives

- **ZFS**
  - Planning to deploy ZFS for our non-content drives
Some interesting context on data...

- Data is “disposable”
  - Multiple redundant copies throughout the CDN
- We pre-position almost all content
- Content placement is incredibly important for efficiency
  - You need the right number of copies...
  - ...in locations close enough to the members...
  - ...spread across the right servers...
  - ...and across the correct disks in each server
Position on Disk

Faster

Slower
We try to be smart about disk I/O

- Using readahead
- Keeping heavily accessed files in memory
- Being careful about mixing reads and writes
- Pacing out very disruptive operations like trims
OCA Operating System Optimizations
Driving efficiency

- Reducing memory bandwidth and/or making sure things are hot in the cache
- Using hardware offloads to reduce memory bandwidth and/or CPU usage
- Using PCIe bandwidth and/or I/O controller resources efficiently
- Enabling new platforms/designs
Reducing memory bandwidth: Async Sendfile

User Space

Webserver Process

Kernel Space

Sendfile System Call

OK, that's scheduled

Please send 16KB from FD 27

VM and I/O Systems

Storage

VM Page Cache

Socket Buffers
Reducing memory bandwidth: Async Sendfile + KTLS

User Space

Webserver Process

Kernel Space

Sendfile System Call

VM and I/O Systems

Storage

VM Page Cache

KTLS Encryption

Socket Buffers

Please send 16KB from FD 27

OK, that’s scheduled
Reducing memory bandwidth: Optimized data structures

<table>
<thead>
<tr>
<th>Flags</th>
<th>Data Ptr</th>
<th>Size</th>
<th>Offset</th>
<th>State</th>
<th>Next Ptr</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>Next Ptr</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Flags</th>
<th>Data Ptr</th>
<th>Size</th>
<th>Offset</th>
<th>State</th>
<th>Next Ptr</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Netflix: Streaming Entertainment
USENIX FAST '21
Reducing memory bandwidth: Coalesce data structures
Reducing memory bandwidth: Sorted/RSS-Assisted LRO
Reducing memory bandwidth: VM Optimizations from Upstream FreeBSD

- Per-CPU page caches
- Batched frees
Hardware offloads:
TCP Segmentation Offload (TSO)
Hardware offloads: Inline TLS Encryption
PCle Optimizations: Coalescing small memory segments

Record 1

Header | 4 KB page | Trailer

Record 2

Header | 4 KB page | Trailer

Header | 4 KB page | Trailer | Header | 4 KB page | Trailer
Enabling new platforms/designs: NUMA enablement
Typical Netflix OCA Workload

- RAM
- Disks
- Network Ports/Offloads
- CPU
- Data
- Encrypted Data
Thank you

Contact: jtl@netflix.com