What’s Behind 8K Video Streaming on Cloud Edge?

Jinghua Gao, Zhi Chang (Staff Researcher, Lenovo Research)
Shaohe Feng (Senior Cloud Engineer, Intel)
2018-11-15
Agenda

01  8K Video Background
02  Edge Optimized Architecture
03  Key Edge Technologies
04  Demo
05  Summary
1. 8K Video Background
New Buzzword

Aug 2012
ITU approved NHK’s 8K SHV standard

Jan 2013
Sharp announced first 8K TV in CES

Aug 2016
NHK kicked off the first 8K satellite broadcasts

Sep 2018
Samsung, LG, and Sony all displayed 8K TVs in IFA consumer technology show.

Mar 2018
Alibaba released 8K video cloud solution.

July & Aug 2020
Tokyo Olympics will be shot and broadcast in 8K by NHK
### 8K Video Overview

<table>
<thead>
<tr>
<th></th>
<th>HD</th>
<th>4K</th>
<th>8K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution</td>
<td>1920*1080</td>
<td>3840*2160</td>
<td>7680*4320</td>
</tr>
<tr>
<td>Audio Channel</td>
<td>5.1</td>
<td>5.1</td>
<td>22.2</td>
</tr>
<tr>
<td>Viewing Angle</td>
<td>30</td>
<td>60</td>
<td>100</td>
</tr>
<tr>
<td>Main Coding Format</td>
<td>MPEG, H.264, H.265, VP9</td>
<td>H.265, VP9</td>
<td></td>
</tr>
<tr>
<td>Network bps</td>
<td>20Mbps</td>
<td>40-50Mbps</td>
<td>120-150Mbps</td>
</tr>
</tbody>
</table>

*Currently: Higher Resolution
UHD-2 (7680 x 4320) "8k"
UHD-1 (3840 x 2160) "QuadHD"
HD (1920 x 1080) "HD"*
Usage Scenarios

- 8K offers **stronger sense of presence and realness**.
  - Commercial TVs
  - Immersive video applications, such as **panoramic video** and **virtual reality** (AR/VR)
  - Real-life applications, such as **remote healthcare** and **high-precision monitoring**

Panoramic Video | AR/VR | Remote Healthcare | High-precision monitoring
--- | --- | --- | ---
![Panoramic Video](image1.png) | ![AR/VR](image2.png) | ![Remote Healthcare](image3.png) | ![High-precision monitoring](image4.png)

Internal structures of blood vessels, Boundaries between cancer and normal tissues. Tiny sutures
Challenges

- More parallel computing in codec algorithm (8K: H.265, VP9)
- Cameras are separated and far away from the datacenter
- Need to manage devices and topology of different separated edge nodes
- Need edge computing technologies.
- Extreme high bandwidth and low latency (8K: 150Mbps)
- Add complexity to data processing of DL/ML model
- Need to manage different accelerators to empowering the DL/ML models and video codec algorithms.

Codec
- Encoding
- Decoding
- Transcoding

Storage
- ...

Analytics

Transmission

2. Edge Optimized Architecture
Edge Computing Solution Architecture
3. Key Edge Technologies
Interaction Between Datacenter and Edge

- **Data Center**
  - Monitor every edge cluster
  - Store edge cluster’s info

- **Edge Cluster**
  - Register into Data Center
  - Report its status and topology info periodically
Physical Topology Management

- Display and automatic discover network topology.
  - After enable LLDP protocol, Proton can find all network switches and servers automatically and the relationships between them.
  - If network topology is changed, Proton can detect the change within few minutes.
Physical Topology Management

- Manage physical network devices and display their current status.
  - Get detailed info about switches, such as software version, system name, interfaces status
Edge Devices Management

REST API

- Vender-based plugin
- Device lifecycle management
- Device monitoring
Intel is Unleashing Innovation in Visual Cloud

Access hardware accelerated codecs and programmable graphics on the latest Intel® processors.

Four core building blocks – Five Key Workloads – Multiple Use Cases

Building block selection & sequence are definitional.
Intel is Unleashing Innovation in Visual Cloud

**Software accelerated workloads, faster time to market**
- **Encode / Decode**
- **Inference**
- **render**
  - OpenVINO™
  - Remote Display SDK
  - Embree
  - Open SWR

**Scalable Software Supporting Industry Frameworks**

**Transformed Network**
- Workload convergence
- Cloud-ready networks
- 5G Infrastructure

**Platform portfolio – scalability for power, density, quality**
- Intel® Xeon® Processor
- Intel® Xeon-D Processor
- Intel® Visual Cloud Accelerator
- Intel® FPGA
- Intel® Optane Persistent Memory
- Intel® Gen Graphics
- Intel® 8th Gen Core™ Processor
- Radeon graphics

**Intel Architecture**
- VNFs
- NFVi
- MANO
Power 8k video processing with hardware

Xeon® Scalable Processor with Advanced Vector Extensions (AVX512)
  • Mesh Architecture delivers low latency and high bandwidth among cores, memory, and I/O controllers
  • 1.72x video stitch
  • 1.9x HEVC video encoding
  • 1.5x transcoding
  • 2.2x AI deep learning

• Higher performance
  • Bandwidth, IOPS, latency
• Higher endurance
QAT (QuickAssist) provides hardware acceleration for compute-intensive workloads
- 100 Gbs Cryptography and Public key encryption workload acceleration.
- 100 Gbs data compression
- 100 kops RSA
- 2k Decrypt

Power 8k video processing with hardware

high-performance demands of high-throughput systems
- 10 TFLOPS of floating-point performance
- Up to 28.3 Gbps Transceiver support
- Over 2.3 Tbps bandwidth for parallel memory interfaces.
- Inference, HPC
Power 8k video processing with hardware

Cascade Glacier (SmartNic) accelerates network performance
• 2x25GbE Today, 100GbE In Future
• Full Open vSwitch Acceleration On SmartNIC
• Programmability with Ease of Deployment
• Offload networking functions (NFs) from host processors
Accelerator Upstream Status

- FPGA in Cyborg  https://review.openstack.org/#/c/531129/
- QAT Crypto in Ceph https://github.com/ceph/ceph/pull/15168
- QAT Compress in Ceph https://github.com/ceph/ceph/pull/19714
- Persistent Memory in Nova https://review.openstack.org/#/c/601596/
- Persistent Memory for Read Cache in Ceph https://github.com/ceph/ceph/pull/22573
- Persistent Memory for Write Cache in Ceph https://github.com/ceph/ceph/pull/20375
- AVX512 VNNI in libvirt https://www.spinics.net/linux/fedora/libvir/msg139936.html
- QAT/GPU/FPGA in Kubernetes https://github.com/intel/intel-device-plugins-for-kubernetes
Accelerator Upstream Status

• FPGA plugin in Kubernetes
  – Support two mode: af and region
  – Multi pod can share FPGA by SR-IOV

• GPU plugin in Kubernetes
  – Multi pod can share GPU by i915 driver

• QAT plugin in Kubernetes
  – Data Plane Development Kit (DPDK) drivers must be loaded and configured
  – Intel QuickAssist Technology software for Linux must be installed and configured
  – Multi pod can share QAT by SR-IOV
Cyborg at the edge

- Align Cyborg data model with DPI before 1.13 release
- Cyborg DPI Plugin releads when DPI GA
- Consider the possibility of a CRD Acc controller
- Could be utilized by Kubeflow
4. Demo

1. Management of Edge Clusters with Multiple Cameras
2. Object Detection
Management of Edge Clusters with Multiple Cameras

• DC: Lenovo ThinkCloud 4.2 Version
  – 3 nodes, 1 controller node and 2 compute nodes.
• Edge:
  – 1 node with 1 NVIDIA Titan XP, video camera and other devices
Object Detection

• Software:
  - CUDA 9.2
  - OpenCV 3.2.0
  - FFmpeg 3.4.4
  - **Object detection model: darknet + yolov3-spp**
  - Training data: [coco.data]

• Comparison:
  • 1) Processing
    - Decode: GPU, Detection: GPU
    - Decode: CPU, Detection: GPU
    - Decode: CPU, Detection: CPU
  • 2) 8K video vs. 1080p
5. Summary
Summary

• Conclusion
  – Edge optimized architecture based on StarlingX.
  – Manage devices & network topology at the edge.
  – Manage different accelerators using cyborg.
  – Video Demo: 8K video provide much more details in analysis. & For the real-time analysis, 8K video should be accelerated in every processing procedure.

• Future Work
  – Cyborg containerized implementation.
  – Using user-space network stack to accelerate edge networking.
Q&A

• Jinghua Gao
  – Email: gaojh4@lenovo.com
  – Twitter: @Miss_Coco_Gao
  – OpenStack Cyborg Contributor
  – Lenovo research, staff researcher

• Zhi Chang
  – Email: changzhi@lenovo.com
  – Twitter: @changzhi1990
  – OpenStack Neutron Contributor
  – Lenovo research, Cloud computing researcher

• Shaohe Feng
  – Email: shaohe.feng@intel.com
  – IRC: Shaohe_feng
  – OpenStack Cyborg Contributor
  – Intel Cloud Team
thanks.
Different is better