CineGrid Networking
CG-2009

Cees de Laat

University of Amsterdam
Contents

1. Use cases CineGrid & Networks
2. Formats - Numbers - Bits
3. Global Lambda Integrated Facility
4. A LightPath
5. Transport Protocol issues
6. End System Issues
7. Network Storage
8. Q/A
CineGrid Mission

To build an interdisciplinary community that is focused on the research, development, and demonstration of networked collaborative tools to enable the production, use and exchange of very-high-quality digital media over photonic networks.

http://www.cinegrid.org/
Keio/Calit2 Collaboration: Trans-Pacific 4K Teleconference

Like High-Def? Here Comes the Next Level

By JOHN MARKOFF
Published: September 26, 2005

The New York Times

Keio University
President Anzai

UCSD
Chancellor Fox

Used 1Gbps Dedicated

Sony NTT SGI

Keio University
President Anzai

UCSD
Chancellor Fox

iGrid 2005
First Remote Interactive High Definition Video Exploration of Deep Sea Vents

Source John Delaney & Deborah Kelley, UWash
US and International OptIPortal Sites

SIO  NCMIR  USGS EDC  NCSA & TRECC
SARA  KISTI  AIST  RINCON & Nortel
TAMU  UCI  UIC  CALIT2
Why is more resolution better?
1. More Resolution Allows Closer Viewing of Larger Image
2. Closer Viewing of Larger Image Increases Viewing Angle
3. Increased Viewing Angle Produces Stronger Emotional Response

Yutaka TANAKA
SHARP CORPORATION
Advanced Image Research Laboratories
Contents

1. Use cases CineGrid & Networks
2. Formats - Numbers - Bits
3. Global Lambda Integrated Facility
4. A LightPath
5. Transport Protocol issues
6. End System Issues
7. Network Storage
8. Q/A
Formats - Numbers - Bits
## Format - Numbers - Bits (examples!)

<table>
<thead>
<tr>
<th>Format</th>
<th>X</th>
<th>Y</th>
<th>Rate /s</th>
<th>Color bits/pix</th>
<th>Frame pix</th>
<th>Frame MByte</th>
<th>Flow MByte/s</th>
<th>Stream Gbit/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>720p</td>
<td>1280</td>
<td>720</td>
<td>60</td>
<td>24</td>
<td>921.600</td>
<td>2.8</td>
<td>170</td>
<td>1.3</td>
</tr>
<tr>
<td>1080p</td>
<td>1920</td>
<td>1080</td>
<td>30</td>
<td>24</td>
<td>2.073.600</td>
<td>6.2</td>
<td>190</td>
<td>1.5</td>
</tr>
<tr>
<td>2k (24)</td>
<td>2048</td>
<td>1080</td>
<td>24</td>
<td>36</td>
<td>2.211.840</td>
<td>10</td>
<td>240</td>
<td>1.2</td>
</tr>
<tr>
<td>2k (48)</td>
<td>2048</td>
<td>1080</td>
<td>48</td>
<td>36</td>
<td>2.211.840</td>
<td>10</td>
<td>480</td>
<td>2.4</td>
</tr>
<tr>
<td>SHD</td>
<td>3840</td>
<td>2160</td>
<td>30</td>
<td>24</td>
<td>8.294.400</td>
<td>25</td>
<td>750</td>
<td>6.0</td>
</tr>
<tr>
<td>4k</td>
<td>4096</td>
<td>2160</td>
<td>24</td>
<td>36</td>
<td>8.847.360</td>
<td>~ 40</td>
<td>960</td>
<td>7.6</td>
</tr>
<tr>
<td>8k</td>
<td>7680</td>
<td>4320</td>
<td>24</td>
<td>36</td>
<td>33.177.600</td>
<td>~ 150</td>
<td>4478</td>
<td>36</td>
</tr>
</tbody>
</table>

Note: this is excluding sound!
Note: these are raw uncompressed data rates ex overhead!
Formats - Numbers - Bits

- **Formats:**
  - uncompressed unreadable (UMF) 3/4 GBytes/sec
  - compressed unreadable (jpeg2000) 300 - 700 Mbit/s
  - uncompressed readable (eg TIFF) 1.2 GB/s, 4.3 TB/h
  - compressed readable (eg DXT) 300 - 800 Mbit/s

- **Do not compress away the science!**

- **Storage**
  - Holland festival taking uncompressed about 12 TByte
Number, numbers and more numbers!

- **Digital Motion Picture for Audio Post-Production**
  - 1 TV Episode Dubbing Reference 1 GB
  - 1 Theatrical 5.1 Final Mix 8 GB
  - 1 Theatrical Feature Dubbing reference 30 GB

- **Digital Motion Picture Acquisition**
  - 6:1 up to 20:1 shooting ratios
  - 4k @ 24 FPS @ 10bit/color: ~48MB/Frame uncompressed
  - ~8TB for Finished 2 Hr Feature

- **Digital Dailies**
  - HD compressed MPEG-2 @ 25Mb/s
  - Data Size: ~22GB for 2 Hours

- **Digital Post-production and Visual Effects**
  - Terabytes, Gigabytes, Megabytes To Select Sites Depending on Project

- **Digital Motion Picture Distribution**
  - Film Printing in Regions
    - Features ~8TB
    - Trailers ~200GB
  - Digital Cinema to Theatres
    - Features ~200 - 300GB DCP
    - Trailers ~2 - 4GB DCP

- **Online Download**
  - Features ~1.3GB
  - TV Shows ~600MB
1. Use cases CineGrid & Networks
2. Formats - Numbers - Bits
3. Global Lambda Integrated Facility
4. A LightPath
5. Transport Protocol issues
6. End System Issues
7. Network Storage
8. Q/A
GLIF Mission Statement

- GLIF is a world-scale Lambda-based Laboratory for application and middleware development on emerging LambdaGrids, where applications rely on dynamically configured networks based on optical wavelengths.

- GLIF is an environment (networking infrastructure, network engineering, system integration, middleware, applications) to accomplish real work.
GLIF 2008

Visualization courtesy of Bob Patterson, NCSA
Data collection by Maxine Brown.
Calit2 is Partnering with CENIC to Connect California Industries and Researchers Into CineGrid

Partnering with SFSU’s Institute for Next Generation Internet

Calit2’s CineGrid Team is Working with Cinema Industry in LA and SF

In addition, 1Gb and 10Gb Connections to:
- Seattle then to Asia, Australia, Canada
- Chicago, Amsterdam, Europe, Russia, Asia
- Tijuana, Rosarita Beach, Ensenada

Extending SoCal OptIPuter to USC School of Cinema-Television

Digital Archive of Films

Prototype of CineGrid

Laurin Herr, Pacific Interface Project Leader

Calit2 UCI

Calit2 UCSD

SFSU

UCB

USC
CENIC Connects to 10Gb Research and Education Networks Nationwide and Worldwide
Contents

1. Use cases CineGrid & Networks
2. Formats - Numbers - Bits
3. Global Lambda Integrated Facility
4. A LightPath
5. Transport Protocol issues
6. End System Issues
7. Network Storage
8. Q/A
What is a LightPath

• A LightPath is a circuit like connection that connects end systems to each other. This uses usually the same infrastructure as the Internet, but a LightPath gets dedicated resources next to Internet.

• A LightPath can be a combination of:
  – A color in a fiber (Lambda)
  – Sonet/sdh circuit in a sonet infrastructure
  – Vlans and dedicated ports in an ethernet switch
  – Etc.

• Aim is to get predictable and knowable connection characteristics

• Let us look at examples setups used recently!
Very constant and predictable!
Network for “4K Pure Cinema” Trial

DCP is directly transferred from GDMX in LA to distribution centers in Japan via fiber network. Within Japan, DCP is distributed from the distribution centers to TOHO theaters. Key is distributed from Osaka center, based on the contract between WB Japan and TOHO cinemas.
Contents

1. Use cases CineGrid & Networks
2. Formats - Numbers - Bits
3. Global Lambda Integrated Facility
4. A LightPath
5. Transport Protocol issues
6. End System Issues
7. Network Storage
8. Q/A
Internet Transport Protocols

- **IP = Internet Protocol**
  - Connectionless packet transport service
  - Datagrams of max 64 kByte
  - Can be fragmented down the way
  - Packets can get lost, duplicated or out of order!

- **TCP/IP = Transmission Control Protocol**
  - Reliable byte-stream over potentially unreliable packet service
  - Connection oriented, exactly once and in order, end to end duplex

- **UDP = User Datagram Protocol**
  - Packet service up to 64 kByte
  - Connectionless, unidirectional, L2 switches may start flooding
  - Unreliable delivery, can get out of order, duplicated, lost
Flow control vs Congestion control

- **Flow control**
  - To prevent a fast sender overflowing a slow receiver
  - Receiver signals sender so it can adapt

- **Congestion control**
  - Traffic jams in the Internet: packets may get lost
  - For TCP protocol control loops via ack’s and ICMP packets
  - TCP is friendly protocol, can adapt but performance usually takes severe hit
  - RTT is reaction and recovery time
Windows and buffering for reliable protocols

• Round Trip Time (rtt) is time it takes to send a shortest message and get the answer back (unix tool ping)
• That is the shortest time the sender can know that traffic arrived at the other end
• Sender can only discard old data after receiving ack’s
• Lightspeed in fiber = 200000 km/s
• 100 km = 200 km round trip = 1/1000 sec = 1 ms rtt
  – Amsterdam - Geneve ≈ 20 ms
  – Amsterdam - Chicago ≈ 90 ms
  – Amsterdam - San Diego ≈ 160 ms
  – Amsterdam - Tokyo ≈ 250 ms
  – Amsterdam - Sydney ≈ 300 ms
## Buffer space

Window = RTT * BW

<table>
<thead>
<tr>
<th>RTT</th>
<th>100 Mbit/s</th>
<th>1 Gbit/s</th>
<th>10 Gbit/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12.5 kB</td>
<td>125 kB</td>
<td>1.25 MB</td>
</tr>
<tr>
<td>2</td>
<td>25 kB</td>
<td>250 kB</td>
<td>2.5 MB</td>
</tr>
<tr>
<td>5</td>
<td>62.5 kB</td>
<td>615 kB</td>
<td>6.15 MB</td>
</tr>
<tr>
<td>10</td>
<td>125 kB</td>
<td>1.25 MB</td>
<td>12.5 MB</td>
</tr>
<tr>
<td>20</td>
<td>250 kB</td>
<td>2.5 MB</td>
<td>25 MB</td>
</tr>
<tr>
<td>50</td>
<td>625 kB</td>
<td>6.25 MB</td>
<td>62.5 MB</td>
</tr>
<tr>
<td>100</td>
<td>1.25 MB</td>
<td>12.5 MB</td>
<td>125 MB</td>
</tr>
<tr>
<td>200</td>
<td>2.5 MB</td>
<td>25 MB</td>
<td>250 MB</td>
</tr>
<tr>
<td>500</td>
<td>6.25 MB</td>
<td>62.5 MB</td>
<td>625 MB</td>
</tr>
<tr>
<td>1000</td>
<td>12.5 MB</td>
<td>125 MB</td>
<td>1250 MB</td>
</tr>
</tbody>
</table>
TCP Tuning (if not auto-tuning)

- 1 Gbit/s on 160 ms RTT (= Amsterdam - San Diego) :
  - sysctl -w kern.ipc.maxsockbuf=50000000
  - sysctl -w net.inet.tcp.sendspace=21000000
  - sysctl -w net.inet.tcp.recvspace=21000000
  - sysctl -w net.inet.udp.maxdgram=57344
  - sysctl -w net.inet.udp.recvspace=74848
  - sysctl -w net.local.stream.sendspace=32768
  - sysctl -w net.local.stream.recvspace=32768
  - sysctl -w kern.ipc.somaxconn=512
  - sysctl -w net.inet.tcp.mssdflt=1460
  - sysctl -w net.inet.tcp.delayed_ack=2
  - sysctl -w net.inet.tcp.rfc1323=1
  - sysctl -w net.inet.tcp.rfc1644=1
  - sysctl -w net.inet.tcp.newreno=1
Other issues & protocols

• When using UDP, watch for bottleneck!
• About 10 other non standard protocols
• FAST TCP
  – Modified receiver algorithms
• RBUDP
  – Runs on top of UDP, simple back-off and retransmission scheme
Contents

1. Use cases CineGrid & Networks
2. Formats - Numbers - Bits
3. Global Lambda Integrated Facility
4. A LightPath
5. Transport Protocol issues
6. End System Issues
7. Network Storage
8. Q/A
End System Issues

• Ethernet card interface to computer bus system
  – PCI-X
    • 32/64 bit 66/133/266 MHZ -> about 8 Gbit/s max in 133 MHZ mode
  – PCI-Express
    • 2.5 Gbit/s per lane, 4, 8, 16 lanes

• Memory organization

• CPU cache
  – Effect when things go out of cache (small windows, etc.)

• CPU core
  – Takes 1 core to handle network (affinity may help)

• Disk raid subsystem
  – raid0 twice as fast as raid5
  – One disk does typically 40 MB/s write, 60 MB/s read
Contents

1. Use cases CineGrid & Networks
2. Formats - Numbers - Bits
3. Global Lambda Integrated Facility
4. A LightPath
5. Transport Protocol issues
6. End System Issues
7. Network Storage
8. Q/A
Amsterdam CineGrid  S/F node
“COCE”

DAS-3 @ UvA

DP AMD processor nodes
comp node

head node

bridge node

bridge node

bridge node

bridge node

bridge node

bridge node

bridge node

storage node
100 TByte

10 Gbit/s

GlimmerGlass
photonic switch

NetherLight, StarPlane
the cp testbeds and beyond

Rembrandt Cluster
total 22 TByte disk space
@ LightHouse

Opteron 64 bit nodes

head node

comp node

comp node

comp node

comp node

comp node

comp node

comp node

comp node

comp node

Option 64 bit nodes

head node

comp node

comp node

comp node

comp node

comp node

comp node

comp node

comp node

comp node

F10 L2/3 switch

streaming node
8 TByte

Node 41

suitcases & briefcases

10 Gbit/s

10 Gbit/s

suitcases & briefcases
RDF describing Infrastructure

Application: find video containing x, then trans-code to it view on Tiled Display

See Ralph Koning’s talk on tuesday
**CineGrid portal**

100 Tbyte

Cache & Store & Forward

---

CineGrid Amsterdam

Welcome to the Amsterdam CineGrid distribution node. Below are the latest additions of super-high-quality video to our node.

For more information about CineGrid and our efforts look at the about section.

**Latest Additions**

- **Wypke**
  - Available format: 4k dot (4.8 KB)
  - Duration: 1 hour and 8 minutes
  - Created: 1 week, 2 days ago
  - Author: Wypke
  - Categories:

- **Prague Train**
  - Available format: 4k dot (3.9 KB)
  - Duration: 27 hours and 46 minutes
  - Created: 1 week, 2 days ago
  - Author: CineGrid
  - Categories: delta prague train

- **VLC: Big Buck Bunny**
  - Available format: 1080p MPEG4 (1.1 GB)
  - Duration: 1 hour and 9 minutes
  - Created: 1 month, 1 week ago
  - Author: Blender Foundation
  - Categories: animation Blender bunny CGI