Real-time multicast p2p video streaming architecture based on scalable multiple descriptions

Alfio Lombardo, Diego Reforgiato, Giovanni Schembra
Università degli studi di Catania
Catania, Italy

Lorenzo Favalli, Marco Folli
Università degli studi di Pavia
Pavia, Italy
Outline

- System description
- Part 1: network construction and management
  - Tree construction algorithm
  - Temporal analysis of the protocol behaviour
- Part 2: Video coder structure
  - H.264/SVC
  - MDC
- Case study
Live broadcast platform. CIF video subsampled in 4 QCIF

Video source distributes video stream to peers in multipoint fashion distributed in different multiple descriptions trees

Video encoded with an H.264 multiple description encoder

Multipoint achieved with P2P approach configuring a multiple tree-structured overlay network

$N=85$ peers (5 classes) organized in $M$ trees. $F$ is the fan-out parameter
No peer churn but:

• Application is considered “stable”
• Focus on bandwidth fluctuations
• Churn handled in other multicast tree applications with local/centralized tree control at the cost of some additional signalling (CoopNet/SlitStream) or Zag+ (??)
Exponentially-weighted moving average (EWMA)

\[ \overline{B}_{P_i}(n) = \alpha \cdot \overline{B}_{P_i}(n-1) + (1 - \alpha) \cdot B_{P_i}(n) \]
NETWORK CONSTRUCTION AND MANAGEMENT
Tree construction algorithm (1/2)

- $N$ peers belonging to 5 classes and $M$ trees
- Each peer is internal in one tree and leaf in $M-1$ trees
- Example for $M=2$. nDST and nDRT (non-distributed structured/random tree)

**Main tree**
- Peers estimates uplink bandwidth towards all the others.
- Every 10 secs they send all values to Topology Manager
- Every 30 secs it chooses the $F$ peers with highest uplink bandwidth and connects them as children of the source
Secondary tree

- First, each peer leaf of the main tree is chosen and sorted in decreasing order of uplink bandwidth
- Then, they are connected in a top-down manner filling each level

\[ N=24, \quad M=2, \quad F=3 \]
Bandwidth generation process

- We generate uplink bandwidth behaviour capturing first and second-order statistics for a given peer.
- We used a modified version of the Switched Batch Bernoulli Process (SBPP), the most general Markov modulated process in the discrete-time domain.
- 4 states for Markov transition matrix, 1 sec slot, 60 secs mean permanence
- Standard deviation = 2.4 secs and mean value = \[((h-(C-1))\cdot 10\%+1)\cdot W_i\]
- EWMA control parameter = 0.8
Temporal analysis (1/3)

- nDST dataset with $F=2$ and $N=85$
- Generated trees have 6 levels where the last one contains 22 leaves
Temporal analysis (2/3)

Levels

Levels – Peer:12

Levels – Peer:3

Levels – Peer:6

Levels – Peer:12

Levels – Peer:3

Levels – Peer:6

NEM Summit 2011 - Turin
Temporal analysis (3/3)

Number of Children

Number of Children – Peer:12

Number of Children – Peer:3

Number of Children – Peer:6
VIDEO CODER STRUCTURE
As first interpretation, each layer can be coded independently from the others.

It is possible to exploit the redundancy by predict each layer from the lower layer.

Techniques applied:
- prediction of macroblocks
- prediction of motion vectors
- prediction of residual

Provides spatial and coarse grain scalability.
Multiple Description Coding

- From the original sequence, several subsequences are obtained in different ways
- Each independently decodable subsequence is called description
- Receiving all description, ideally recovers the sequence with original quality
- Most common multiple description algorithms
  - Spatial MDC
  - Temporal MDC
  - SNR MDC
Redundancy

- Adjacent pixels are clearly correlated
- How to reduce redundancy?
  - Differential encoding of some sub-sampled descriptions
  - Exploit some form of more complex prediction among some of the descriptions
Translated onto H.264

• Use scalability options of H.264/SVC
• Subsampling performed before the coder to be (almost) fully compatible
• H.264/SVC is a layered coder with several tools to remove redundancy among layers
  – Prediction of macroblocks using up-sampled lower resolution signals
  – Prediction of motion vectors using up-sampled lower resolution motion vectors
  – Prediction of residual signals using up-sampled residual signals of the lower resolution layer
ILPS – MDSC algorithm

- Start from the PSS-MD scheme
- Instead of independently code and transmit four descriptions, it couples two subsampled subsequences to form a description
- Use the ILPS to exploit the redundancy between the subsequences that form the description
RESULTS
Case Study

- Coder used: H.264/SVC ver. 8.1 with these parameters:
  - YUV 4:2:0 CIF sequence with 24 fps
  - ¼ pixel accuracy for motion estimation
  - Single reference frame
  - GOP size of 8
  - I frame every second
  - 16x16, 16x8, 8x16, 8x8 inter-prediction block with SAD metric
  - Context-based binary arithmetic coding
Unbalanced descriptions

<table>
<thead>
<tr>
<th>Descr, A / Descr. B</th>
<th>Nothing received</th>
<th>Base Layer ($BL_A$)</th>
<th>Full Desc. ($FD_A$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nothing received</td>
<td>0</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Base Layer ($BL_B$)</td>
<td>2</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Full Desc. ($FD_B$)</td>
<td>5</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>

Different cases of received descriptions

<table>
<thead>
<tr>
<th>($BL_A$), ($FD_A$), ($BL_B$), ($FD_B$)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>50799, 62946, 16659, 33224</td>
<td>32,3717</td>
<td>27,5162</td>
<td>30,101</td>
<td>32,4353</td>
<td>28,128</td>
<td>30,5528</td>
<td>29,3483</td>
<td>29,7342</td>
</tr>
<tr>
<td>55468, 92317, 55468, 64992</td>
<td>32,4887</td>
<td>32,1685</td>
<td>33,2221</td>
<td>33,236</td>
<td>32,3151</td>
<td>33,192</td>
<td>32,64</td>
<td>32,865</td>
</tr>
<tr>
<td>111000, 271000, 166000, 369000</td>
<td>34,437</td>
<td>34,772</td>
<td>36,3055</td>
<td>36,6678</td>
<td>37,869</td>
<td>36,7247</td>
<td>37,0749</td>
<td>38,5796</td>
</tr>
<tr>
<td>166000, 332000, 166000, 610000</td>
<td>35,3175</td>
<td>34,772</td>
<td>36,9787</td>
<td>37,6563</td>
<td>38,5855</td>
<td>37,394</td>
<td>38,0457</td>
<td>39,9398</td>
</tr>
<tr>
<td>332939, 536038, 299791, 366344</td>
<td>36,564</td>
<td>35,7214</td>
<td>38,7844</td>
<td>39,2534</td>
<td>37,7199</td>
<td>39,0156</td>
<td>38,1726</td>
<td>40,1907</td>
</tr>
<tr>
<td>443000, 629000, 295000, 370000</td>
<td>36,8566</td>
<td>36,6964</td>
<td>39,0228</td>
<td>39,5533</td>
<td>38,0545</td>
<td>39,2093</td>
<td>38,5453</td>
<td>40,7863</td>
</tr>
<tr>
<td>507000, 629000, 166000, 332000</td>
<td>37,0395</td>
<td>34,772</td>
<td>38,2234</td>
<td>39,2173</td>
<td>37,7023</td>
<td>38,3849</td>
<td>38,3094</td>
<td>40,2963</td>
</tr>
<tr>
<td>554000, 923000, 166000, 407000</td>
<td>37,0784</td>
<td>34,772</td>
<td>38,2608</td>
<td>40,4825</td>
<td>38,0735</td>
<td>38,9851</td>
<td>38,6326</td>
<td>41,6407</td>
</tr>
<tr>
<td>616000, 629000, 201000, 320000</td>
<td>37,1703</td>
<td>35,1369</td>
<td>38,7145</td>
<td>36,4008</td>
<td>37,7744</td>
<td>36,8783</td>
<td>38,4299</td>
<td>37,7201</td>
</tr>
<tr>
<td>1665489, 4440419, 1670975, 6499794</td>
<td>37,5768</td>
<td>36,7378</td>
<td>41,2275</td>
<td>43,4108</td>
<td>43,4702</td>
<td>42,3304</td>
<td>42,5185</td>
<td>57,7418</td>
</tr>
<tr>
<td>5549671, 9231988, 1665083, 4070112</td>
<td>37,6577</td>
<td>36,7378</td>
<td>41,357</td>
<td>43,7324</td>
<td>43,3287</td>
<td>42,506</td>
<td>42,5706</td>
<td>58,9646</td>
</tr>
</tbody>
</table>

Performances with unbalanced descriptions
Results (1/2)

Two children, non distributed

Two children, random
Conclusions

- Increasing the number of information received, the performances of the multiple description algorithm increase
- Best when both description are fully received
- The proposed tree construction algorithm (nDST) outperforms the random trees (nDRT)
- The fifth class shows worst performances due to the lower bandwidth available to the peers
THANKS