Multiple View Scalability of Presentations Distributed to Heterogeneous Devices

Raja S. Kushalnagar and Jehan-François Pâris
Department of Computer Science, University of Houston
Houston, TX 77204-3010
{rajah, paris}@cs.uh.edu

Abstract—**We present a novel approach of distributing lecture video by capturing multiple region of interest video streams from a high definition camera and then distributing via a peer-to-peer network to participants' viewing devices. We show this approach is scalable and supports a broader range of viewing devices for participants. Our approach is more flexible compared with the traditional solution of a traditional camera operator capturing a single high bandwidth video stream. In addition, the solution supports a more active learning and inclusive environment by enabling views of additional streams such as audio captioning, video descriptions or optical character recognition of overheads and whiteboards for sensory disabled participants using personal devices.**

Keywords- Online Learning, Mobile Devices, Peer-to-Peer Live Streaming, Multiple Video Streams, Simulation

I. INTRODUCTION

Online presentations and lectures have been traditionally captured by a single video camera manually aimed at the presentation. The video is then distributed via cable TV, videotapes or internet streaming. Although this model has worked well for simple presentations, modern presentations usually include multiple auxiliary information sources, such as whiteboards, multimedia overheads, demonstrations, and may include text captions of the audio or audio descriptions of the presentation for deaf or blind participants. Although these multiple information sources aid in learning and retention, it is difficult to receive and view on personal viewing devices all these sources and associated regions of interest simultaneously over a single video stream due bandwidth and resolution limits. Specifically, while high definition video streams can capture the multiple regions with sufficient resolution, they are difficult to stream to or read on small screen personal devices like the iPad.

The usual solution is to have a camera operator zoom in on the auxiliary source as needed, which is time and labor expensive. This camera zooming technique does not take advantage of the fact that these auxiliary sources have smaller regions of interest, and are more likely to slowly change, which is amenable to greater compression. Another weakness is that this does not support active learning, in which participants actively search and zoom into their field of view for essential features to answer specific questions as they participate in the presentation in a problem-driven, selective and active way.

In this paper, we propose a technique which combines a novel multiple video regions of interest streaming approach that extracts multiple regions of interest with a prioritized sliding window extension of BitTorrent. This technique creates a scalable and robust system that breaks a large single video stream into multiple non-overlapping smaller video streams with customized video settings. This approach enables individual peers to select which streams they will subscribe to. This combination allows each peer to fit the streams it will receive within its available bandwidth, thus maintaining a reasonable quality of service and an effective learning experience.

II. RELATED WORK

A. *Regions of Interest Capture and Viewing*

With the advent of inexpensive personal viewing devices or netbooks that can download and play back video, there is a need for distribution of scalable video feeds cheaply. For that to happen, presentation capture, broadcasting and viewing solutions should be portable, passive, require no pre or post processing, capture synchronized high resolution audio and video of the lecture automatically [1].

Automating the capture and transmission of distinct regions of interest of a presentation to participants offers several advantages over the traditional camera operator. First, separate video streams for each region of interest enable participants to select from an array of regions of interest with an appropriate resolution and bandwidth. This eliminates the need for a camera operator. Second, participants take control of their learning process by prioritizing their views of the multiple video streams to accommodate their visual learning preferences, rather than depending on the camera operator's preferences [2]. Third, grouping the regions of interest on their computer reduces their visual dispersion demands, especially for deaf participants dependent on visual translations of audio information [3]. Fourth, participants have the option to rewind real-time recording of each region of interest to review any missed information while still keeping up with the presentations in other regions of interest. Finally, additional, accessible streams can be obtained by transformation of the instructor's audio stream or overhead information to text streams that enable searching and indexing capabilities. These accessible streams promote participation inclusiveness by hearing, vision or accent impaired participants. Another possible accessible stream is to have audio descriptions of the overhead slides or other visual information, which benefits blind participants.

Video streaming requires timely data delivery guarantees, which is a challenge for current peer-to-peer systems, for they have inherent network delivery variations, varying connection lengths, churn, and peer inhomogeneity. Though current P2P solutions are most effective for static content sharing, solutions have been developed to multicast high quality real-time video streams over the Internet, by utilizing a limited data delivery time guarantee.

Given that video bandwidth is independent of the network and peer bandwidth, application level adaptation is necessary to match it with the varying network bandwidth. Moreover, modern video viewing peers vary widely in download bandwidth and video resolution viewing capacity: connection speed can vary from gigabit to 4G speeds, and video resolution can vary from high definition resolution (1920x1080) on home theater systems to standard definition (640x480) or less (320x240) on portable viewing device peers. Supporting receivers at a single video streaming rate or resolution is not appropriate, as it can either overwhelm slower receivers, or provide insufficient quality to larger resolution receivers. An elegant solution to adapting to varying device and network bandwidth demands is to divide the video stream into sub-parts that can be independently viewed, and combined to reconstitute the original. The sub-streams can be divided in terms of temporal, spatial or lossy compression in terms of signal-to-noise resolution. One of the earliest attempts at supporting receiver heterogeneity was ESM [4] which encoded video at multiple bit-rates in parallel and broadcasted them simultaneously; it also prioritized audio over video streams, and prioritized lower quality video over higher quality video. However, this approach incurs a lot of overhead both in terms of bandwidth and stream processing. A subsequent streaming approach attempted to mitigate the bandwidth overhead by means of scalable video coding (SVC) [5]. In this approach, a cumulative layered coder re-codes the video stream into multiple layers, by scaling a video attribute such frame rate, size, or quality. A receiver, depending on its capability, can subscribe to the base layer only with the basic playback quality, or subscribe to additional layers that progressively refine the reconstruction quality. Another approach utilizes Multiple Descriptive coding (MDC), in which a multiple description coder generates multiple streams, also referred as descriptions, from the source. Any subset of the descriptions, including each single one, can be used to reconstruct the video. A simple implementation of multiple description coding can be achieved by splitting even and odd numbered frames. In addition to temporal interleaving, more recent implementations have implemented scalar quantization. The descriptions are then distributed over multiple paths to enhance robustness and to accommodate user heterogeneity [6]. However, both approaches incur some bandwidth penalty. SVC encoding requires extra bits for synchronizing layers, and MDC also requires extra bits for each description to carry sufficient information about the original video. For both approaches, these extra bits reduce compression efficiency. Finally, both approaches also require extra processing power for receivers to assemble and decode the video stream [7], but as mobile devices gain more processing power, this issue becomes less important. All these algorithms utilize the fact that video is loss tolerant by incorporating looser statistical and perceptual guarantees. These media coding strategies enable transparent scaling to variable network and device bandwidths and resolutions.

B. Peer-to-Peer Solutions

Peer-to-peer solutions offload distribution bandwidth from the servers to participating clients, which enables available bandwidth to viewers to grow proportionally with the number of clients. Therefore, server capacity to remains constant even as the number of clients grows.

Current P2P streaming approaches are grouped into (1) tree-based approaches that push content over multiple trees, (2) mesh-based approaches that pull content over a randomly connected mesh, or (3) tree/mesh hybrids.

Tree-based approaches connect peers top-down via a rigid overlay such that each peer forwards incoming data packets to all children peers. As the tree depth is logarithmic to total number of peers, information is rapidly transmitted to all network members with little overhead, but they poorly utilize peers on tree edges and leaves. Newer approaches ameliorate this poor utilization by transmitting multiple descriptions over multiple trees, such as SplitStream [8]. However, these approaches still handle poorly low bandwidth connections, which propagate the bottleneck downstream [9].

Unstructured meshes connect peers bottom-up, via a swarm, in which each peer maintains a list of its neighbors, which improves churn response and reduces overhead of constructing and maintaining trees. Due to this flexibility, the mesh approach dominates peer-to-peer non-real-time content sharing, such as BitTorrent [10]. But due to lack of global structure, peers have to explicitly request needed packets, called pulling, which increases overhead and delays packet delivery, which makes real-time content delivery difficult. BiToS [11] modifies BitTorrent's piece selection policy of rarest-first piece selection with a need for real-time delivery. However, BiToS does not include any backing servers to guarantee real-time delivery. As a result, the measured results for all variants indicate that at least 5% of pieces are not received in time, degrading quality of service.

Tree/mesh hybrids attempt to keep advantages and avoid disadvantages of both approaches, and several different approaches have been explored. For example, CoolStreaming [12] adopts a push-based mechanism over a mesh, whereas ChunkySpread [13] splits a stream into distinct slices and transmits over separate trees.

A comparison of tree and mesh approaches [9] showed that meshes were more stable to churn and better utilized low bandwidth peers. Given the modern trend of the popularity of multimedia capable mobile devices such as smart phones, netbooks, iPads all with varying bandwidth capabilities depending on whether they are connected over phone data network or local area wire/wireless network.

Figure 1. HD view of the classroom.

Figure 2. SD region of interest of the overhead, whiteboard, and real-time transcript respectively.

III. ANALYSIS

A. *Multiple Video Regions of Interest Streaming (MVR)*

In addition to the advantages in terms of affordable off-the-shelf technology and efficient use of perceptual and learning responses in presentation capture, the Multiple Video Regions of Interest (MVR) encoding approach can be regarded as a content-aware video encoding and delivery technique that reduces bandwidth demand by capturing and delivering multiple non-overlapping small targeted video streams (see Figure 2), as opposed to a single large video stream (see Figure 1). This approach is possible and effective because presentations have non-overlapping regions of interest with different sizes and compression needs. This approach also benefits from the fact that video streaming is loss-tolerant, unlike file sharing.

Survey responses were collected from two student viewers in a computer science class lecture and two other student viewers in an intellectual property law lecture at the University of Houston. The responses indicated that these participants were equally satisfied with viewing three standard definition resolution video of regions of interest focused on the whiteboard, overhead slides and real-time transcript respectively, as compared with viewing a single high definition resolution video that captured all three with equal clarity. Since the contents of the overhead, white board and real-time transcripts tend to change slowly and be more static, higher compression settings were used. This enables significant streaming bandwidth savings: the single high quality H.264 encoded HD stream bandwidth was 8 Mbps, while the total stream bandwidth for all three highly compressed H.264 encoded SD streams was 1.2 Mbps, a ratio of approximately 7:1.

```
for each stream in streams
    if peer.bw ≥ stream.bw
        peer.get(stream)
        peer.bw -= stream.bw
    end if
end for each
```

Figure 3. The stream prioritization policy.

The Multiple Video Regions of Interest approach has the property of being tolerant and scalable, for as more streams are received, video coverage improves. This approach is largely orthogonal to, and can be combined with another content-aware delivery approach called Multiple Description Coding (MDC). In MDC, a single stream is encoded into multiple unique descriptions or streams that can be independently displayed. These descriptions are usually generated by a global video division method, such as separating by time frames, i.e., interleaving, or by spatial frequency distribution. Like MVR, MDC is also tolerant and scalable, for as more descriptions are received, video quality improves. MVR guarantees high fidelity, both spatial and temporal, of critical sections of a presentation's view, whereas MDC guarantees minimum quality of the overall view. Therefore, combining MVR and MDC guarantees a minimum quality of critical regions of interest, while providing tolerance and scalability over networks with low bandwidth connections. This is possible due to the fact that viewers can choose the number of descriptions and regions of interest proportional to their inbound bandwidth. The process of splitting into multiple video streams reduces playback delay as follows:

1. By decreasing the minimum bandwidth required to display the video stream, we proportionally decrease the playback buffer size and hence playback delay.
2. Also, by increasing the number of seeds, we decrease the peer playback delay by a factor of log N, which is proportional to the diameter of the swarm.

B. *Sliding Window*

We extend the sliding window approach proposed by Shah and Pâris [14] in which each peer keeps a sliding window array of the N streams. Peers only request chunks that are within a sliding window containing the next w chunks to be consumed. As a result, peers never download chunks that arrive after their playback times, because they would be useless. To keep the window moving according to the playback rate, we pick the size w of the sliding window, expressed in the number of chunks, per the relation:

$$w = db_v/c \qquad (1)$$

where d is the playback delay, b_v is the video consumption rate, and c is the chunk size.

IV. EXPERIMENTAL SETUP

We encoded video capture files of a presentation with multiple video regions of interest (MVR), and multiple description codes (MDC). We tested the transmission and reception of these capture files by peers that utilized prioritized stream selection in the GPS [15] simulator.

We use GPS' model, which considers network transmission and queuing delays, but neglects network propagation delays and TCP connection times, since the multimedia streaming time is dominated by the chunk exchange traffic. This model also assumes idealized TCP performance, and assumes connections traversing a link share bandwidth equally.

Each multimedia streaming session consists of N sources streaming distinct video streams to all peers, with N being the number of video capture devices. The single view video stream file was a 50 minute long lecture obtained from a single high definition camera recording at the standard HDTV resolution of 1920x1080 pixels to a raw AVI format so that no frames were lost. The multiple view video stream files were obtained from three regions of interest files at the standard SD resolution of 640x480 pixels using raw AVI so that no frames were lost. The processed single view high definition video capture file size was 3GB and the processed standard definition capture file sizes were 150 MB each. Table 1 summarizes the settings of the simulator:

Table 1. Simulation Parameters.

Chunk size	256 KB
Maximum concurrent upload transfers	4
Rechoking interval	5 s
Optimistic unchoking interval	15 s
Number of random peers returned by tracker	50
Number of neighbors of each peer	10
Playback delay	60 s

Our simulation assumes a well-behaved BitTorrent swarm in which each peer uploads until the complete video stream is received, and then leaves the swarm. At the start of each simulation run, we created a 90 peer swarm with three equal sets of peers. The first set consists of peers with 100 Mbps upload and download links, the second set consists of peers with 10 Mbps upload and download links, and the third set consists of peers with 1 Mbps upload and download links. The peer upload links were chosen so as to represent modern device bandwidths. The first set represents powerful multimedia capable computers with big displays, large bandwidth and high video processing capacity; the second set represents netbooks or iPad devices that have mid-sized displays, wi-fi connections and medium processing capacity, and the third set represents mobile peers with small displays, bandwidth and CPU power. We examined performance results for multiple video streams and piece request policy algorithms.

Like previous simulation studies we assumed that bandwidth bottlenecks only occurred at the edge and did not model shared bottleneck links in the interior of the system. We believe alternate techniques utilizing the physical topology to improve the system performance are complementary to this work.

A. *MDC Video Processing*

The single high definition video presentation was encoded with the AVCHD 1080i format, a H.264/AVC profile at a bit rate of 8 Mbps. The three multiple description were obtained by splitting into three equal resolution and bandwidth descriptions by extracting every 3rd frame and re-encoding in the same format, which yields three files with a bit rate of 2.67 Mbps. The video server streams each of the three multiple description code files such that a properly equipped peer can display the video as long as it received at least one of the descriptions.

B. *MVR Video Processing*

Three standard definition multiple video regions of interest capture files were manually extracted from the single high definition video by cropping fixed pre-defined areas. This was possible because the locations of the whiteboard, power point display and captions did not change over the duration of the presentation. We encoded the three video capture files using appropriate high compression settings to get a bit rate of 0.3, 0.4 and 0.5 Mbps for regions of interest centered around the white board, power point and caption displays, for a total bit rate of 1.2 Mbps at the standard definition resolution of 640x480 pixels using the settings for Constrained Baseline Profile (CBP) H.264/AVC.

C. *Presentation Streams*

Two streaming presentation scenarios were tested: streaming of a single high definition video clip at 8 Mbps from a single server with 100 Mbps upload capacity. The second scenario streamed multiple standard definition video streams of the selected regions of interest. Peers were assigned symmetric bandwidth download capacities of 100, 10 or 1 Mbps, which is roughly matched by upload/download bandwidths over Ethernet, Wireless-G or Sprint's 4G [16] networks respectively. The total link capacity, i.e., upload plus download capacity of the peers in the simulator were 200, 20 and 2 Mbps respectively.

V. RESULTS

When the presentation was captured and encoded using the traditional approach of a single high definition stream, only 200 Mbps link capacity peers consistently timely received all chunks of the stream; 20 Mbps link capacity peers had unacceptable reception in terms of timely chunk rate arrival, and 2 Mbps link capacity peers received nothing at all, as shown in Figure 4.

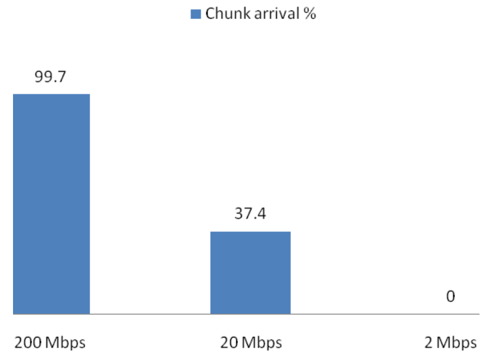

Figure 4. Average on-time chunk percentage arrival rates at peers for a presentation captured and encoded using HD.

With three multiple video regions of interest streams, 200 and 20 Mbps peers received all streams, but 2 Mbps peers only received a single region stream, as shown in Figure 5.

When the presentation was captured and encoded using multiple video regions of interest (MVR) and multiple description codes (MDC), all peers could receive at least one sub-stream. Specifically 200 Mbps peers consistently received all descriptions, while 20 Mbps peers received all descriptions and 2 Mbps peers consistently received one or more description of one or more regions of interest.

With multiple description codes streams, 200 Mbps peers received all descriptions, while 20 Mbps peers could not get all descriptions and 2 Mbps peers were unable to receive any stream, as shown in Figure 6.

The results shown in Figure 7 indicate that combining both the MDC and MVR approaches allows even the smallest bandwidth peers (2 Mbps) to consistently receive and display at least some combination of regions of interest and descriptions of the presentation video, here two regions and two descriptions (2R, 2D) or less. By combining the tit-for-tat BitTorrent property with sub-stream selection property, peers pair up with peers of similar bandwidth and utilize available sub-streams generated by the MDC and MVR encoding schemes. This enables peers to transparently segregate themselves and request sub-streams with total bandwidth of less than their network capacity.

A. *Findings*

The first significant bandwidth savings was obtained by substituting multiple standard definition streams in place of the single high definition single stream: while the single H.264 HD stream bandwidth was 8 Mbps, the total stream bandwidth for all three H.264 VGA streams was 1.1 Mbps, a ratio of approximately 7:1.

The second significant bandwidth savings was achieved by implementing multiple coding descriptions using time division multiplexing, which reduces the peer's minimum bandwidth requirement correspondingly.

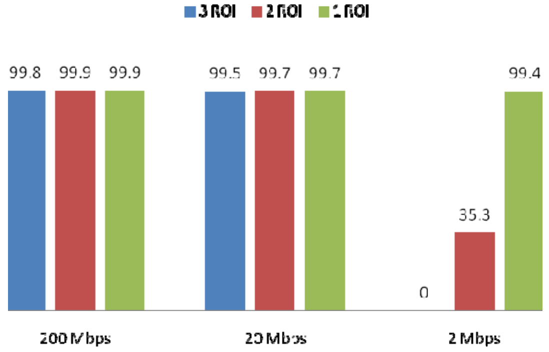

Figure 5. Average on-time chunk percentage arrival rates at peers for a presentation captured and encoded using MVR.

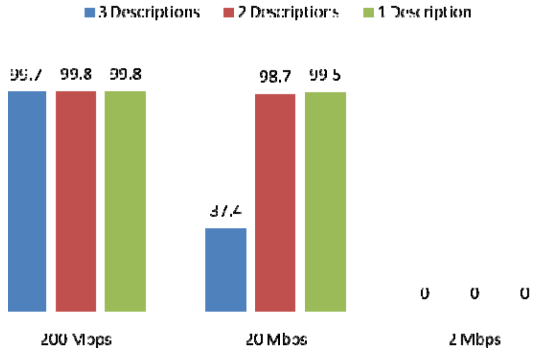

Figure 6. Average on-time chunk percentage arrival rates at peers for a presentation captured and encoded using MDC.

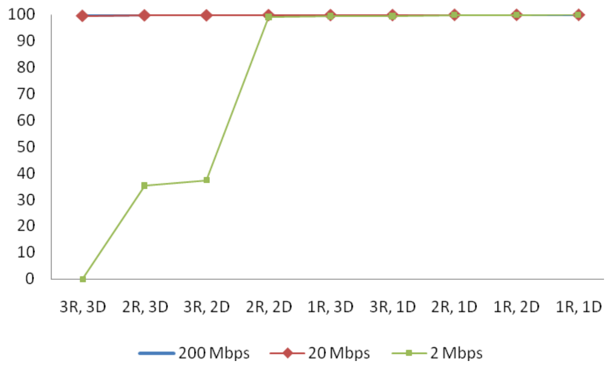

Figure 7. Average on-time chunk arrival rates using both MVR and MDC.

The bandwidth savings from these two approaches is gained mainly because modern presentations contain many non-interesting regions between the regions of interest, which are usually the overhead slides, the white board, or other physical demonstrations. Moreover, each of these regions of interest may have different sizes, resolutions and angles, which make it difficult for a single global video stream to find an optimal compression tradeoff for each of the various contents in the regions of interest within it. The approach of extracting individual regions of interest by either manual or automatic means allows content aware processing of each stream's contents so as to be able to separately applying optimal compression ratios to each of these regions of interest. Another benefit in bit rate and bandwidth savings is obtained by selecting interesting regions of interest and discarding non-interesting regions of interest. This approach reduces file size and bandwidth demand by half or more. These two approaches yield close to an order of magnitude improvement in the substantial savings in bit rate and bandwidth usage. Combining these two approaches enable a presentation to be streamed effectively to a heterogeneous network consisting of a wide variety of end-consumer peers with varying bandwidth and resolutions. Modern peers in such a network can range from home theater systems with high speed connections, laptops with wireless connections, and smart phones with mobile data connections.

VI. CONCLUSIONS

We have proposed a Multiple Video Regions of Interest streaming approach for capturing and streaming online presentations that is less expensive and is more inclusive, scalable, flexible and easier to deploy than traditional video streaming, especially for mobile peers. We analyzed issues in extending streaming to a range of consumer peers and proposed solutions: (1) Multi-video streaming with Multiple Description Encoding, and (2) stream prioritization policy.

This study proposes the MVR idea and also shows clear evidence that MVR yields bandwidth savings of approximately 7:1 in a real classroom study. This study also provides additional support for further bandwidth savings of 20:1 when MVR is combined with MDC. This magnitude of scaling enables us to provide meaningful quality of service for end-users who are viewing presentations with multiple regions of interest, even if they have limited screen resolution and bandwidth like smart phones, portable wireless devices and notebooks.

Our results show that incorporating multiple description codes and multiple video regions of interest at the video capture and encoding stage enables streaming BitTorrent to transparently support a wider range of heterogeneous bandwidth peers, when combined with the incorporation of a stream prioritization and selection policy. This stream prioritization and selection policy adds to peers the capability of prioritizing and selecting one or more multiple sub-streams from the video encoding using multiple description codes and/or multiple video regions of interest. The combination of a stream prioritization strategy with the standard tit-for-tat BitTorrent strategy enables peers to segregate with other peers of similar bandwidth. Therefore, each segregated group of peers is able to effectively select MVR and MDC encoded sub-streams that are within their upload and download bandwidth limits.

REFERENCES

[1] C. Zhang, *et al.*, "An automated end-to-end lecture capture and broadcasting system," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP),* vol. 4, p. 6, 2008.

[2] R. S. Kushalnagar and J. F. Pâris, "Scalability of Prioritized Multiple Video Perspectives in BitTorrent," Submitted to *IEEE International Conference on Computer Communications and Network (ICCCN) 2010*, Zurich, Switzerland, 2010.

[3] A. C. Cavender, *et al.*, "ClassInFocus: Enabling Improved Visual Attention Strategies for Deaf and Hard of Hearing Students," in *International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2009)*, Pittsburgh, PA, USA, 2009, pp. 67-74.

[4] Y. Liu, *et al.*, "A survey on peer-to-peer video streaming systems," *Peer-to-Peer Networking and Applications,* vol. 1, pp. 18-28, 2008.

[5] H. Schwarz, *et al.*, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 17, pp. 1103-1120, 2007.

[6] S. S. Panwar and K. W. Ross, "A Peer-to-Peer Video-on-Demand System Using Multiple Description Coding and Server Diversity," in *International Conference on Image Processing, 2004 (ICIP '04)*, Singapore, 2004, pp. 1759-1762.

[7] Y. Shen, *et al.*, "Peer-driven video streaming: Multiple descriptions versus layering," in *IEEE International Conference on Multimedia and Expo*, Amsterdam, The Netherlands, 2005.

[8] M. Castro, *et al.*, "SplitStream: high-bandwidth multicast in cooperative environments," in *19th ACM Symposium on Operating Systems Principles*, Bolton Landing, NY, USA, 2003.

[9] M. Nazanin and R. Reza, "Understanding mesh-based peer-to-peer streaming," in *2006 International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV 2006)*, Newport, Rhode Island, 2006.

[10] B. Cohen, "Incentives build robustness in BitTorrent," in *First Workshop on Economics of Peer-to-Peer Systems*, Berkeley, CA, 2003.

[11] A. Vlavianos, *et al.*, "BiToS: Enhancing BitTorrent for Supporting Streaming Applications," in *25th IEEE International Conference on Computer Communications*, 2006, pp. 1-6.

[12] Z. Xinyan, *et al.*, "CoolStreaming/DONet: a data-driven overlay network for peer-to-peer live media streaming," in *24th IEEE International Conference*

on Computer Communications (INFOCOM 2005), Miami, FL, 2005, pp. 2102-2111 vol. 3.

[13] V. Venkataraman*, et al.*, "Chunkyspread: Multi-tree unstructured peer-to-peer multicast," in *14th IEEE International Conference on Network Protocols (ICNP '06)*, Santa Barbara, CA, 2006, pp. 2-11.

[14] P. Shah and J. F. Pâris, "Peer-to-Peer Multimedia Streaming Using BitTorrent," in *26th International Performance of Computers and Communications Conference (IPCCC 2007)*, New Orleans, LA, USA, 2007, pp. 340-347.

[15] W. Yang and N. Abu-Ghazaleh, "GPS: a general peer-to-peer simulator and its use for modeling BitTorrent," in *13th IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS 2005)*, Atlanta, GA, 2005, pp. 425-432.

[16] (2009, March 1). *Sprint 4G Benefits and the WiMAX Experience*. Available: http://developer.sprint.com/site/global/home/4g/wimax_experience/wimax_experience.jsp