Diagnosing Network-wide P2P Live Streaming Inefficiencies

Chuan Wu
Department of Computer Science
The University of Hong Kong
cwu@cs.hku.hk

Baochun Li
Dept. of Electrical and Computer Engineering
University of Toronto
bli@eecg.toronto.edu

Shuqiao Zhao
Multimedia Development Group
UUSee, Inc.
zhaosq@uusee.com

Abstract—Large-scale live peer-to-peer (P2P) streaming applications have been successfully deployed in today's Internet. While they can accommodate millions of users simultaneously with hundreds of channels of programming, there still commonly exist channels and times where and when the streaming quality is unsatisfactory. In this paper, based on more than two terabytes and one year worth of live traces from UUSee, a large-scale commercial P2P live streaming system, we show an in-depth *network-wide* diagnosis of streaming inefficiencies, commonly present in mesh-based P2P streaming systems. We first identify an evolutionary pattern of low streaming quality in the system and the distribution of streaming inefficiencies across various streaming channels. We then carry out an extensive investigation to explore the causes to such streaming inefficiencies over different times and across different channels at specific times. The original discoveries we have brought forward include the two-sided effects of peer population on the streaming quality in a channel, the significant impact of inter-peer bandwidth bottlenecks at peak times, and the inefficient utilization of server capacities across concurrent channels. We conclude with a number of suggestions to improve real-world large-scale P2P streaming.

I. INTRODUCTION

Large-scale live peer-to-peer (P2P) streaming systems have been successfully deployed in the Internet, delivering channels of live multimedia content to millions of users [1], [2]. Practical experiences with these streaming systems, however, have shown that they may not be able to provide a satisfactory viewing experience to all the users at all times. Since the streaming quality experienced by the peers is of pivotal importance to the success of a P2P streaming system, it is critical to study existing commercial systems to look for any possible streaming *inefficiencies*, *i.e.,* low streaming qualities experienced by participating users. If such inefficiencies exist, it is also important to find reasons that may have caused them.

To this end, we have collected more than two terabytes and one year worth of live traces from a large-scale commercial P2P live streaming system, UUSee [2], one of the top three commercial P2P streaming systems in China, along with PPLive and PPStream. This paper presents our in-depth investigation of the network-wide streaming inefficiencies in UUSee, that may adversely effect user experiences.

As the first highlight, we have identified an evolutionary pattern of low streaming qualities in the system, and the distribution of streaming inefficiencies across various streaming channels. We then conduct an extensive study of the causes to such inefficiencies, by investigating the impact of factors such as the number of peers, peer upload bandwidth contribution, inter-peer bandwidth availability, server bandwidth consumption, the level of reciprocity among peers, and many more. In particular, we explore the correlation between the evolutionary pattern of streaming inefficiencies and the variation of various factors over time, and also zoom into snapshots of the entire system to investigate causes to the low streaming qualities experienced in specific channels at each specific time.

The original discoveries brought forward in our study include: *First,* the population of peers has two-sided effects on the streaming quality in a streaming channel: while peers in a larger channel typically enjoy a better streaming quality, the streaming quality degrades as the peer number increases in the channel during daily peak hours. *Second,* server capacity still plays an indispensable role in P2P streaming over today's Internet, and the inefficient supply of server capacity has largely led to the low streaming quality at specific times. *Third*, smaller channels tend to obtain less server capacity per peer than large channels, which contributes significantly to their low streaming qualities. *Fourth,* inter-peer bandwidth availability represents a more significant bottleneck than the upload bandwidth at the peers during daily peak hours. *Finally,* increasing peer indegree in the topology does not help enhancing the peer streaming quality. Based on these insights, we conclude with a number of suggestions on improving the design of real-world large-scale P2P streaming systems.

The remainder of this paper is organized as follows. In Sec. II, we discuss trace collection methodologies and basic statistics of UUSee. In Sec. III, we investigate the distribution of streaming inefficiency across different channels and their evolutionary pattern over time. We then search for the causes to the temporal pattern of streaming inefficiencies in Sec. IV, and diagnose the distribution of streaming inefficiencies across channels in Sec. V. We discuss related work in Sec. VI, and summarize and discuss our discoveries in Sec. VII.

II. BACKGROUND

A. UUSee P2P streaming solution and trace collection

Starting September 2006, we have continuously monitored the performance of UUSee, a P2P streaming platform with legal contractual rights with mainstream content providers in China. The UUSee streaming framework consists of over 100 streaming servers that simultaneously broadcast about 800 live streaming channels (mostly encoded to streams around 500 Kbps) to hundreds of thousands of peers.

Similar to all state-of-the-art P2P streaming applications, UUSee utilizes a mesh-pull based protocol, that allows peers in each channel to serve other peers ("partners") by exchanging media blocks in their playback buffers. The playback buffer at each UUSee peer has a buffer size of 500 media blocks, where each block represents $1/3$ second of media playback. A new peer obtains a number of partners from one of the tracking servers, exchanges buffer availability bitmaps (*i.e.*, buffer maps) with them, and retrieves new blocks from them, prioritizing blocks based on their playback deadlines. The number of consecutive blocks received and cached in the playback buffer, starting from the current playback time, is used in UUSee protocols to represent the current streaming quality of each peer, referred to as the *buffering level*.

To inspect the run-time behavior of UUSee streaming, we have implemented detailed measurement and reporting capabilities within its P2P client software. Each peer collects a set of its vital statistics, and reports to dedicated trace servers every 5 minutes. The statistics include its IP address, the channel it is watching, its current buffer map, the buffering level, its instantaneous aggregate sending and receiving throughput, as well as a list of all its partners, with their corresponding IP addresses, and current sending/receiving bandwidth to/from each partner. Each streaming server in UUSee utilizes a similar P2P protocol as deployed on regular peers, is routinely selected to serve the peers, and reports its related statistics periodically as well. Details on the UUSee protocols and measurement methodologies can be found in our technical report [3].

B. Trace summary

From September 2006 to September 2007, we have collected more than two terabytes of traces, representing time-continuous snapshots of UUSee streaming network every five minutes. At any time, there are more than $100,000$ concurrent peers and over 1 million simultaneous P2P flows in the system. The UUSee peers span most of the major and regional ISPs in China, and more than 40 countries across the world.

With respect to the distribution of concurrent peer population across various streaming channels, our measurement has revealed a heavy-tailed distribution: There exist a small number of most popular channels ($\approx 2\%$ out of 795) with an average peer population of more than 5000, a small percentage of channels ($\approx 12\%$) have fewer than 100 peers each, and the majority of channels accommodate a per-channel peer population in the range of 500 to 3000 ($\approx 51\%$). Interested readers are referred to our technical report [3] for the detailed peer distributions across channels and regions.

III. DISTRIBUTION OF STREAMING INEFFICIENCY

Before starting our investigation of streaming inefficiencies in UUSee, we first define how we evaluate the streaming quality of a channel. The streaming quality in a channel at each time is evaluated as the percentage of high-quality peers in the channel, where a high-quality peer has a buffering level of more than 80% of the total size of its playback buffer. The criterion of the buffering level (i.e., the number of consecutive

blocks received and cached in the current playback buffer of a peer) has been extensively used in UUSee protocols to evaluate the current streaming quality of a peer; and the 80% buffering level benchmark has empirically been shown to be effective in reflecting the playback continuity of a peer in the following few minutes. Accordingly, we also use the peer buffering level as our basic streaming quality criterion.

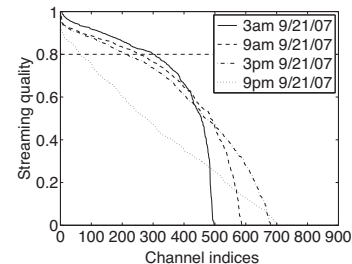### A. Distribution across different channels

Fig. 1. Distribution of streaming quality across all the channels in UUSee.

Fig. 1 plots the distribution of streaming quality across all the channels at four representative snapshots of the system on September 21, 2007 (GMT+8). Each curve is plotted in a descending order of the channels' streaming quality. The indices on the x-axis may not correspond to the same channels for different curves. We observe that, while $50 - 60\%$ of the channels have a streaming quality higher than 0.8 (*i.e.*, 80% of the peers in the channel enjoy smooth playback) at earlier times of a day, the percentage decreases towards later times of the day, and it is significantly lower in the evening (667 out of 795 channels at 9pm have a streaming quality lower than 0.8). These observations exhibit the existence of streaming inefficiency (low streaming quality) in channels at each time, which becomes more evident at later times of a day.

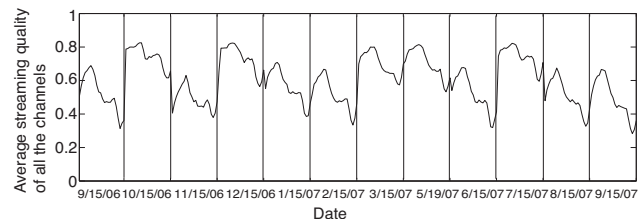### B. Evolutionary pattern over the trace period

Fig. 2. Evolution of the average streaming quality of all the channels.

The above observations reveal a time of the day effect on the streaming quality in the system. To inspect any evolutionary pattern of streaming quality, we plot in Fig. 2 the evolution of the average streaming quality over all the channels on a representative date in each month, over the 12 months of trace span (Note that April 2007 is skipped due to lack of traces in the month caused by the trace server upgrade).

We can clearly observe a daily evolutionary pattern: the average streaming quality is generally better at early hours of a day, degrades to a low value around the noon time, then improves slightly, and drops to its daily lowest value before midnight. The daily lowest value can be $0.1 - 0.3$, meaning that the majority of peers in UUSee are regularly experiencing a downgrade of streaming quality in the evenings.

The patterns of streaming inefficiencies have raised an intriguing question: what has caused such low streaming qualities at peers at evening times and in specific channels? In what follows, we extensively explore the influential factors.

IV. DIAGNOSING STREAMING INEFFICIENCY: THE DAILY EVOLUTION

We first investigate the cause to the daily evolutionary pattern of streaming qualities, by exploring the correlations between the evolution of streaming quality in the streaming channels and the variation of the following influential factors:

Number of peers in a channel, which decides the level of bandwidth demand and supply in a channel.

Average upload bandwidth per peer, which constitutes the main source of upload bandwidth in a P2P streaming channel.

Average server bandwidth consumption per peer. Server upload bandwidth plays an indispensable role in cases of insufficient peer bandwidth contribution and peer dynamics.

Intra-ISP/Inter-ISP per-link bandwidth availability. We investigate bandwidth availability along P2P links within an ISP (the intra-ISP case) and across ISP boundaries (the inter-ISP case), which reflects possible inter-peer bandwidth bottlenecks (Peers' ISPs are derived by mapping from their IP addresses using a mechanism provided by UUSee).

Average peer indegree, which is related to the level of download bandwidth at the peers.

Average buffer map difference at a peer, which represents the level of content availability along each incoming P2P link at a peer, and is computed in the following fashion: the total number of non-duplicated new blocks a peer's partners can provide it is counted by comparing their buffer maps, and then divided by the number of partners. In a mesh streaming system, a larger buffer map difference represents less chance of content bottleneck and a higher probability to saturate the available bandwidth on each P2P link.

A. Evolution of the influential factors

With the example of a popular channel CCTV1 and a less popular channel CCTV12, Fig. 3 plots the evolution of streaming quality and values of the factors in the channels. A daily variation pattern can be observed in all the evolution series. A closer look reveals that the drop of streaming quality in the evenings is generally accompanied by the increase of peer population, and the decrease of available server capacity per peer and per-link bandwidth availability. On the other hand, a bit surprisingly, we notice that the peer upload bandwidth contribution, the buffer map difference and the peer indegree may have increased at the evening peak hours.

To verify our observations, we compute the *cross-correlation* between the evolution series of each factor and that of the streaming quality in each streaming channel. The cross-correlation between two time series, $x_t, t = 1, 2, \ldots$ and $y_t, t = 1, 2, \ldots$, at a delay of d is evaluated as the correlation between the corresponding members of x_t and $y_{t+d}, t = 1, 2, \ldots$, in the two series. For example, we may compute the cross-correlation between the time series of the
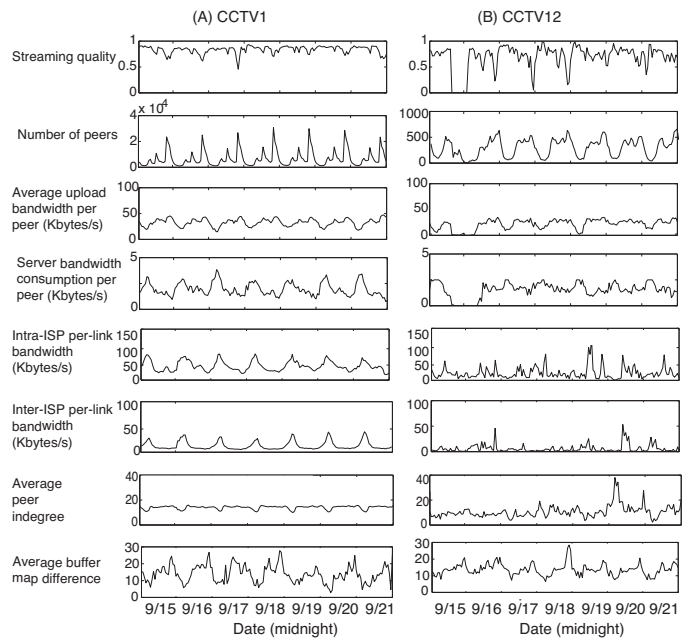

Fig. 3. Evolution of streaming quality and influential factors in September 15 — 21, 2007 (GMT+8).

streaming quality and the number of peers for CCTV1 and CCTV12, respectively, at a delay of 0.

For CCTV1, we obtain a significant negative cross-correlation at a delay of 0, which validates a strong negative correlation between the peer population and the achieved streaming quality in this channel over time. For CCTV12, we obtain a less significant negative value at a delay of 0, which shows a weaker negative correlation between its series.

B. Cross-correlation with the influential factors

To investigate whether our observations for CCTV1 and CCTV12 generally exist among other channels, we have extensively investigated the cross-correlations between evolution series of streaming quality and each of the influential factors for all the channels in UUSee. As we have observed different levels of cross-correlations for channels of different sizes (*i.e.*, different average peer populations), we show the cross-correlation in each streaming channel by plotting it against the size of the channel, as in Fig. 4, in order to discover any possible relationship between the two. In each sub figure in Fig. 4, the y-axis represents the cross-correlation at a delay of 0 between the two series of the streaming quality and one respective factor (we have used "cross-correlation with the respective factor" for short in the y labels), and each sample in the plot denotes one streaming channel.

Fig. 4(A) shows that for most channels in UUSee (as denoted by the dense areas), the cross-correlation between the evolution of streaming quality and that of the peer population is negative, meaning that the streaming quality is worse with more peers in the channels. Such a negative cross-correlation is more significant for popular channels with larger peer populations; only in channels with less than a few hundreds of peers, can the cross-correlation be positive. Such a negative impact of peer population in most channels has *conflicted* with
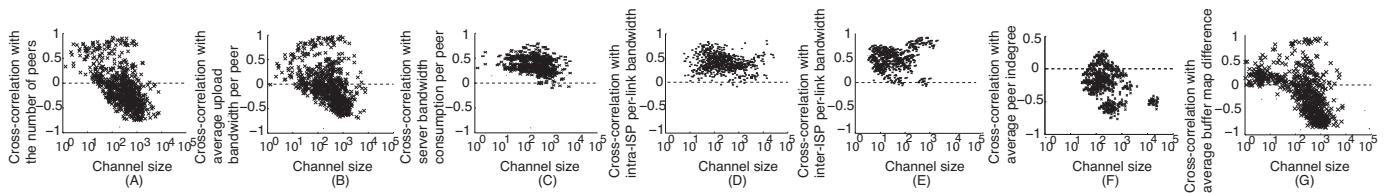
Fig. 4. Cross-correlation between the series of streaming quality and each influential factor for all the streaming channels in UUSee.

the common belief that a P2P streaming system may scale better with more peers in the channel. But why is it so?

Fig. 4(B) reveals a negative cross-correlation between the streaming quality and average peer bandwidth contribution for most channels. This has made the question more interesting: why shall the streaming quality in most sizable channels still degrade at daily peak hours with more peers and larger bandwidth contribution from each peer?

Fig. 4(C), (D) and (E) bring us useful insights towards the answer. A positive cross-correlation is exhibited in these figures for almost all channels, revealing the evolution of streaming quality in each channel is largely caused by the evolution of those three factors: at daily peak hours, the less available server capacity to each peer and the lower level of available bandwidth along the P2P links, have produced a negative impact on the streaming quality. However, why would the available server capacity become less sufficient with more peer bandwidth contribution at peak hours? We believe the reason is that: even when peers contribute more at peak hours (as shown in Fig. 3), the level of peer bandwidth contribution is still lower than the required streaming bandwidth (*i.e.*, 500Kbps), and the insufficient amount has to be covered by the upload capacity from the servers. As the total server capacity is limited and the increase of peer population is more significant than that of peer bandwidth contribution, the average server capacity available to each peer becomes less with more peers.

Fig. 4(F) plots a negative cross-correlation between the streaming quality and the average peer indegree in most channels. This brings another interesting discovery: at daily peak hours, peers tend to obtain more partners (possibly as a result of trying to boost their degraded streaming quality), but nevertheless the streaming quality is poorer, due to the lower bandwidth availability along each link.

Fig. 4(G) shows a negative cross-correlation for most large channels. It shows that at peak hours in large channels, there is no content bottleneck among peers, as the number of available blocks to exchange is actually larger than that at other times. Therefore, the content availability does not constitute a significant cause to the streaming quality downgrade in large channels. On the other hand, in small channels, the positive cross-correlations reveal the existence of content bottlenecks among peers, when the low streaming quality occurs.

In summary, the above observations bring us a number of intriguing insights, with respect to the causes to the streaming inefficiencies commonly existing on a daily basis:

In large channels with hundreds or thousands of peers, the lack of server capacity per peer and the downgrade of inter-peer bandwidths contribute significantly to the downgrade of

streaming quality with more peers in the channel.

In small channels with less than a few hundreds of peers, streaming inefficiency usually occurs at times with a smaller peer population, due to a general lack of server, link and peer bandwidths and the available content blocks to exchange.

The increase of partner numbers in both large and small channels may not help boosting the peer streaming quality, as the available bandwidth along the links is the key.

In a practical P2P streaming system in today's Internet, server capacity is still indispensable to guarantee the streaming quality of mid-quality streams around 500Kbps.

V. DIAGNOSING STREAMING INEFFICIENCY: FOCUSING ON ONE SNAPSHOT

We have observed that there exist channels that regularly experience a lower streaming quality than others at any specific time in Fig. 1. We now inspect the causes by focusing on one representative snapshot at 9pm September 21 2007.

The influential factors we investigate for their impact on the distribution of streaming inefficiency across channels include: *number of peers* (channel size), *average peer upload bandwidth*, *average server bandwidth consumption per peer*, and *average buffer map difference at a peer* in each channel. The average peer indegree and per-link bandwidth availability are not included as their values do not differ much among channels at the same time.

Fig. 5 plots the correlation between channel streaming quality and each of the factors. Each sample in the plots represents one channel. A positive correlation is observed with the Pearson product-moment correlation coefficients, *rho*, marked in each plot. It shows that the streaming quality is generally better in channels with a larger peer population, larger average peer bandwidth contribution, higher server capacity availability per peer, and larger buffer map differences.

To dig deeper on the rationale behind the positive effect of peer population on streaming quality across channels, we have explored the correlation between the channel size and the other factors. Interested readers are referred to our technical report [3] for the correlation plots. Our investigation shows that the average peer bandwidth contribution is larger in a larger channel due to (1) the higher percentage of Ethernet peers in larger channels and (2) the larger buffer map difference in larger channels, such that the available peer upload bandwidth is more efficiently utilized. Our investigation also shows that the average server capacity used by each peer is larger in larger channels. In the UUSee protocol, there is no allocation of server capacity to different channels, and servers can be selected to serve peers from any channels in a similar way as any regular peers are selected. In this case, our investigation
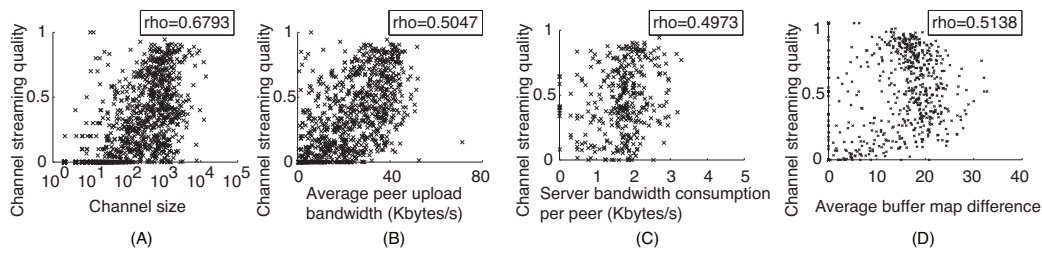
Fig. 5.   Correlation of channel streaming quality with influential factors: for all the channels in UUSee at 9pm September 21 2007 (GMT+8).

shows that popular large channels may grasp not only more server bandwidth in total, but also larger server capacity per peer, which is not a fair situation for the less popular channels.

In summary, we can derive the following insights towards the distribution of streaming inefficiency across channels at each specific time:

At each specific time, less popular channels tend to experience lower streaming qualities, due to a lower percentage of high bandwidth peers, less efficient utilization of peer upload bandwidth due to largely overlapping buffer maps among a small peer population, and a lower level of achievable server capacity in competition with large channels.

VI. RELATED WORK

With respect to measurement studies targeting commercial P2P live streaming, Hei *et al.* [4], [5] studied *PPLive*, using traces collected by crawling and passive sniffing. Ali *et al.* [6] studied PPLive and SopCast with traces collected in 2-3 hour durations using passive sniffing. Silverston *et al.* [7] compared four P2P streaming applications, PPLive, PPStream, SopCast, and TVAnts, using passive sniffing during the broadcast of the 2006 FIFA World Cup. These studies mainly investigated peer distributions, control/data traffic, peer partnership, etc., but have not focused on exploring the causes to the low streaming qualities in P2P streaming.

Li *et al.* [8], [9] studied CoolStreaming using a similar trace collection method as ours, by deploying an ActiveX component at peers, which reports peer statistics periodically to a logging server. Their trace collection lasted for a few days, instead of over a long period of time as in this paper. Their focus of study has not been on an extensive investigation of the causes to streaming inefficiencies. To the best of our knowledge, our work represents the first extensive measurement study targeting at diagnosing network-wide low streaming qualities in large-scale P2P live streaming.

VII. CONCLUDING DISCUSSIONS

This paper presents a first effort to extensively explore network-wide streaming inefficiencies in large-scale mesh-based P2P live streaming applications, with abundant traces from a real-world P2P streaming system. We have identified an evolutionary pattern of streaming inefficiencies in the system over different times, and its distribution in channels with a smaller population.

Our extensive investigation of causes to such streaming inefficiency patterns has brought forward the following intriguing discoveries: (1) The population of peers has two-sided effects on the streaming quality in a streaming channel: while peers in a larger channel typically enjoy a better streaming quality, the streaming quality degrades when peer number increases in the channel during daily peak hours. (2) Server capacity still plays an indispensable role in P2P streaming over today's Internet, and the inefficient supply of server capacity has largely led to the low streaming qualities at peak times and in less popular channels. (3) Smaller channels tend to obtain less server capacity per peer, when the servers do not explicitly allocate capacity among channels. (4) Inter-peer bandwidth availability represents a more significant bottleneck than peer upload bandwidth during peak hours. (5) Increasing peer indegree does not help enhancing peer streaming quality, where it is the per-link bandwidth availability that matters.

These discoveries have pointed us to a number of possible directions to improve the mesh-based P2P streaming protocol. First, the available server capacity should be explicitly and dynamically allocated across different streaming channels, based on their current bandwidth demand and supply at each time; and it is desirable to dynamically adjust the total amount of server capacity deployed in the system, both to meet the demand at peak hours and to save the cost at off-peak times. Second, the locality of partners should be explored especially during daily peak hours, when the inter-ISP bandwidth bottleneck is more evident. Third, the peer selection protocol should include more judgment of good partners with high inter-peer bandwidth, instead of finding many more partners. As ongoing work, we are designing better streaming protocols that make our suggestions more concrete.

REFERENCES

[1] *PPLive*, http://www.pplive.com/.
[2] *UUSee*, http://www.uusee.com/.
[3] C. Wu, B. Li, and S. Zhao, "Diagnosing Network-wide P2P Live Streaming Inefficiencies," http://cairo.csl.toronto.edu/papers/diagnose.pdf, ECE, University of Toronto, Tech. Rep., January 2009.
[4] X. Hei, C. Liang, J. Liang, Y. Liu, and K. W. Ross, "A Measurement Study of a Large-Scale P2P IPTV System," *IEEE Trans. on Multimedia*, vol. 9, no. 8, pp. 1672–1687, December 2007.
[5] X. Hei, Y. Liu, and K. W. Ross, "Inferring Network-Wide Quality in P2P Live Streaming Systems," *IEEE Journal on Selected Areas in Communications, Special Issue on Advances in Peer-to-Peer Streaming Systems*, vol. 25, no. 9, pp. 1640–1654, December 2007.
[6] A. Ali, A. Mathur, and H. Zhang, "Measurement of Commercial Peer-To-Peer Live Video Streaming," in *Proc. of Workshop in Recent Advances in Peer-to-Peer Streaming*, August 2006.
[7] T. Silverston and O. Fourmaux, "Measuring P2P IPTV Systems," in *Proc. of NOSSDAV 2007*, June 2007.
[8] B. Li, S. Xie, G. Y. Keung, J. Liu, I. Stoica, H. Zhang, and X. Zhang, "An Empirical Study of the CoolStreaming+ System," *IEEE Journal on Selected Areas in Communications, Special Issue on Advances in Peer-to-Peer Streaming Systems*, vol. 25, no. 9, pp. 1627–1639, December 2007.
[9] B. Li, Y. Qu, Y. Keung, S. Xie, C. Lin, J. Liu, and X. Zhang, "Inside the New Coolstreaming: Principles, Measurements and Performance Implications," in *Proc. of IEEE INFOCOM*, April 2008.