

# DISTRIBUTED RATE-DISTORTION OPTIMIZATION FOR RATELESS CODED SCALABLE VIDEO IN MOBILE AD HOC NETWORKS

Thomas Schierl<sup>1</sup>, Stian Johansen<sup>2</sup>, Cornelius Hellge<sup>1</sup>, Thomas Stockhammer<sup>3</sup>, and Thomas Wiegand<sup>1</sup>

<sup>1</sup>Fraunhofer Institute for Telecommunications  
Heinrich-Hertz-Institut (HHI)  
Einsteinufer 37, D-10587 Berlin, Germany  
{schierl|hellge|wiegand}@hhi.de

<sup>2</sup>Norwegian University of Science  
and Technology  
NO-7491 Trondheim, Norway  
stianjo@q2s.ntnu.no

<sup>3</sup>Nomor Research GmbH  
Tannenweg 25  
D-83346 Bergen, Germany  
stockhammer@nomor.de

## ABSTRACT

Recent advances in forward error correction and scalable video coding enable new approaches for robust, distributed streaming in Mobile Ad Hoc Networks (MANETs). This work presents an approach for distribution of real time video by different uncoordinated peer-to-peer relay or source nodes in an overlay network on top of a MANET. The approach proposed here allows for distributed, rate-distortion optimized transmission-rate allocation for competing scalable video streams at relay nodes in the overlay network. Furthermore the approach has the desirable feature of path/source diversity for enhancing reliability in connectivity to serving nodes. Signaling overhead within the overlay network is kept at a minimum, since optimizations are done at relay nodes and clients rather than at servers.

**Index Terms**— Channel coding, Video coding, Multimedia systems, Transport protocols, Network reliability

## 1. INTRODUCTION

Recently, MANETs [1] e.g. based on WLAN IEEE 802.11 ad hoc or on the upcoming IEEE 802.16j Mobile Multihop Relay techniques have gained interest for delivery of multimedia content and other mobile services. Similar to push services in 3G networks, new services can be introduced based on ad hoc groups built on top of MANETs. MANETs are attractive due to low infrastructure costs, especially in areas with high user density. The coverage area for mobile services can generally be extended through cooperation with neighboring nodes. In MANETs, user terminals in a mobile network are conceptually not assumed to be receivers only, but can also be used as routing nodes in order to build a dynamic network infrastructure.

User nodes building a MANET are assumed to be highly mobile, which results in the dynamic characteristics of this network type. Thus a topology built upon a MANET cannot be truly robust against network separation as well as against route or path loss. Therefore, clients typically experience loss of connection to serving nodes [2].

Multimedia delivery services in MANETs can be implemented using non real-time downloads or real-time streaming. Download delivery in general lacks to meet of timing constraints for media data. However, by appropriate end-to-end protocols [3], one could more easily deal with connectivity loss and longer outages in MANETs in order to provide full reliability. However, if real-time delivery is an essential service requirement and streaming delivery needs to be used due to the associated delay constraints [2], reliability is much harder to achieve. With common point-to-point transmission techniques such as link layer forward error correction or retransmission protocols, sufficiently good service quality in MANETs is often not

possible, because of entire link outages and disconnections. Hence, solutions are desirable for satisfying the different connectivity requirements for real time streaming, most importantly timing constraints and the requirement of continuous media presentation at the client. Our suggested solution to this problem is to enhance source connectivity by using a diversity of serving nodes combined with the use of a rateless forward error correction code also known as fountain codes, as described in [4].

The approach presented in this paper combines the benefits of cooperative interaction of client and peers (also referred to as sources or relay nodes) in an overlay network on top of a MANET for enhancing reliability in connectivity. Due to the network dynamics as described above, most of the problems can only be solved with appropriate end-to-end service provisioning. For this purpose, we rely on end-to-end application layer techniques, but also assume that peers cooperate by applying optimized rate allocation for their competing video streams. For suitable application layer QoS, we rely on two technologies: Scalable Video Coding (SVC) and application layer forward error correction (AL-FEC). The scalable video stream allows for rate adaptation at peers in the overlay network by just forwarding or not forwarding a network stream. Further, by using an efficient and flexible FEC code, the Raptor code [4], in combination with SVC, a distributed reception of the real time video data from uncoordinated peers is realized. The basic approach for the distribution of the media streams has been shown in [2]. This work extends the work in [2] in the sense that each layer is encoded by a separate FEC encoding process similar to the proposal in [5]. Furthermore, we investigate the case where source and/or relay nodes perform locally optimized rate-allocation for the FEC encoded scalable media for different connected clients. The approach takes into account local competing traffic, characteristics of the video streams as well as reception feedback from the client. The local optimization results in a purely distributed optimization approach which is a key feature for operation in a MANET. By that, reliable communication in unreliable environments like MANETs can be achieved, while at the same time maximizing the decoded video quality at receivers.

## 2. MEDIA TRANSPORT IN MANETS

This section provides a brief overview of the proposed techniques for reliable real-time media distribution in MANETs. The detailed approach has been introduced in [2]. Assume that a video stream is encoded in different layers, preferably applying H.264/AVC-based Scalable Video Coding (SVC) [6, 7]. Furthermore, assume that each sub-stream (layer) is encoded by applying a separate FEC encoding process. We apply the rateless FEC Raptor code as described below. In the following, we refer to the process of generating in-

dependent encoding symbols for each video substream as Rateless Scalable Video Coding (RSVC). In [2], the generated FEC code symbols of the different video layers are distributed into network packets based on Priority Encoding Transmission [8] (PET). In this work, the RSVC process is extended by transporting the different RSVC streams on different network transport streams, which is similar to layered transmission as for example shown for SVC in [9]. In this framework, appropriate rate assignments of the rateless or fountain encodings of the different layers allows for layer-specific QoS through unequal error protection.

The Raptor code [4] is an erasure correction code mainly used in environments with packet losses. The rateless or fountain property of the Raptor code implies that a virtually infinite amount of independent encoding (output) symbols (ESs) can be generated from a limited number of source (input) symbols (SSs). Transmitting these ESs intelligently over different paths using different sources, significantly enhances the reliability of streaming sessions in MANETs. For the multiple source scenario, a randomization mechanism has been proposed in [2] for making the different Raptor encodings at different sources linearly independent without the need for coordination among the sources. Because of this property, a Raptor decoder at a receiver does not need to be aware from which source a symbol originates from, only the amount of received symbols has to be sufficient.

Fig. 2 shows the RSVC network stream aggregation. A source block (SB) of source symbols corresponding to one timeframe of the scalable real time video data with duration  $t_{SB}$  is encoded with different Raptor encodings per media layer  $l$ . This theoretically allows for producing an unlimited number  $n_l$  of ESs per SB and layer  $l$  from  $k_l$  SSs. Assume now that for a source block of length  $t_{SB}$ , a receiver receives  $\tilde{n}_l^s$  encoding symbols from each source  $s$  for substream  $l$ , corresponding to video layer  $l$ . The efficiency of the Raptor code is such that if, on average, the sum of received symbols for layer  $l$  from  $S$  sources  $\sum_S \tilde{n}_l$  is only slightly greater than the number of SSs,  $kl$ , then the media data of layer  $l$  within this SB of length  $t_{SB}$  can be recovered [2].

### 3. DISTRIBUTED RATE-DISTORTION OPTIMIZATION FOR RSVC IN MANETS

This section gives an overview of the use-case for the proposed approach and the assumptions made for this work. Furthermore the section presents the details and the realization of the distributed optimization approach for RSVC.

#### 3.1. Use case — overlay swarms in MANETs

In order to overcome the connectivity problem in MANETs [2], it is proposed to use an overlay network based on a swarm approach. The swarm approach was initially introduced in [10] and we integrate the swarm approach for lowering traffic on overlay paths, but also for enhancing connectivity in case of connection loss or ungraceful peer disappearance. We extend the approach for Real Time Streaming in MANETs as introduced in Section 2 by the application of overlay swarms. This requires that sources are synchronized on a timeframe/source block accuracy, e.g. by forwarding a live stream coming from a different downlink connection. Furthermore, the extension of the swarm approach to our scenario requires that all sources are contributing to the same timeframe with independent encoding symbols for reconstructing the original data. This way of requesting data in a distributed peer-to-peer fashion is also known as "Gossiping".

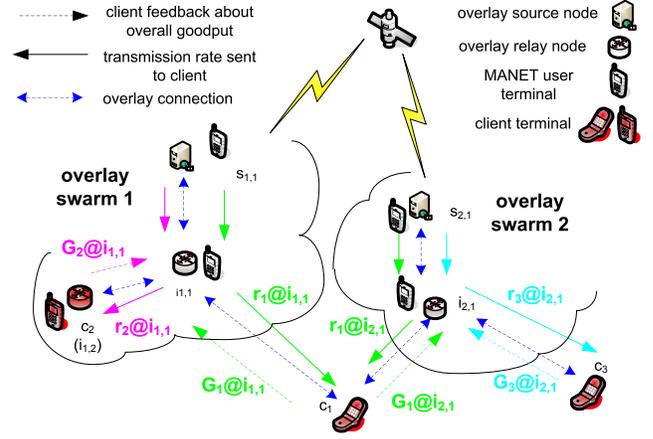


Fig. 1. Overlay networks for MANET multiple source media distribution based on RSVC.

Fig. 1 shows the principal view on overlay networks on top of a MANET, where  $s_{x,y}$  and  $i_{x,y}$  represent source respectively relay node  $y$  of swarm  $x$  and  $c_z$  represents client  $z$ . An overlay swarm consists of all connected sources, relay nodes and clients that can act as relay node due to their network interface, processing power and battery capabilities. We assume that source nodes in the swarms are fed by additional access networks, e.g. by a reliable wireless downlink connection possibly provided through a terrestrial or satellite transmitter with a scalable SVC live video signal. Raptor encoding is applied at the source nodes only.

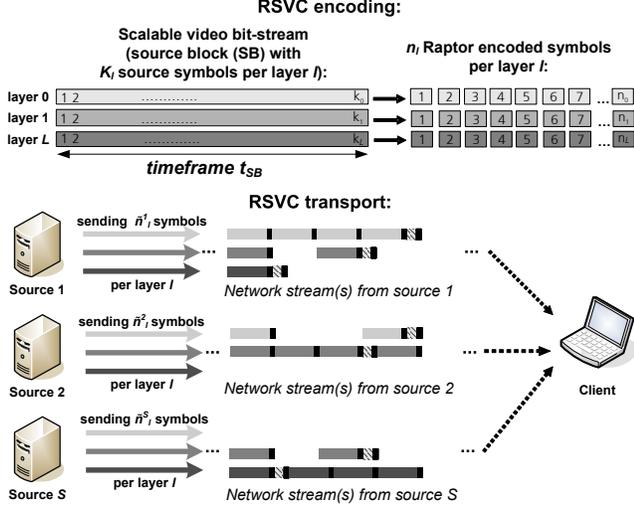
For ease of exposition we only refer to source-to-client connections in the remainder, since all connections of type sources-to-relay, relay-to-relay and relay-to-client can be viewed as a source-to-client connection but without Raptor re-encoding in the overlay. The optimization described in the next subsection is valid for both, source and relay nodes and we assume that the same procedure is carried out at both entities.

We further assume in the following that the rate available for transmission on an overlay path is known. This may be achieved by techniques as proposed in [11] which observes the time sharing and contention of the wireless channel at each MANET relay for rate allocation.

#### 3.2. Rate distortion optimized streaming of RSVC streams

In this section, we describe our distributed approach for rate-distortion optimized resource sharing. The limited capacity at overlay nodes in MANETs is shared in such a way that the total (sum) of video qualities experienced at receivers is maximized. The approach is inherently receiver-driven, since the connections to the multiple sources are requested by the receiver. For simplicity and lack of space, we explain the proposed procedure through an example.

Referring to Fig. 1, we have a situation with two sources ( $s_{1,1}$  and  $s_{2,1}$ ), three clients ( $c_1$ ,  $c_2$  and  $c_3$ ) and two relay nodes ( $i_{1,1}$  and  $i_{2,1}$ ). Source  $s_{1,1}$  and overlay node  $i_{1,1}$  belong to swarm 1 while source  $s_{2,1}$  and overlay node  $i_{2,1}$  belong to swarm 2. Client  $c_1$  sends a request for the green video stream to the two neighboring overlay nodes. The request from  $c_1$  contains its current receiving rate (measured in terms of goodput, i.e. the reception rate of correct packets) summed over connected sources. It is proposed to signal the goodput in sufficiently small intervals for allowing proper rate adaptation within  $t_{SB}$  and to overcome losses of these messages. Since  $i_{1,1}$



**Fig. 2.** Rateless Scalable Video Coding (RSVC) source block aggregation for uncoordinated sources.

and  $i_{2,1}$  are connected to  $s_{1,1}$  and  $s_{2,1}$ , respectively, they are able to relay the video stream to  $c_1$ . However, since there are concurrent requests or already existing streams to clients  $c_2$  and  $c_3$ , there may not be sufficient capacity for carrying the full rate of all video streams to all clients. How to share the available rate for the different users and streams is decided at the serving overlay nodes (this optimization is described later in this section). For overlay nodes to be able to do such rate allocation, rate-distortion information about the layers of SVC encoded video is required. It is proposed that this information is sent at safe intervals within each SB.

Consider the situation where  $N$  clients are requesting video streams from a source. Without loss of generality we assume that each client is receiving exactly one video stream. Further assume that the sending capacity at the source  $s$  in question is limited to  $R_{s,avail}$ , and that  $R_{s \rightarrow n}^{path}$  denotes the available rate on the path to client  $n$ . Both values (or estimates of them) are assumed to be known to the source node. A video stream is characterized by a set of distortion points  $d_l$ , which represent some measure (like MSE) of the difference between encoded and original video, with  $\rho_l$  being the corresponding encoding rate of the scalable video stream at layer  $l \in \{0, \dots, L-1\}$ . A function  $D$  maps the total goodput  $g$  being received by a client to discrete distortion points  $d_l$  of the video stream received by the client in question.

$$D : g \rightarrow d \quad g \in \mathbb{R}^+ \quad d \in \{d_0, \dots, d_{L-1}\} \quad (1)$$

A client  $n$  experiencing a goodput  $g_n$  is assumed to operate at distortion point  $d_l$  (corresponding to distortion at layer  $l$  of the scalable video stream received by client  $n$ ) given by  $D(g_n)$ , where goodput is defined as

$$g_n = \sum_{s=1}^S (1 - \gamma_{s \rightarrow n}) r_{s \rightarrow n} \quad (2)$$

for client  $n$  receiving data from  $S$  sources, each sending at rates  $r_{s \rightarrow n}$ , over paths characterized by packet loss rates  $\gamma_{s \rightarrow n}$ . Note that the above only holds for constant packet size, which is typically the case for the FEC-encoded packets used in this work. We assume in the following that the packet loss rate  $\gamma_{s \rightarrow n}$  is known and independent of the transmit rate.

The source attempts to minimize the average distortion experi-

enced at all connected clients as follows:

$$\min_{\{\Delta r_{s \rightarrow 1}, \Delta r_{s \rightarrow 2}, \dots, \Delta r_{s \rightarrow N}\}} \left( \sum_{n=1}^N D_n(g_{n,opt}(\Delta r_{s \rightarrow n})) \right) \quad (3)$$

The optimal goodput  $g_{n,opt}$  for client  $n$ , as stated in (4), is calculated with  $r_{s \rightarrow n}$  being the rate the source  $s$  is sending to client  $n$ , and  $\Delta r_{s \rightarrow n}$  being the change in sending rate (which is the result of the optimization in (3)). The set of sending rate changes  $\{\Delta r_{s \rightarrow 1}, \Delta r_{s \rightarrow 2}, \dots, \Delta r_{s \rightarrow N}\}$  for the  $N$  clients are the parameters subject to optimization.

$$g_{n,opt} = \underbrace{g_n - (1 - \gamma_{s \rightarrow n}) r_{s \rightarrow n}}_{\text{Old goodput from this source}} + \underbrace{(1 - \gamma_{s \rightarrow n}) (r_{s \rightarrow n} + \Delta r_{s \rightarrow n})}_{\text{New goodput from this source}} + \underbrace{\text{Goodput from other sources}} \quad (4)$$

which simplifies to

$$g_{n,opt}(\Delta r_{s \rightarrow n}) = g_n - (1 - \gamma_{s \rightarrow n}) \Delta r_{s \rightarrow n} \quad (5)$$

Optimization is done under constraints (6) through (8).

$$\sum_{n=1}^N (r_{s \rightarrow n} + \Delta r_{s \rightarrow n}) \leq R_{s,avail} \quad (6)$$

$$r_{s \rightarrow n} \leq R_{s \rightarrow n}^{path} \quad (7)$$

$$(1 - \gamma_{s \rightarrow n}) r_{s \rightarrow n} \leq \Delta r_{s \rightarrow n} \leq R_{s \rightarrow n}^{path} - r_{s \rightarrow n} \quad (8)$$

Here, (6) constrains the rate increase at source node  $s$ , (7) restricts the rate on the path to the receiver, and condition (8) gives the upper and lower bounds on the rate change  $\Delta r_{s \rightarrow n}$ .

Solving (3) will lead to a convex optimization problem described by lagrangian cost functions as  $D + \lambda R$ .

When relay nodes have carried out the RSVC rate allocation for its requesting clients, information about the allocated rate  $r_{s \rightarrow n}$  is propagated to connected clients. Based on these messages from the serving overlay nodes, clients decide which rates for each media layer should be requested from each overlay node. In other words, the client is partitioning its total allocated rate to subscriptions for the video layers at the available overlay nodes.

This optimization procedure fulfills two important aspects of MANET communication: Cooperation and distributed processing. Each participating node carries out its own optimization and propagates the decisions which other nodes use for their own optimization. The stability of this heuristic algorithm needs to be investigated, since it simultaneously depends on the dynamics of the network and dynamics of the involved video streams. Nevertheless, extensive simulations show that stable operation and video quality with excellent bit rate sharing is achieved.

#### 4. SELECTED SIMULATION RESULTS

In this section we present a set of selected simulation results showing the benefits of the proposed approach. We encoded three different ITU-T video sequences (repeated forward and backward) City, Crew and News in QCIF res. at 15 frame per second with a total length of about 100sec. All sequences are encoded using SVC reference software JSVM 8.8, with an H.264/AVC base layer and two SVC fidelity enhancement layers (ELs) with MGS [7], a group-of-picture (GOP) size of 16 and one IDR-frame in each GOP for random access. All streams are encoded at a rate of about 160kbit/sec, if decoding the highest enhancement layer. The rate points are achieved by removing NAL units of the enhancement layer from the bi-stream starting

	City		Crew		News	
	Rate	PSNR	Rate	PSNR	Rate	PSNR
Base layer	53.4	34.5	58.8	29.2	45.0	36.1
Enh. layer 1	81.8	37.2	77.0	31.0	69.2	38.7
Enh. layer 2	88.9	37.4	87.4	31.4	78.7	39.4
Enh. layer 3	142.3	39.9	125.8	33.5	128.5	42.4
Enh. layer 4	169.4	40.9	150.4	34.7	159.5	44.3

**Table 1.** PSNR and rate values for base layers and enhancement layers for the 3 transmitted SVC coded video sequences

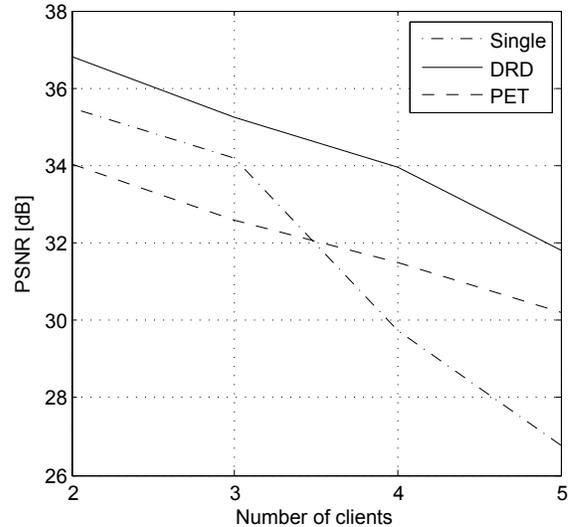
with the lowest temporal priority. The resulting PSNR values are shown in Tab. 1.

We generated one SB every two GOPs, i.e. the minimum adaptation interval is about 2.13 sec. Raptor performance is evaluated by applying the simulation approach introduced in [4]. The system has been integrated into the ns-2 simulation environment presented in [2]. We simulated scenarios within a MANET with 30 nodes moving on random waypoint patterns at a maximum speed of 3m/s within an area of 650x650m. In each scenario the number of available sources was kept constant at 2, with each source having a fixed (maximum) sending rate of 200kbit/s. Clients were selected randomly, varying the number of clients from 2 to 5. For each different number of clients in the system, we simulated 40 scenarios, each having a simulation length of 100sec. Each client selects a video stream from the set of available sequences (see table 1) in a round-robin fashion, and picks server(s) depending on transport method and server connectivity. For comparison, two other transport methods are simulated as well as the method proposed here. Specifically, the PET approach of [2] and a state-of-the-art single server approach with rate-adaptation were simulated. Throughput limitations on the paths through the overlay are emulated by enforcing transmission rate limitations at the serving overlay nodes. We note that the available bandwidth on an overlay path could be dynamically estimated as proposed in [11].

Fig. 3 shows avg. received video quality over all clients in the overlay in terms of PSNR for the different methods. DRD denotes the distributed RD-optimized method proposed in this work, PET refers to the method of [2] and SINGLE to the single server system. The results show that the RD-optimized approach performs consistently better than the other two (up to approx. 4dB better than the single-server approach and always approx. 2dB better than PET). When the number of clients increases there are more degrees of freedom for doing the RD-optimization. Therefore, the performance gain experienced by the DRD approach increases with the number of clients. Due to the connectivity-preserving property of PET, it performs better than the single server, but performs worse with a low number of clients due to the PET rate overhead. The DRD approach gives an average performance gain over the other two systems, since the connectivity of clients and the RD-information about the video streams is taken into account.

## 5. CONCLUSIONS

We present an approach for robust real-time video transmission in MANETs based on an extended peer-to-peer swarm approach. This approach uses a rateless forward error correction code in combination with scalable video coding for distribution of layered video to different sources in an overlay network on top of a MANET. In particular, a distributed mechanism for rate allocation at relay nodes is presented. The rate allocation and by that the adaptation of the scalable video stream is done in a rate distortion optimized manner. That is, information about the possible rate distortion points for the layered video are taken into account in order to minimize the overall



**Fig. 3.** Average PSNR as a function of the number of clients.

distortion experienced at clients connected to the relay node in question. In future work we will examine the impact of traffic caused by the proposed scheme on other nodes in the network. By this we want to allow for optimized load balancing and reduction of contention experienced by nodes not part of the overlay.

## 6. REFERENCES

- [1] IETF, "IETF mobile ad-hoc network (MANET) working group, see <http://www.ietf.org/html.charters/manet-charter.html>,".
- [2] T. Schierl, K. Gaenger, C. Hellge, T. Stockhammer, and T. Wiegand, "SVC-based multisource streaming for robust video transmission in mobile ad-hoc networks," *IEEE Wireless Communications Magazine*, vol. 13, no. 5, 2006.
- [3] T. Stockhammer, T. Gasiba, A. Samad, T. Schierl, H. Jenkac, W. Xu, and T. Wiegand, "Nested harmonic broadcasting for scalable video over mobile datacast channels," *Wireless Communications and Mobile Computing*, vol. 7, no. 2, pp. 235–256, 2007.
- [4] M. Luby, T. Stockhammer, M. Watson, and T. Gasiba, "Reliable multimedia download delivery in cellular broadcast networks," *IEEE Transactions on Broadcasting*, vol. 53, no. 1, pp. 235–246, 2007.
- [5] J.P. Wagner, J. Chakareski, and P. Frossard, "Streaming of scalable video from multiple servers using rateless codes," in *Proceedings of ICME '06*, Toronto, ON, Canada, 2006, IEEE.
- [6] T. Wiegand, G. Sullivan, J. Reichel, H. Schwarz, and M. Wien, Eds., *Joint Draft 9: Scalable Video Coding (Doc JVT-V202)*, Joint Video Team (JVT), Marrakesh, Morocco, January 2007.
- [7] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, To appear.
- [8] A. Albanese, J. Blomer, J. Edmonds, M. Luby, and M. Sudan, "Priority encoding transmission," *IEEE Transactions on Information Theory*, vol. 42, no. 6, 1996.
- [9] S. Wenger, Y.-K. Wang, and T. Schierl, "Transport and signaling of SVC in IP networks," *IEEE Transactions on Circuits and Systems for Video Technology*, 2007, To appear.
- [10] B. Cohen, "Incentives to build robustness in bittorrent," in *1st Workshop of Economics of Peer-to-Peer Systems*, Berkeley, CA, USA, May 2003.
- [11] X. Zhu and B. Girod, "Distributed rate allocation for video streaming over wireless networks with heterogeneous link speeds," in *Proc. International Symposium on Multimedia over Wireless, (ISMW-07)*, Honolulu, HI, USA, September 2007.