DISTRIBUTED RATE ALLOCATION FOR VIDEO STREAMING OVER WIRELESS NETWORKS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Xiaoqing Zhu
May 2009

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Prof. Bernd Girod) Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Prof. Fouad Tobagi)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Dr. Peter van Beek)

Approved for the University Committee on Graduate Studies.

Abstract

Video streaming over wireless networks is compelling for many applications, ranging from home entertainment to surveillance to search-and-rescue operations. When multiple video streams share a wireless network, careful rate allocation is needed to prevent congestion, as well as to balance the video qualities among the competing streams. In this dissertation, we present a distributed media-aware rate allocation protocol, and evaluate its performance in the application example of streaming high-definition (HD) and standard-definition (SD) video over 802.11-based wireless home networks.

Our optimization framework incorporates heterogeneity in wireless link speeds and video rate-distortion (RD) characteristics, as well as traffic contention among neighboring links. The goal of the protocol is to minimize the total video distortion of all participating streams while limiting network utilization. It relies on cross-layer information exchange between video rate controllers at the end hosts and link state monitors at the intermediate relay nodes. Results from various network simulations confirm that the media-aware allocation outperforms TCP-Friendly Rate Control (TFRC) in terms of average video quality and fairness among the streams.

The protocol is further extended for the scenario of application-layer video multicast over wireless. Following the same mechanism of congestion price updates at relaying wireless nodes and video rate updates at sending peers, the multicast extension of the protocol is designed to support either non-scalable or scalable video streams. As in unicast video streaming, the proposed media-aware protocol achieves lower average video distortion of all participating peers than a TFRC-based heuristic scheme.

Acknowledgments

This dissertation results from several years of graduate study at Stanford. During this period, I enjoyed the inspiration, guidance, and support of many wonderful people.

My dissertation advisor, Professor Bernd Girod, constantly amazes me with his vision in research and his natural style to lead. As his student I have been profoundly influenced by his habits in thinking big and thinking outside the box, his appreciation of intellectual merits, as well as his mindfulness of practicality. It has been a great privilege working with and learning from him.

I started my first project on wireless video streaming in Professor Fouad Tobagi's class. Since then I have received much encouragement and many insightful suggestions from him. I was fortunate to have Dr. Peter van Beek as a knowledgable and caring mentor during my summer internship at Sharp Labs, and continued to learn from him in our discussions afterwards. I greatly appreciate the support of them both, for serving as my dissertation readers and providing valuable feedback. I am also grateful to Professor Nick McKeown for chairing my defense.

My graduate life was brightened by the many hours spent with the Image, Video, and Multimedia Systems (IVMS) group, in work or in play. I am indebted to Prashant, Chuo-Ling, Mark, and Markus for the many research skills they have patiently taught me. I cherish the sisterly bond with Anne, Houda, and Sangeun, built over various team projects together. Chatting with Eric was always fun. Our topics ranged from Linux programming to Chinese cooking. Over many cups of coffee together, Shantanu, Pierpaolo, Aditya, and David Varodayan provided me with tasteful recommendations on research papers and restaurants alike. I thank the other group members, Yao-Chung, Jeonghun, Zhi, Vijay, Gabriel, Sam, and David Chen, for constantly reminding me of a passion to learn, a spirit to explore. I thank Kelly Yilmaz for graciously facilitating all administrative details for me.

Over various joint projects I had the fortune to learn from many brilliant colleagues. The joint NSF project in collaboration with Taesang Yoo and Sachin Adlakha from Professor Andrea Goldsmith's group ignited my interests in cross-layer design for wireless video networking. The collaboration with Dr. Thomas Schierl was key to the multicast extension of the rate allocation protocol proposed in this thesis. I enjoyed working with Piyush Agrawal, Dr. Jatinder Pal Singh, and Dr. Tansu Alpcan in a team multi-homed in India, Berlin, and Stanford. As I wrapped up this dissertation at the Advanced Architecture and Research Group at Cisco, other group members, especially Rong, Nandita and Flavio, provided spiritual support for my efforts along with valuable technical inputs.

Many fellow graduate students well compensated the academic stress and challenges at Stanford with their delightful presence. Life on the Farm was much more fun with them around. I cherish my friendship with each one of them.

I owe a very special thanks to my caring husband Keji, who was always there to listen, to share, and to help. My deepest gratitude goes to him, and my beloved parents, Shulan Wang and Xiuchang Zhu. Their unconditional love and constant support empower me to pursue a path of my own. To them, I dedicate this thesis.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Video streaming over wireless networks is compelling for many applications, ranging from news and multimedia messaging services for cell phones, extended broadband Internet access in corporate or community networks, to wireless home entertainment or surveillance camera networks, to audiovisual communication in search-and-rescue operations.

In spite of the growing networking capabilities of modern wireless devices and the sophisticated techniques used by today's video coding and streaming systems, video streaming over wireless networks remains a challenging task. The wireless radio channel is subject to interference from other nearby transmitters, multipath fading, and shadowing, causing fluctuations in link capacities and sometimes an error-prone communication environment. The traffic patterns of compressed video streams typically change over time due to content variations and dynamic user behavior, and the received video quality may degrade drastically in the presence of packet losses, due to error propagation in the compressed bitstream. Moreover, video streaming applications typically have high data rates and stringent latency requirements, at odds with the limited bandwidth resources in a wireless network. Simultaneous streaming of multiple video sessions can easily lead to network congestion without careful rate allocation. The lack of centralized control in a wireless network, on the other hand, requires that the task of multi-user resource allocation be performed in a distributed manner.

This dissertation focuses on the problem of distributed rate allocation among multiple simultaneous video streaming sessions, so that they can efficiently share a wireless network without incurring excessive congestion. Since neighboring links compete for the same wireless radio channel, the rate of a video stream will not only affect the links along its own route, but also contend with traffic over other nearby links. The rate allocation problem is further complicated by heterogeneity in both the video rate utilities and the wireless link qualities. In this thesis, we take into consideration all the above factors to design a practical, distributed rate allocation protocol for video over wireless. The contributions are summarized as follows:

- *Formulation and analysis of a mathematical framework for multi-stream rate allocation over wireless networks.* In this framework, a wireless network model explicitly captures the effect of traffic contention among neighboring links and heterogeneous link transmission speeds. A parametric video distortion model is used to represent the utility of allocated rate for each stream. The multi-stream rate allocation problem is formulated within the convex optimization framework, with the goal of minimizing total video distortion while avoiding excessive network utilization. It is shown that the globally optimal solution can be achieved in a distributed fashion, by iteratively updating video source rates and link congestion prices. We further analyze dynamics of the proposed distributed solution, and establish system stability under proper parameter choices.

- *Design and simulation study of a distributed media-aware rate allocation protocol.* The proposed protocol allows cross-layer information exchange between link state monitors at the relaying wireless nodes and video rate controllers at the end hosts, so that each stream can quickly adapt its rate to various changes in the wireless networks and the video streams. Consequently, each video end host can regulate its stream rate according to explicit congestion prices accumulated along its path, instead of reacting to inferred congestion from packet losses or excessive delay. In comparison with conventional schemes such as TCP-Friendly Rate Control (TFRC), the proposed media-aware allocation leads to lower average video distortion and more balanced qualities among the streams.

- *Extension of the distributed rate allocation protocol for video multicast over wireless, supporting both scalable and non-scalable video streams.* Each relay node in the multicast tree actively maintains local link state information and a congestion price reflecting the impact on congestion of traversing video streams. Each parent peer in the multicast tree either actively performs video rate adaptation when relaying a scalable video stream, or collects an aggregated congestion price over its entire subtree when streaming non-scalable video streams. As in unicast video streaming, the multicast extension of the rate allocation protocol outperforms the TFRC-based heuristic scheme in terms of average video quality received by all participating peers.

Most simulation results presented in this dissertation are collected from scenarios of high-definition (HD) and standard-definition (SD) video streaming over 802.11a networks, a concrete example being a home media network. Nevertheless, we believe that the general principles of the proposed distributed rate allocation protocol carry over to other types of networks, and expect similar performance gains.

The rest of this dissertation is organized as follows. The next chapter reviews research in the related areas of wireless networking, congestion control, and video coding and streaming systems. Chapter 3 presents our optimization framework, together with stability analysis and numerical illustrations of the distributed rate allocation algorithm. In Chapter 4, we explain the design of a practical media-aware rate allocation protocol, based on cross-layer information exchange between video rate controllers at the end hosts and link state monitors at the relay nodes. Performance of the protocol is compared against a conventional media-unaware scheme based on TFRC, in network simulations involving various network topologies and different types of video content. Chapter 5 extends the media-aware rate allocation protocol for wireless video multicast. For delivery of non-scalable video streams, congestion prices are accumulated in a recursive manner and passed along to the root node for video rate adaptation. For delivery of scalable video streams, graceful quality reduction at intermediate nodes within the multicast tree becomes possible, again based on accumulated congestion prices. Finally, in Chapter 6, we summarize lessons learned from this dissertation and discuss future research directions.

Chapter 2

Background

The prospect of improving the performance of video streaming over wireless networks has motivated efforts across several research communities. This dissertation builds upon recent advances in wireless networking, congestion control, as well as video coding and streaming techniques. The following sections review the relevant state-of-art in these related areas.

2.1 Wireless Networking

2.1.1 802.11 Networks

Overview of 802.11 standards

To support communications over wireless Local Area Networks (LANs), the IEEE 802.11 standard provides specifications for both the physical layer and the Media Access Control (MAC) sublayer [7]. Later versions of the standard differ in their choice of modulation and frequency bands at the physical layer: 802.11b uses Direct Sequence Spectrum Spreading (DSSS) and operates at the 2.4 GHz Industrial, Scientific, and Medical (ISM) band [3]; 802.11a and 802.11g have adopted Orthogonal Frequency-Division Modulation (OFDM) and use spectrums centered at 5 GHz and 2.4 GHz respectively [1, 2]. The tutorials in [59], [97] and [228] provide more details on these standards.

The 802.11 MAC protocol has two operation modes: the contention-based Distributed Coordination Function (DCF) and the centralized Point Coordination Function (PCF). PCF is an optional mode, and it is rarely used in practice. On the other hand, DCF has attracted much attention both in industry and within the research community due to its fully distributed nature. In this dissertation, we assume that all wireless nodes operate in DCF mode, and describe its procedures in Appendix A.

Given increasing popularity of real-time voice and video traffic over wireless LANs, the 802.11e protocol has been developed to address the growing need for Quality-of-Service (QoS) support [8, 252, 179]. The Enhanced Distributed Channel Access (EDCA) scheme allows traffic classification at the MAC layer, and serves different traffic categories differently according to their priority levels by tuning their channel access parameters [21]. Despite the enhancements introduced by 802.11e, supporting QoS over 802.11 networks remains a challenging problem [144, 263].

New standardization efforts for 802.11n are devoted to increasing both the data rate and throughput in wireless LANs [4, 170, 172]. The IEEE 802.11n amendment promises transmission rates up to 600 Mbps by applying Multiple-Input-Multiple-Output (MIMO) technology across multiple antennas and bonding multiple frequency channels for transmission. The amendment is also designed to reduce MAC-layer overhead by aggregating transmissions of multiple consecutive packets, thereby improving throughput of payload data [249]

In the IEEE 802.11s draft standard for wireless mesh networking [9], the basic distributed MAC procedures in 802.11 are extended to support packet relays, meanwhile addressing many performance issues arising from a multi-hop environment, such as the exposed node problem [29]. An overview of this ongoing project can be found in [100].

Performance analysis

The saturating throughput of an 802.11 wireless LAN has been derived based on Markov modeling of the DCF procedures [25, 26, 28]. The analysis has also been extended to scenarios with non-saturating traffic [65], and to differentiated media access in 802.11e [183, 217].

General performance limits of ad hoc wireless networks have also been the aim of a number of information-theoretical studies. In a landmark paper, the capacity of a static wireless ad hoc network is shown to asymptotically vanish as the number of users increases [90]. Following the same methodology, upper bounds of the transport capacity over ad hoc networks are derived in [121] and capacity for energy-constrained networks are calculated in [184]. Later studies suggest that mobility can increase the capacity of wireless networks [89]. In addition, the scheme in [10] presents a mechanism to achieve the tradeoff between delay and throughput in a mobile network. In [54], it is shown that advanced signal processing techniques such as multiuser detection significantly improves the capacity of mobile ad hoc networks with delay constraints. The capacity region achieved by time-sharing in a wireless ad hoc network has also been characterized in [219] and [220].

Another thread of research has explored the practical capacity limits of 802.11 networks [27, 33]. Experimental studies show that throughput is typically significantly lower than the theoretical prediction in a multi-hop network, due to contention among adjacent links along the path [143]. Schemes for bandwidth estimation along a path over an ad hoc network are proposed in [200] and [43].

Performance issues

Many performance issues have been identified when the 802.11 MAC protocol is used for ad hoc networking [41]. It has been pointed out in [99] that the presence of one stream traversing a slow link significantly reduces the throughput achieved by other streams traversing faster links. More generally, this anomaly can be attributed to the design objective of maintaining max-min fairness among all competing wireless stations in 802.11 protocols [177].

For multi-hop networks, it has been reported that the congestion control procedures of TCP interact poorly with the exponential random backoff mechanism of the 802.11 MAC protocol [253, 62, 165]. As revealed in [24], [122], [42] and [237], serious fairness issues arise in certain network topologies, where TCP leads to partial or complete starvation of some of the flows.

2.1.2 Cross-Layer Design

Unlike in the conventional network structure with protocols independently designed for each layer, cross-layer design allows information sharing across the different layers for efficient utilization of network resources [48]. Recent research has studied joint optimization of physical layer power allocation, MAC layer link scheduling, network layer routing and transport layer flow congestion control [218, 60, 248, 173, 171, 246]. The joint optimization in [246] can be achieved by a distributed scheme based on local price updates and message exchanges among wireless nodes [245]. For multicast, network coding techniques [13, 53] can be combined with cross-layer design of power allocation, medium access and routing to maximize throughput, minimize power consumption or minimize network congestion [185, 259, 242, 243].

For video streaming over wireless networks, cross-layer design is both promising and challenging [226, 196]. The importance of adapting application layer video streaming rate and error protection parameters according to time-varying wireless channel conditions has been recognized fairly early [241, 222, 261, 260]. The studies in [224, 145, 155, 204] have unveiled potential benefits of adjusting lower layer parameters, such as 802.11 MAC layer retransmission limits and priority queueing, based on relative importance and urgency of media packets. Similarly, many research efforts manifest the performance gain from joint consideration of application-layer and link-layer adaptation techniques [201, 213, 169, 92]. In [88] and [157], multipath routing is combined with multiple description of video streams to leverage path diversity for better error resilience over a wireless mesh network.

A cross-layer design framework for video streaming is presented in [196]. By allowing information exchange and joint optimization of key parameters across different layers in the protocol stack, the framework allows greater flexibility for media and network adaptation while keeping the computational complexity tractable within the layered structure. Significant performance gain can be achieved over schemes with oblivious layers by exploring joint capacity and flow assignment, congestion-distortion minimized routing and packet scheduling, as well as media- and network-aware video rate allocation [257, 197, 199, 12].