Ad Hoc Networks 13 (2014) 282–295

Contents lists available at ScienceDirect

Ad Hoc Networks

journal homepage: www.elsevier.com/locate/adhoc

Optimal rate allocation in wireless networks with delay constraints

Fan Qiu *, Jia Bai, Yuan Xue

Institute for Software Integrated Systems, Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN, USA

ARTICLE INFO

Article history:
Received 17 May 2012
Received in revised form 11 August 2013
Accepted 14 August 2013
Available online 3 September 2013

Keywords:
Rate allocation
End-to-end delay
Wireless networks

ABSTRACT

The optimal rate allocation problem with end-to-end delay constraints in wireless networks is highly challenging due to the dynamics incurred by delay related factors, such as traffic arrival pattern, queuing process, and wireless resource sharing mechanism. In this paper, we solve the problem through a utility maximization framework with two sets of constraints: (1) capacity and schedulability constraints and (2) end-to-end delay constraints. The end-to-end delay of a flow can be adjusted by controlling the per-link delays via a novel parameter called *Virtual Link Capacity Margin (VLCM)*, which is the measurement of the gap between the schedulable link capacity and the maximum allowable flow rate over a link.

This optimization problem can be solved via its dual decomposition through two prices derived with regard to the constraints: the *link congestion price* that reflects the relationship between the traffic load and the capacity of a link, and the *flow delay price* that reflects the margin between the average packet delay and the delay constraint of a flow. We prove that the algorithm converges to the optimal solution under the $M/M/1$ queuing model. We have also discussed using our algorithm under generalized traffic patterns, where we use link delay functions to formulate the relationship between *VLCM* and the average link delay. The algorithm is implemented distributedly via joint wireless link scheduling, *VLCM* adjustment and congestion control. Extensive experiments are conducted over diverse network topologies and traffic arrival models. The experimental results show that our algorithm is able to achieve optimal rate allocation while satisfying the end-to-end delay constraints.

1. Introduction

End-to-end delay is highlighted as a critical performance requirement for numerous network applications, such as multimedia live streaming [1] and networked control systems [2]. Providing delay assurance, which is notably difficult even in wireline networks [3,4], has been a long lasting challenge for wireless networks. The main difficulty comes from the complex interactions between traffic arrival and departure, shaped by the scheduling mechanisms that deal with the location-dependent interference of wireless networks. The studies of [5,6] have focused on the delay analysis of wireless networks, where delay bounds are derived for given data flows. Understanding the delay bounds and minimizing the end-to-end delay, however, are barely enough for real world applications, which have strict delay requirements. On the other hand, traffic *elasticity* is exhibited by many networking applications, in which the flow rates can be adjusted to operate at multiple quality levels [7]. Thus, identifying an operating point, where wireless network resource is efficiently and fairly allocated to flows while satisfying their

* Corresponding author. Tel.: +1 615 3436216.
E-mail addresses: fan.qiu@vanderbilt.edu (F. Qiu), jia.bai@vanderbilt.edu (J. Bai), yuan.xue@vanderbilt.edu (Y. Xue).

1570-8705/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.adhoc.2013.08.010

end-to-end delay constraints, carries practical value, especially in resource-constrained wireless networks.

In this paper we investigate the optimal rate allocation problem in wireless networks with delay constraints. We adopt a utility maximization framework [8–11], where the utility of a flow is defined as a concave function of its rate. The objective is to maximize the aggregate utility of all flows subject to two constraints: the schedulability constraint and the end-to-end delay constraint. We introduce a novel parameter called *Virtual Link Capacity Margin (VLCM)*, which is the gap between the schedulable link capacity and the maximum allowable rate over a link. The end-to-end delay of a flow can be tuned via the per-link delays along its route by adjusting the *VLCM*s. Through the dual decomposition of the optimization problem, we employ a double-price scheme, where the *link congestion price* reflects the relationship between the traffic load and the capacity of a link, and the *flow delay price* reflects the margin between the average packet delay and the delay constraint of a flow. This naturally leads to a cross-layer design consisting of *VLCM* adjustment, wireless link scheduling and congestion control.

The main contributions of this paper are summarized as follows: (1) By introducing a novel parameter *Virtual Link Capacity Margin (VLCM)* that can be adjusted to jointly control the delay and the rate of a wireless link, this paper presents a tractable optimization formulation for maximizing the network utility under the end-to-end delay constraints. (2) We present a double-price based algorithm which converges to the optimal solution in a distributed way. (3) Our approach can be extended to handle generalized network models with unknown packet arrival patterns. Since this approach only requires the first order derivative of the delay-margin function, it can be applied to a large range of application domains.

The remainder of the paper is organized as follows: In Section 2, we briefly overview the related works. Section 3 introduces the system model. Section 4 describes the delay control mechanism. In Section 5 we propose a cross-layer rate allocation algorithm for wireless networks with end-to-end delay constraints, under the $M/M/1$ queuing model. Section 6 presents a distributed implementation of the optimal rate allocation algorithm. We introduce an extended algorithm to deal with generalized packet arrival patterns in Section 7. Section 8 presents the simulation results. Finally, Section 9 concludes the paper.

2. Related work

Considerable research efforts have recently been devoted to resource management in wireless networks, for instance, scheduling, rate allocation and routing, under delay constraints. One common objective of these works is to achieve performance optimality based on certain metrics, such as throughput, system utility and service disruption probability, while satisfying the delay requirements.

A large class of existing studies seek to solve the delay constrained QoS problems in wireless networks via scheduling. For example, the delay based scheduling algorithms in [12–14] use the delay information from the head-of-line packets as link weight, to achieve throughput optimality or network utility maximization in single-hop wireless networks. The studies of [15,16] make further attempts to solve the delay constrained QoS using game theory. The approach in [15] addresses the utility maximization problem under packet level delay constraints in wireless networks consisting of one access point and multiple clients. The optimization problem is decoupled into an access point sub-problem and a client sub-problem, which are jointly coordinated by a bidding game framework. In addition, a weight based scheduling policy is proposed to solve the access point sub-problem, where the weight of a client is related to its history of scheduling. The algorithm proposed by [16] maximizes network utility with respect to service rate under time varying channels, the central idea behind their solution is max-weight scheduling and VCG-based auction. The study in [17] is another example of delay constrained scheduling, which employs a dynamic program for scheduling decision with packet arrival deadline.

Because the aforementioned works mainly investigate the delay problem in single-hop wireless networks, their techniques cannot be easily applied to multi-hop networks, in which the delay problem is further complicated by the process at the intermediate nodes. The work of [18] extends the above works by proposing a new delay metric, which is the delay difference between a queue's head-of-line packet and the head-of-line packet of its previous hop. The throughput optimality is achieved by using a back-pressure scheduling in multi-hop wireless networks. The study of [6] investigates the impact of scheduling policies on delay performance in multi-hop wireless networks. Nevertheless, these works do not investigate how to satisfy the given delay requirements by adjusting flow rates.

In addition to the scheduling oriented delay solutions, algorithms based on rate allocation or even cross-layer resource management have also been proposed. The work of [4] investigates the delay-aware cross-layer algorithm of rate allocation, routing and scheduling. Especially, they propose delay-aware routing and back-pressure scheduling to jointly improve the end-to-end delay performance. However, this work does not provide any assurance of satisfying explicit delay requirements.

In great contrast to the existing works of delay control in wireless networks, one salient feature of our approach is that it only requires the knowledge of the derivative of link delay with respect to link capacity margin, which can be either derived analytically for known arrival processes (*e.g.*, Poisson process), or measured online over each link for unknown processes. Therefore it is applicable to a wide range of applications.

This work extends our preliminary works [19,20] with a local packet scheduling algorithm and a rate allocation algorithm that handles generalized traffic arrival models. Moreover, we present simulation study using a new packet-level simulator. We conduct extensive simulations with more realistic and sophisticated network settings to validate our theoretical claims.