Cross Layer based Resource Allocation in Wireless Ad-Hoc Networks

PhD Annual Progress Seminar

by

Hemant Kumar Rath
Roll No: 04407001

under the guidance of

Prof. Abhay Karandikar

Department of Electrical Engineering
Indian Institute of Technology, Bombay
Mumbai - 400076

August 2005

Acknowledgment

I would like to express my deep sense of gratitude to Prof. Abhay Karandikar for giving me an opportunity to work in this area and his motivation and fresh ideas he provides during the course of this research work. Also, I thank Prof. Anirudha Sahoo, Prof. S A Soman and my colleagues of Infonet Lab for their inputs during that period.

Hemant Kumar Rath
Roll No.: 04407001

Abstract

In this report we first discuss a cross layer congestion control technique of TCP Reno-2 in wireless networks, in which TCP layer and PHY layer jointly control congestion. The PHY layer changes transmission power as per the channel condition, whereas the TCP layer controls congestion using Reno-2 windowing flow control. We formulate congestion control mechanism of TCP Reno-2 in terms of control system equations with a feedback system model. We use transmission power of wireless nodes as a function of cost and derive a utility function. Using optimization techniques, we determine the maximum aggregate utility (and hence, throughput) of all the sources, subject to link capacity constraints and maximum transmission power of wireless nodes. Our simulations show that power control using the optimization technique provides performance improvement in terms of throughput and window size variations.

Further, we formulate a resource allocation scheme for wireless ad-hoc networks which takes care of energy cost of transmission of wireless nodes and congestion in the network. The aim of this problem formulation is to find out the optimal rate of transmission of a source-sink pair and optimal power transmission to maintain the transmission rates in an individual as well as social optimization scenario. Nodes are fed back by the links with the "network cost" (combination of energy cost and congestion cost) of the link in the process to compute the optimal transmission rates. This is also a cross layer resource allocation problem, where the PHY layer controls the power transmission and the transport layer controls the transmission rate jointly.

Contents

List of Figures

Chapter 1

Introduction

1.1 Introduction

An ad-hoc wireless network is a network of self-configurable wireless mobile nodes. The nodes are connected by wireless links. The communication between nodes is either single-hop or multi-hop in nature. The multi-hop networking improves the power efficiency of the network, as intermediate nodes help in relaying/forwarding traffic to the final destination. The ad-hoc[1] network is advantageous as it is easy to build the network and is cost effective in terms of deployment and maintenance. But, the ad-hoc networks are more complex to manage than the traditional wired networks. This is because of the time varying and broadcasting nature of wireless links. Also, as the number of hops in an ad-hoc networks increases, managing the ad-hoc network become more complex. The number of hops are also directly related to the scalability [1] of the network.

Both wired and wireless networks are meant for providing two different kind of broad services, viz. *Guaranteed Services* and *Best-effort Services*. The guaranteed services are characterized by some guarantee in terms of delay or/and bandwidth by the network, whereas the best-effort services are characterized by usual Internet traffic without any guarantee in terms of bandwidth or/and delay by the network. In this report, we are dealing with best-effort services in an ad-hoc network. In the best-effort services, since there is no guarantee, individual users[2] try to maximize their data rate in a competitive manner, without bothering others' requirement. This results in congestion in the network, with delay and packet loss. In wired network, the links are assumed to be reliable and of fixed capacities. So, if there is any packet loss or delay, then it is due to congestion in the network. There are various techniques, used in modern Internet to counter the congestion problem. They are either to avoid congestion (proactive) or to control congestion (reactive). These techniques are implemented in Transport Control Protocol (TCP) which is used as a reliable transport layer protocol in the wired network. The congestion control/avoidance techniques of TCP are proven to be effective in the wired network of the modern Internet.

[1] We used the term ad-hoc networks for wireless ad-hoc networks

[2] We used nodes and users alternately in our report and should not be get confused

But, this is not true in wireless networks. Packet loss or/and delay can result due to congestion as well as due to time varying nature of the wireless channel. Hence, applying the congestion control of wired networks in wireless networks may not be suitable [2]. Also, the direct mapping of layered nature of wired networks into wireless networks is not effective. This is because, the time varying nature of wireless channel necessitates that the Physical (PHY) and Media Access Control (MAC) layers should interact with the higher layers for effective use of resources like bandwidth, energy (battery power in wireless nodes) and buffer capacity. This interaction of layers requires a redesigning of layered network into a cross layered network.

Similarly to provide QoS based services, one needs to allocate his/her resources in an efficient and fair manner. This in turn call for redefining the allocation schemes of several QoS attributes like bandwidth, delay and packet loss etc. The allocation schemes are reasonably different for wired and wireless networks, because of the nature of the wireless networks, in which the channel is time varying and broadcasting in nature. Another important factor which needs to be explored is the co-operation in ad-hoc networks. This is very much required, because the nodes can't reach their destinations in one hop due to low transmission power and multiple access in the channel. So, QoS based resource allocation in wireless networks needs co-operation between nodes and may require admission control to maintain some outage probability of the network. Further, this scheme should be adaptive to the channel variations.

1.2 Our Approach

In this report, we discuss a cross layer congestion control technique involving TCP, MAC and PHY layer. TCP layer performs the windowing flow control and PHY layer varies transmission power of wireless nodes depending on the channel condition and interference. We have extended the approach of [3] for TCP Reno-2 and used maximum transmission power as a constraint in the optimization phase. Further, we have derived the utility function for Reno-2 in a wireless network and have simplified the cost function for wireless network for TCP Reno-2 by an $M/M/1/B$ queue model. The window based congestion control is modeled as a rate based optimization problem. Simulations are used to verify the cross layer congestion control technique for TCP Reno-2 in a wireless network. The details of this are discussed in Chapter 2.

Further we have modified the above problem with inclusion of minimum battery power consumption in a link called energy cost within the optimization framework. We have modeled a QoS based resource allocation scheme for wireless ad-hoc networks using the concept of "network costs". The details of this are discussed in Chapter 3. A similar approach based on iterative solution method can be found in [4]. The basic difference between their approach and our approach is that, ours is along the line of [3] and [5] and theirs is based on an iterative solution similar to that of [6]. The system in their case needs to know the utility function of all users before hand, whereas in our case, it is not required.

1.3 Congestion Control in Wired Networks

When the aggregate demand for resources (e.g., bandwidth) exceeds the capacity of the link, congestion occurs. Congestion is characterized by delay and loss of packets in delivery. In TCP, congestion is said to have occurred when the sender receives three duplicate acknowledgments (*dupacks*) or when a timeout (packet loss) occurs, resulting in wastage of resources. *Congestion Control* and *Congestion Avoidance* are two known solutions which address the above problem. In congestion control [7], system controls the network parameters after realizing congestion (reactive); whereas, in congestion avoidance, system controls the network parameters before congestion (proactive). The congestion control techniques are further divided into three broad types, viz., (i) window based, (ii) equation based, and (iii) rate based. These are discussed below.

1.3.1 Window Based Congestion Control

The congestion control algorithm followed by most of the variants of TCP is an end-to-end window based congestion control technique [8]. They follow an *Adaptive Window Management* technique in which increase and decrease of congestion window *cwnd* is based on packet drops and *dupacks*. The increase and decrease of *cwnd*s are based on the principle of Adaptive Increase and Multiplicative Decrease (AIMD). Different versions of TCP uses different methods to increase/decrease *cwnd* during congestion. These are discussed below.

TCP Tahoe, Reno, Reno-2 and Vegas

In TCP Tahoe [9], congestion is detected by timeouts only. On timeouts, it decreases *cwnd* size to one, known as *Fast Retransmit*. Then it increases *cwnd* size from one to *ssthresh* (which acts as limit point) in an exponential manner in each Round Trip Time (RTT). This phase is known as *Slow-Start* phase. After reaching *ssthresh*, it increases *cwnd* linearly in each RTT till next congestion occurs. This phase is known as *Congestion Avoidance* phase.

In TCP Reno [10], congestion is detected by triple *dupacks* and timeouts. On timeouts it works similar to TCP Tahoe. But, for triple *dupacks*, it follows *Fast Recovery* technique and decreases *cwnd* value by half of the current *cwnd* and then increases linearly until next congestion. Though it is better than TCP Tahoe for dealing with single packet loss and triple *dupacks*, the performance is not good when multiple packets are lost within one RTT. This problem is solved in its newer version called TCP Reno-2.

In TCP Reno-2 [11], *cwnd* value is not decremented for every packet loss, rather it is decremented in an intelligent manner: *cwnd* is decremented by half after RTT period if there is one or more packet loss (not for each packet loss in one RTT).

In TCP Vegas [12, 13], *cwnd* size is increased or decreased depending on the difference between the expected throughput (ratio of current window size and *baseRTT*, which is the minimum of all measured RTTs) and actual throughput (ratio of number of bytes transmitted during the last RTT, known as *rttLen* and averaged RTT of the last segment). It decreases the *cwnd* size before experiencing any congestion in the network in a more aggressive manner than Reno and Tahoe. TCP Vegas is proactive, whereas TCP Tahoe, Reno and Reno-2 are reactive.

1.3.2 Equation Based Congestion Control

Equation Based Congestion Control [14] is appropriate for real-time traffic, as TCP's congestion control using AIMD leads to a noticeable degradation of throughput. This is because, when congestion is experienced, TCP reduces its window size to half or one depending upon the type of TCP. Reducing to half or to one will not experience any quality difference in bulk-data transfer, but will result in a degradation of quality in streaming traffic. So, instead of reducing the window size to half or one, equation based congestion control adapts its sending rate based on the loss event rate. The main aim of equation based congestion control is to maintain a relatively steady sending rate, while still being responsive to congestion. It does not reduce the sending rate aggressively as TCP does. Also, it increases the sending rate in a slow rate, in response to a decrease in loss event rate.

1.3.3 Rate Based Congestion Control

The window based congestion control is unfair [15] for the connections with higher round-trip delays, when multiple connections share a link. This is the basic nature of TCP dynamics in a bottlenecked link, where the delay-bandwidth product is large. Analysis in [15] shows that the instantaneous throughput of a connection is inversely proportional to the square of round-trip delay of the connection. This bias is somewhat alleviated, if the delay-bandwidth product is lesser than the buffer in the link which is true in a high speed network with buffers. In that case, the instantaneous throughput or rate of transmission of an user can be approximated by $x_r = \frac{w_r}{\tau_r}$, where x_r is the instantaneous rate, w_r is the window size and τ_r is the RTT of user r at that instant. So, for a high speed network, the congestion control can be modeled as rate based instead of usual window based one and instead of changing the window size in each RTT, instantaneous rates can be regulated to avoid congestion.

The rate based congestion control scheme has been further investigated in [16]. In this the authors formulate the end-to-end congestion control as a global optimization problem of a class of minimum cost flow control (MCFC) algorithms for controlling the rate of a connection/session or window size. The window based congestion control of TCP is shown as a special case of MCFC algorithm in that paper.

1.4 Pricing in Congestion Control: Optimization Flow Control

What is Price?

Price is the amount that one node needs to pay, either to an operator or to any other node for transmitting its traffic. In real sense, individual node can either charge other nodes for forwarding their traffic or use that as a debit for future (forwarding its own traffic). In this report, we consider cost to be the price that one node needs to pay to other for using their nodes for relaying. Similarly, we consider incentive to be the price that one gets from other users for forwarding their traffic.

Optimization Flow Control

Modeling congestion control as a optimization flow control problem has been addressed in [5, 17, 6, 7]. The authors in these papers, have modeled a optimization based flow control in which the sources with different valuation of bandwidth reacts differently to congestion. They try to model the flow control problem as a constraint based optimization problem. The solution to this constraint based optimization problem gives the solution to congestion control or flow control in networks. The authors differ basically in defining the objective functions of the constraint based optimization problem and solution methodologies. [6] is similar to [5] in defining the objective functions, but differ in solving the objective function. In [6], the users decide their rates based on the charges received from the network, whereas in [5], the users decide the payments they can make to the network and receive the rates allocated by the network. We have discussed these two different approaches in this report. We name the two approaches as *Kelly's Model* [5] and *Low and Lapsley Model* [6] which are discussed below.

1.4.1 The Basic Model

The basic flow control model used in [5] and [6] is explained here.

We consider that a set I of sources i shares a network of L unidirectional links l. The capacity of the individual links l are c_l, where $l \in L$. Each source $i \in I$ has a utility function $U_i(x_i)$ associated with its transmission rate x_i. The utility function $U_i(x_i)$ is concave, increasing and double differentiable [18] for elastic[3] traffic. Let the route matrix R consists of all possible links and $R_{li} = 1$ if, source i uses link l and $R_{li} = 0$ otherwise.

The goal of both [5] and [6] is to find the source transmission rate x_i, which maximizes the aggregate utility of all sources subject to capacity constraints of the links. This corresponds to a system optimality and is expressed as:

[3]Elastic traffic consists of traffic for which users do not necessarily have an minimum requirements, but would get as much as data through to their respective destinations as quickly as possible [19]

$$
\begin{aligned}
&\max_{x_i \geq 0} \sum_i U_i(x_i), \\
\text{s. t.,} \quad &\sum_{i \in I(l)} R_{li} x_i \leq c_l;
\end{aligned} \tag{1.1}
$$

The solution of Eqn. (1.1) requires the knowledge of utility functions of all sources and coupling of sources through shared links. Both [5] and [6] solve Eqn. (1.1) in a de-centralized manner as discussed below.

1.4.2 Kelly's Model

Here the users choose the charges per unit time that they can pay and the network allocates the rates according to a proportional fairness criterion applied to the rate per unit charge. A system equilibrium is achieved when the users choice of charges and networks choice of allocated rates are in equilibrium. The system comprises both the users with utility functions and a network of links with capacity constraints. In Kelly's formulation Eqn. (1.1) is known as $SYSTEM(U, R, C)$ optimization. To solve this, it considers two simple subproblems as $NETWORK(R, C; W)$ and $USER(U_i(x_i), q_i)$. These two subproblems are expressed as follows.

$USER(U_i(x_i), q_i)$

$$
\max_{x_i \geq 0} U_i(x_i) - q_i x_i, \tag{1.2}
$$

where $q_i = \sum_{l \in L} R_{li} p_l$, is the charge per unit flow for user i. This is done by choosing w_i as the amount to pay per unit time and receiving x_i in return, proportional to w_i. So, $w_i = q_i x_i$.

$NETWORK(R, C; W)$

$$
\begin{aligned}
&\max_{X \geq 0} \sum_i w_i log x_i, \\
\text{s. t.,} \quad &RX \leq C; \\
X = \{x_i\} \quad \text{and} \quad &C = \{c_l\}
\end{aligned} \tag{1.3}
$$

Since there exists vectors [5] $q = \{q_i\}$, $w = \{w_i\}$ and $X = \{x_i\}$ satisfying $w_i = q_i x_i$ for $i \in I$, the optimum w_i solves $USER(U_i(x_i), q_i)$ and optimum X solves $NETWORK(R, C; W)$. So, X is the unique solution to $SYSTEM(U, R, C)$ (X maximizes Eqn. (1.1)).

1.4.3 Low and Lapsley Model

In this model, the authors have proposed an adaptive de-centralized scheme to find out the source rate x_i after eliminating the centrally required constraints of Eqn. (1.1). They solve a dual problem of Eqn. (1.1), which is similar to solving a distributive computing system with two processors as source and links. Each processor executes a local algorithm and communicates its computational result to other and repeats for some bounded number of cycles in an iterative manner. This solution is a reactive flow control technique, where each link $l \in L$ computes a price p_l for a unit bandwidth in l, based on all local source rate x_i on that link. A source $i \in I$ is fed-back the total price in its path $q_i = \sum_{l \in L} R_{li} p_l$. The individual profit maximization of a source i becomes $U_i(x_i) - q_i x_i$. Based on this, the source computes the transmission rate x_i such that it achieves maximum profit. All sources computes their transmission rates in an adaptive manner as per the network load in each iteration. Though, this is an iterative method of solving the individual optimization, it is bounded and converges to some optimal transmission rate ${x_i}^*$ for an optimal price ${p_l}^*$ of the links and maximizes Eqn. (1.1).

1.5 Modifications for Wireless Network

As discussed before, the congestion control of the Internet is well modeled within an utility based optimization framework. The fairness and stability of the model has been extensively studied in [5, 20, 21, 22] etc. These models are well suited to traditional wired network. Since the wireless network is mainly dictated by the time varying channel, a direct extension [2] of the congestion control of wired network to wireless network is not feasible [23]. In [2] the authors gave an counter example, where the direct extension is not feasible which was verified by simulations. So, there is a need of modification of the utility based optimization framework for congestion control to accommodate the time varying channel and interference of neighboring nodes. Also, in a multi-hop network, battery power should be accommodated in the optimization process, which is not addressed in any of these papers so far. Few approaches using cross layer model to control congestion in a wireless ad-hoc network are seen in the recent past. Explicit Congestion Notification (ECN) approach of wired network is also applied to wireless ad-hoc networks by [24]. We have discussed these issues in the sections below.

1.5.1 Joint Congestion and Power Control

Efforts were made by various authors to use the congestion control techniques of TCP in a wireless network. [3, 25, 26] have used power control along with congestion control in a wireless network and analyzed it for TCP Vegas. In [27], the author has modeled the joint power control and utility maximization of a wireless network as a sum product problem. This approach is used to design a new joint optimal congestion control and power control (JOCP) algorithm (for a wireless network). In JOCP, the authors have proposed a distributed power control algorithm which works jointly with the existing TCP congestion control algorithm to increase end-to-end throughput and energy efficiency of a multi-hop wireless network. This JOCP algorithm does not change the original TCP and the layered structure of the network. Rather, it is a cross-layer approach, where the interactions among the physical and transport layer is used to increase the end-to-end throughput.

1.5.2 Explicit Congestion Notification

Loss of packet in TCP is considered as an indication of congestion in the network. One can avoid congestion by indicating the users before congestion. This is achieved by the active queue management mechanisms which detect congestion before the buffer overflows, and provides an indication of this congestion to the users. Active queue management allows routers to use the Congestion Experienced (CE) code-point in a packet header as an indication of congestion, instead of relying solely on packet drops. This has the potential of reducing the impact of loss on delay-sensitive flows. This technique is called as Explicit Congestion Notification (ECN) [28] technique. ECN for wired network is a well established technique. There are few studies [24, 29], which exploit the ECN marking of wired network to control the random losses due to the time varying nature of wireless channel. But, these models do not consider losses due to fading and they do not distinguish the packet losses due to time varying nature of channel and due to congestion in the network. Someone can think along the line of ECN to distinguish between the wireless channel loss and congestional loss, and model a flow control problem of window size and hence transmission rate in wireless ad-hoc networks. This can improve the end-to-end throughput of wireless network.

1.6 QoS based Resource Allocation

QoS based resource allocation in wireless ad-hoc networks is attempted up by various authors in the recent past. [30] has discussed various issues of QoS in wireless ad-hoc networks. It has focused on QoS routing and issues related to ad-hoc networks. The main points it covers in it are the necessity of effective QoS based routing and management of channels etc., in wireless ad-hoc networks.

[31, 32] have used convex optimization techniques based on *Geometric Programming* to provide QoS based services in ad-hoc networks. In this paper, the authors convert a non convex and nonlinear constraint optimization problem into a nonlinear convex optimization problem using geometric programming techniques. A global optimal solution can be achieved by using geometric programming, in which a heuristic is used to compute and power control techniques are also used in the resource allocation problem. Though it considered delay as whole as a constraint, it did not consider queuing delay at nodes. Queuing delay is important in the design of multi-hop ad-hoc networks as it is the major factor contributing to the total delay.

Dynamic bandwidth management based single-hop ad-hoc networks have been discussed in [33]. The authors proposed an admission control and dynamic bandwidth scheme to provide fairness and a soft rate guarantee in the absence of distributed MAC layer weighted fair scheduling. The authors provide a mapping scheme to translate bandwidth requirement into a channel time requirements for real-time traffic. Their scheme is more suitable for single-hop multimedia transmissions. But, it is not suitable for multi-hop ad-hoc networks.

In [4], the authors have discussed another price-based bandwidth allocation scheme for ad-hoc networks. In this the authors model resource allocation problem through a set of optimization equations based on utility function. The solution method is an iterative one and needs the utility function of all users before hand. The iterative solution may not converge, if the utility functions are different for different users.

Another price based resource allocation scheme has been proposed in [34]. The proposed scheme in this is used to achieve optimal resource utilization and fairness among the users. Their pricing scheme is based on "maximal cliques" in wireless ad-hoc networks. The algorithm in their case is a two-tier iterative algorithm.

Chapter 2

Cross Layer Congestion Control for Reno-2

In this chapter, we discuss cross layer congestion control based on sum product algorithms of [27] for TCP Reno-2 in a wireless network. Reno-2 is more suitable for wireless network case. This is because, it is more robust than TCP Reno, for multiple packet loss in an window. In Reno-2, even if there is no congestion and multiple packet loss occurs in a window due to time varying nature of the channel, the window size ($cwnd$) reduces only to half and not further. So, intuitively packet loss due to wireless channel is also taken care of by the Reno-2 congestion control algorithm. We adopt the following steps to achieve a cross layer congestion control algorithm for TCP Reno-2 in this chapter.

1. Formulation of TCP congestion control in terms of control system equations.

2. Use of transmission power of wireless nodes as a function of cost in an optimization equation.

3. Use of optimization techniques to determine the maximum aggregate utility of sources, subject to capacity constraints and maximum transmission power [5] of wireless nodes.

2.1 Problem Formulation

We formulate our problem statement initially for a wired network, similar to [20] [21] and then extend it to a wireless network. Our system model is based on fluid-flow abstraction. We consider long TCP connections instead of short ones in our model, as the former generate most of the TCP traffic in the Internet.

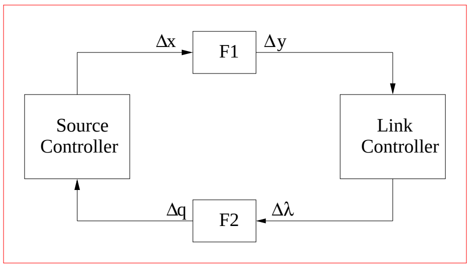

Figure 2.1: Feedback System Model

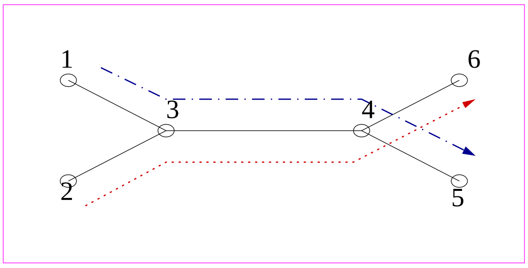

Figure 2.2: System Model/Topology

2.1.1 System Model

For analyzing TCP congestion, we consider a feedback system model consists of a Source Controller and a Link Controller[1] as shown in Figure 2.1. The network consists of N communicating nodes connected by L communicating links, similar to [21]. The communicating links are indexed by $l \in L$, and have finite capacity c_l packets/sec. These communicating links are shared by a set of source-sink pairs indexed by $i \in I$ (Figure 2.2). Now, we write the routing matrix R as:

$$R_{li} = \begin{cases} 1 & \text{if, source-sink pair } i \text{ uses link } l, \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

Each source-sink pair i has an associated transmission rate $x_i(t)$, i.e., the transmission rate of an individual TCP flow of a source-sink pair. The aggregate flow $y_l(t)$ (packets/sec) at each link is defined as:

$$y_l(t) = \sum_i R_{li} x_i(t - \tau_{li}^f), \tag{2.2}$$

where τ_{li}^f is the forward delay between the source and sink (between the Source and Link Controllers through F1 in Figure 2.1). The Source Controller controls the transmission rate x_i based on the change in price (due to change in aggregate losses Δq at F2 in Figure 2.1). The time scale of change of routing is higher than the time scale of our analysis, which is a multiple of RTT (τ_i). Hence, we assume stationary routing.

If each link l is associated with price $\lambda_l(t)$, as a congestion measure, then the aggregate price[2] of all links in the route is:

$$q_i(t) = \sum_l R_{li} \lambda_l(t - \tau_{li}^b), \tag{2.3}$$

where τ_{li}^b is the backward delay in the feedback path (of the acknowledgment packets between the link and the source) between the Link and Source Controllers through F2 in Figure 2.1. The Link Controller controls the individual link prices ($\lambda_l(t)$) based on the

[1]The details of the controllers and the derivations of various equations are described in [35]

[2]Sources are assumed to have an access to the aggregate price of all links in their route

change in aggregate traffic (∇y) in the link at F1 in Figure 2.1. We write the RTT as $\tau_i = \tau_{li}^f + \tau_{li}^b$. In vector form we write Eqn. (2.2) and Eqn. (2.3) as:

$$y = Rx; \quad q = R^T \lambda \tag{2.4}$$

2.1.2 Use of Control System Equations

We use control system equations [21] to determine the equilibrium point (x^*, y^*, q^*) for our system with the following assumptions: (i) resource allocation is characterized by a demand curve $x_i^* = f_i(q_i^*)$, (ii) aggregate flow is less than or equal to the capacity of the link, i.e., $y_l^* \leq c_l$, and (iii) the equilibrium queues are either empty or small enough to avoid the queuing delays.

The demand curve specified above is a decreasing function of price, which is equivalent to use of an utility function $U_i(x_i)$ defined by [18] which is concave and continuously differentiable function of x_i, in the range of $x_i \geq 0$, i.e. $f_i = (U_i^{'})^{-1}$ [18]. At equilibrium, a source chooses its maximum profit by choosing local parameters as:

$$\max_{x_i^*} \Bigg[U_i(x_i^*) - q_i^* x_i^* \Bigg], \tag{2.5}$$

where $q^* = R^T \lambda^*$, and x^*, y^*, q^* are the optimal fixed parameters. Since at equilibrium, each source tries to maximize its profit (individual optimality) by choosing an appropriate rate, the individual optimality of Eqn. (2.5) is re-written as a social optimality equation [36] as follows:

$$\begin{aligned} \max_{X \geq 0} \sum_i U_i(x_i), \\ \text{s. t.,} \quad RX \leq C; \\ X = \{x_i\} \quad \text{and} \quad C = \{c_l\} \end{aligned} \tag{2.6}$$

Using Karush-Kuhn-Tucker (KKT) condition [37] on the social optimal equation (Eqn. (2.6)) as in [20], the constrained global maxima point[3] $X^* = \{x_i^*\}$ is expressed as:

$$\nabla \sum_i U(x_i) - \sum_l \lambda_l \nabla (RX) = 0; \quad \forall \lambda_l > 0, \tag{2.7}$$

where λ_l are the Lagrangian Multipliers and are used as link prices (shadow prices) as discussed in [5] and [38]. Hence, Eqn. (2.7) is reduced to $U_i^{'}(x_i) = q_i^*$ at the global maximum point.

Different variants of TCP have different utility functions. The primal-dual distributed algorithm of the maximization equation Eqn. (2.6) signifies that the price is updated as a

[3]Since the utility function $U(x_i)$ is concave and increasing, KKT conditions are necessary and sufficient for a global maximum

congestion parameter and is a dual variable. We discuss the utility function of Reno-2 in the next section.

2.2 Utility Function for Reno-2

In TCP Reno-2, *cwnd* is decreased by half for one or more mark or incremented by one for no mark in one RTT. We assume the marking probability to be a measure of congestion [20].

If τ_i is the equilibrium RTT (assumed to be constant for all), then transmission rate of a source-sink pair $x_i(t)$ is expressed [16] as:

$$x_i(t) = \frac{w_i(t)}{\tau_i}, \tag{2.8}$$

where $w_i(t)$ is the window size and t, a period for our analysis for Reno-2 which is a multiple of few RTTs. The transmission rate $x_i(t)$ is interpreted as the average rate over t. If there is no mark in one RTT (τ_i), then, $w_i(t+\tau_i) = w_i(t) + 1$, else, $w_i(t+\tau_i) = w_i(t)/2$, for one or more mark in one RTT. Hence, we write the increase in window size is $1/\tau_i$ with probability $1 - \hat{q}_i(t)$ and the decrease is by $\frac{2w_i(t)}{3\tau_i}$[4] with probability $\hat{q}_i(t)$. $\hat{q}_i(t)$ is the end-to-end probability that at least one packet is marked out of all packets of an window, in a period t for the source-sink pair i (the packet marking are independent). We approximate [20] $\hat{q}_i(t)$ for small $q_i(t)$ (aggregate price of all links in the route Eqn. (2.3)) as follows:

$$\hat{q}_i(t) = 1 - (1 - q_i(t))^{w_i(t)} \approx w_i(t) q_i(t) \tag{2.9}$$

So, the average change in window size[5] in the period t is:

$$x_i(t) = \frac{1}{\tau_i}\Big(1 - \hat{q}_i(t)\Big) - \frac{2}{3}\frac{w_i(t)}{\tau_i}\hat{q}_i(t) \tag{2.10}$$

From Eqn. (2.9) and Eqn. (2.10), we determine the equilibrium parameters as follows:

$$\dot{x} = \frac{x_i(t)}{\tau_i} = \frac{1 - w_i(t)q_i(t)}{\tau_i^2} - \frac{2}{3\tau_i} q_i(t) x_i(t) w_i(t) \tag{2.11}$$

At equilibrium, setting $\dot{x} = 0$, $x_i(t) = x_i^*$ and $q_i(t) = q_i^*$ in Eqn. (2.11) we determine q_i^* as follows:

$$q_i^* = \frac{3}{x_i(t)\tau_i(3 + 2x_i(t)\tau_i)} \tag{2.12}$$

Now, using the fact that $U_i^{'}(x_i) = q_i^*$, Eqn. (2.7) the utility function is expressed as:

[4] For a single Reno-2 flow, the window oscillates between $\frac{4w_i(t)}{3\tau_i}$ and $\frac{2w_i(t)}{3\tau_i}$ with an average of $w_i(t)$ [20]

[5] The change in window size due to time-out is neglected here

$$U_i(x_i) = \int q_i^* dx_i = \frac{1}{\tau_i} \log\left[\frac{x_i\tau_i}{2x_i\tau_i + 3}\right] \tag{2.13}$$

Since the utility function of TCP Reno-2, derived above, is concave for $x_i, \tau_i \geq 0$, our problem formulation in Eqn. (2.5) holds good. The utility function derived here is for an average model of TCP as described in Eqn. (2.10).

2.3 Modification for Wireless Network Case

As discussed in [3], price is considered as a function of transmit power and by appropriate power control of all nodes the maximum aggregate throughput can be attained. However, the increase of power in one node has a direct effect on the data rates of other nodes due to interference. Hence, we modify the maximization equation (Eqn. (2.6) as follows:

$$\begin{aligned} &\max_{X\geq 0} \sum_i U_i(x_i), \\ \text{s.t.,} \quad &RX \leq C(P); \quad P = \{P_l\}, \\ &P_l \leq P_{l_{Max}}, \quad \forall l, \\ &P, X \geq 0, \end{aligned} \tag{2.14}$$

where P_l is the transmission power of a node in l^{th} link. Here, the link capacity c_l is a function of transmitting power P_l of nodes. Since, power control significantly increases throughput in both CDMA based and non-CDMA based wireless networks [39] [40], we use the power control mechanism to achieve congestion control by the above optimization techniques. This is a cross layer [41] mechanism and is a joint power and congestion control technique

2.4 Implementation

For implementation, we consider a CDMA-based wireless network. Let $SINR_l$ be the signal-to-interference-and-noise ratio of a time varying link $l \in L$.

$$SINR_l = \frac{P_l G_{ll}}{\sum_{k\neq l} P_k G_{lk} + n_l}, \tag{2.15}$$

where P_l is the transmission power on the link l, of the source, G_{ll} is the path gain from the transmitter of link l to the receiver of the link l and G_{lk} is the path gain from the transmitter on link k to the receiver on link l. n_l is the thermal noise on the link l. Using Shannon's capacity theorem, we determine the maximum capacity attainable in link l as:

$$c_l = \frac{1}{T} log(1 + M.SINR_l) \qquad \text{packets/sec}, \tag{2.16}$$

where T is the symbol period and M is a constant that depends on the modulation scheme used by the node for a successful transmission.

2.4.1 Implementation Framework

We follow the algorithms derived in [20] [3] for our implementation in Reno-2. We associate a Lagrangian Multiplier λ_l for the first constraint in Eqn. (2.14). Then, we solve Eqn. (2.14) using KKT [37] optimality conditions by solving the complementary slackness conditions at equilibrium. Then we determine the stationary points of the Lagrangian as:

$$\phi_{system}(X, P, \lambda) = \sum_i U_i(x_i) - \sum_l \lambda_l \big(\sum_i R_{li} x_i - c_l(P)\big) \tag{2.17}$$

Maximization of ϕ_{system} in Eqn. (2.17) is decomposed[6] as in [3]:

$$\begin{aligned} \max \quad I(X, \lambda) &= \sum_i U_i(x_i) - \sum_l \lambda_l \sum_i R_{li} x_i, \\ \max \quad I(P, \lambda) &= \sum_l \lambda_l c_l(P), \\ \text{s.t.}, \quad X \geq 0; &\quad 0 \leq P_l \leq P_{l_{Max}}, \end{aligned} \tag{2.18}$$

The first maximization equation involving $I(X, \lambda)$ in Eqn. (2.18) is a direct consequence of congestion control of TCP. It is solved by the congestion control mechanism of TCP by increasing/decreasing the window size (and hence the individual data rates) for each flow.

The second maximization equation involving $I(P, \lambda)$ in Eqn. (2.18) is solved by choosing appropriate transmission power of wireless nodes. Both $I(X, \lambda)$ and $I(P, \lambda)$ are related by a common variable λ, which plays a significant role in determining the equilibrium window size (hence the data rate) and transmission power. Any change in λ results in change in throughput and transmission power. Without loss of generality, we assume that the symbol period T and the modulation index M of Eqn. (2.16) as unity and re-write $I(P, \lambda)$ as follows:

$$I(P, \lambda) = \sum_l \lambda_l \log(SINR_l(P)) \tag{2.19}$$

$I(P, \lambda)$ is a concave function of a logarithmic transformed power vector. and is simplified as:

[6] Distributed solution is possible as along as there is an interaction between the two decomposed equations through some information passing (message passing in our case). This is known as sum product algorithm [27]

$$
\begin{aligned}
I(P,\lambda) &= \sum_{l} \lambda_l \log\left(\frac{P_l G_{ll}}{\sum_{k\neq l} P_k G_{lk} + n_l}\right) \\
&= \sum_{l} \lambda_l \left[\log(P_l) + \log(G_{ll}) - \log\left(\sum_{k\neq l} P_k G_{lk} + n_l\right)\right]
\end{aligned} \tag{2.20}
$$

Now, by differentiating $I(P,\lambda)$ with respect to P_l, we evaluate the l^{th} component of the gradient $\nabla I(P,\lambda)$.

$$
\nabla I(P,\lambda) = \frac{\lambda_l(t)}{P_l} - \sum_{j\neq l} \frac{\lambda_j(t) G_{jl}}{\sum_{k\neq j} P_k G_{jk} + n_j} \tag{2.21}
$$

Now, we use the Steepest Descent method as in [3] [37] to solve the maximization problem as:

$$
\begin{aligned}
P_l(t+1) &= P_l(t) + \Delta\Big(\nabla I(P,\lambda)\Big) \\
&= P_l(t) + \Delta\left(\frac{\lambda_l(t)}{P_l(t)} - \sum_{j\neq l} \frac{\lambda_j(t) G_{jl}}{\sum_{k\neq j} P_k(t) G_{jk} + n_j}\right) \\
&= P_l(t) + \Delta\frac{\lambda_l(t)}{P_l(t)} - \Delta\sum_{j\neq l} m_j(t) G_{lj},
\end{aligned} \tag{2.22}
$$

where Δ is a constant, called the step size in the direction of the gradient and $m_j(t)$ is the interference received from node j (similar to message passing in [3]) to the link l and is defined as:

$$
m_j(t) = \frac{\lambda_j(t) SINR_j(t)}{P_j(t) G_{jj}} \tag{2.23}
$$

From Eqn. (2.22), it is evident that the transmission power of a node in the next time slot $P_l(t+1)$ in a link l depends on three parameters, viz., (i) transmission power in the present time slot $P_l(t)$, (ii) shadow price $\lambda(t)$ and (iii) the weighted sum of interference received from all neighboring nodes. The third factor is responsible for decreasing the transmission power of the concerned node in the next time slot, i.e., the transmission power of the concerned node should be such that the interference caused by other nodes is below some threshold. This is known as the co-operation principle in power control of wireless network. $\lambda(t)$ is responsible for increasing power in the next time slot. Intuitively, more the shadow loss, more the congestion, more the transmission power in the next time slot, i.e., transmission power increases with respect to shadow loss. However, this increase is not linear.

In power control techniques, each wireless node needs to advertise its $SINR_l$ requirement either on a separate channel [40] or on the same channel [42]. These nodes update their path gains, noise levels, interference causes by other nodes etc., either after receiving the advertised signal or in periodic manner.

2.5 Shadow Prices in Reno-2

We consider the Lagrangian Multipliers λ_l as the shadow prices of a link. In this section, we derive a closed form expression for the shadow price of a link in Reno-2. The concept of pricing is different for different schemes of TCP, viz., queuing delay in TCP Vegas and loss probability in TCP Reno and Reno-2. We focus on the pricing scheme of TCP Reno-2.

When the aggregate traffic is more than the link capacity, packets are dropped. The probability of dropping in link is modeled by the probability of loss of packets in a $M/M/1/B$ queue [29], where, B is the buffer size at the link. In this kind of queuing model, y_l and c_l serves as average arrival rate and average service rate [7] respectively. The loss probability (p_l) for this kind of queue [43] at equilibrium is defined as:

$$p_l(c_l, y_l) = \frac{(1-\rho)\rho^B}{(1-\rho^{B+1})}, \tag{2.24}$$

where $\rho := \frac{y_l}{c_l}$. By scaling the arrival rate, departure rate and buffer capacity by a factor K as in a many-sources large-deviation scaling and by taking $K \to \infty$ we find out the loss probability as in [29]:

$$\begin{aligned} \lim_{K\to\infty} p_l(c_l, y_l) &= \lim_{K\to\infty} \frac{(1-\rho)\rho^{KB}}{1-\rho^{KB+1}} \\ &= \lim_{K\to\infty} \left[1 - \frac{1-\rho^{KB}}{1-\rho^{KB+1}}\right] \\ &= \frac{(y_l - c_l)^+}{y_l} = \frac{\max\big(0, (y_l - c_l)\big)}{y_l}, \end{aligned} \tag{2.25}$$

This is valid only for $y_l > 0$. When $y_l = 0$, $p_l(c_l, y_l) = 0$. Since the loss probability in TCP Reno-2 is considered as the price, the price λ_l is expressed as:

$$\lambda_l(t) = p_l(c_l, y_l) = \begin{cases} \frac{\max\big(0,(y_l(t)-c_l(t))\big)}{y_l(t)} & \text{if } y_l > 0, \\ 0 & \text{if } y_l = 0. \end{cases} \tag{2.26}$$

2.6 Cross Layer Congestion Control Algorithm

Our cross layer congestion control algorithm for Reno-2 is based on the joint power control of PHY layer and the congestion control of TCP layer. We discuss the cross layer congestion control algorithm for Reno-2 in Algorithm 1.

[7] The maximum capacity of the link is c_l and hence the service rate

Algorithm 1 : Cross Layer Congestion Control Algorithm for Reno-2

1: Set initial window size $w_i = w_{initial}$ (and hence transmission rate x_i)
2: Initialize $P_l = P_{l_{Min}}$
3: Advertise the minimum $SINR_l$ required
4: Update G_{lj} and G_{jj} periodically or after receiving the advertised signals
5: Determine maximum capacity of the link using Eqn. (2.16)
6: Determine $\lambda_l(t)$ using Eqn. (2.26)
7: Determine $m_j(t)$ using Eqn. (2.23)
8: Calculate $P_l(t+1)$ using Eqn. (2.22)
9: **if** $|P_l(t+1) - P_l(t)| \leq \delta$ **then**
10:     Continue transmission at $P_l(t)$
11: **else**
12:     Transmit at $\min(P_l(t+1), P_{l_{Max}})$
13: **end if**
14: Change w_i according to the congestion control algorithm of Reno-2 (which changes x_i according to Eqn. (2.8))
15: Update $SINR_l$ at each node Eqn. (2.15) and go to Step 3

In our algorithm, the initial window size $w_{initial}$, initial power $P_{l_{Min}}$, minimum $SINR_l$ and δ are taken as configuration parameters. We take $w_{initial} = 3$, $\delta = 0.5$ in our simulations. The value of $P_{l_{Min}}$ and $P_{l_{Max}}$ in our simulation is 3 and 15 respectively. The frequency of updating of $SINR_l$ is also a configuration parameter (usually this is a multiple of RTTs). We calculate the data rate x_i from Eqn. (2.8), whereas w_i and τ_i are decided using Reno-2 congestion control principle.

2.7 Simulation

We consider a simple topology with 6 wireless nodes and two pairs of TCP transmitters and receivers as shown in Figure 2.2 for our simulations. The two pairs of transmitters and receivers in our simulation are (1-5) and (2-6). All nodes in our simulation are TCP Reno-2 agents. Depending upon the shadow price of the links, we update the transmission power of the participating nodes. We use the TCP timeout timer as a multiple of four RTTs. We update RTTs using following exponential averaging technique.

$$RTT = \alpha RTT_{estimated} + (1-\alpha) RTT_{measured} \tag{2.27}$$

The value of α is taken as 0.85 for our simulation. Also, for simplicity, we assume that the time required for transmission in each of the segments 1-3, 2-3, 3-4, 4-5 and 4-6 is same and both forward and reverse channel characteristics are same. The channel gains are randomly chosen with a log-normal shadowing of $\sigma = 8dB$. The path loss factor γ is assumed to be 4. We use Matlab software [44] for our simulations.

2.8 Results

We simulate TCP Reno-2 congestion control mechanism using both with power control based on the $SINR$ values and without power control techniques. In the latter case, the transmission power of nodes are fixed at some maximum value. But, in the former case, depending upon the channel condition and interference, the nodes transmit at some optimal power level. Figure 2.3 shows the *cwnd* variation of joint power and congestion control mechanism, whereas Figure 2.4 shows *cwnd* variation without power control. We observe that the fluctuations in *cwnd* with power control mechanism is lower as compared to the fluctuations in *cwnd* without power control mechanism. Also, the average window size of joint power and congestion control scheme is larger than that of congestion control without power control. Hence, power control provides stabilized and better throughput. Intuitively, this occurs because the maximization of utility function with power control (Eqn. (2.14)) is done over a larger set of constraints than without power control (Eqn. (2.6)).

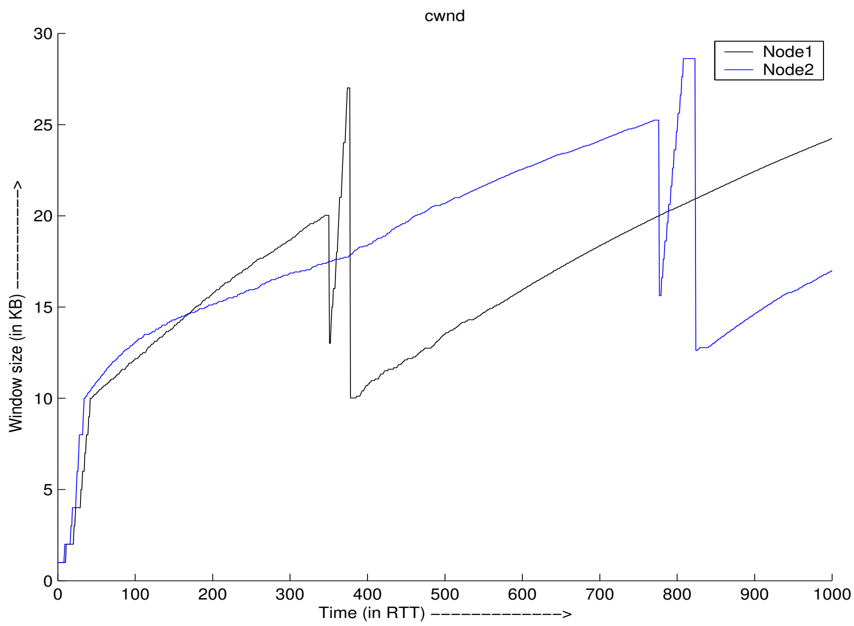

Figure 2.3: Variation of Congestion window - with Power Control

The transmission power of all Reno-2 agents are shown in Figure 2.7 (with power control). We observe that the transmission power is a function of packet loss and hence is a function of $SINR$. The power consumption of nodes in our cross layer scheme are lower as compared to the fixed power scheme. Further, we analyze the pricing mechanism for both fixed and power control schemes in Figure 2.5 and 2.6. The price in Reno-2 is a function of packet loss and hence is a function of congestion window. Price rises at the point of congestion (e.g., in Figure 2.4 and Figure 2.6 at $t = 220, 250, 400, 550, 650, 900$ and 950, *cwnd* is at the peak and hence the price is also at the peak, whereas immediately after

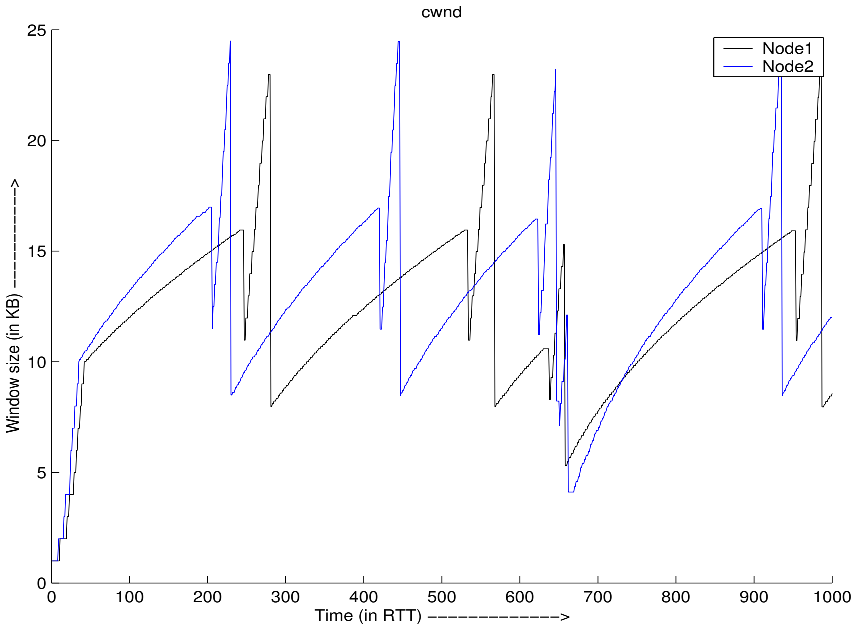

Figure 2.4: Variation of Congestion window - without Power Control

the peaks, the *cwnd* decreases by half of previous *cwnd* and hence the price also decreases) and falls after congestion control (by decreasing *cwnd*).

2.9 Convergence Analysis

In TCP Reno-2, we consider the probability of marking to be a measure of congestion. Indirectly this is controlled by the power control in the wireless channel as the marking probability is used as the common variable between the the two separate optimization problems of (Eqn. (2.18)). Since the transmitted power is bounded by minimum and maximum power levels, and the shadow price is zero at equilibrium state (total incoming traffic and capacity of a link are same, resulting no packet loss) a complimentary slackness condition between the primal variable (x_i) and dual variable (λ_l) of Eqn. (2.18) can be achieved. Now, going along the line of the proof given in [26] (see Theorem 1 of [26]) one can prove that the cross layer congestion control algorithm derived from Eqn. (2.18) will converge to a global optimum point. This requires the minimum $SINR$ should be greater than one, else, $I(P, \lambda) = \sum_l \lambda_l \log(SINR_l(P))$ (Eqn. (2.19)) can not be approximated from the second optimization equation of (Eqn. (2.18)). Also, it is clear from [26], that $I(P, \lambda)$ is a strictly concave function of logarithm of power transmission vector. Hence, the Lagrangian Multiplier λ_l facilitates a global maximization of Eqn. (2.17) and ensures convergence towards that. The step size (Δ) of Steepest Descent method of optimization in Eqn. (2.22) decides the rate of converges towards a global optimum point.

We perform simulations with three different values of Δ (0.1, 0.2 and 0.5) and observe the number of iterations to converge to an optimum power transmission value in all three cases at different $SINR$ level. For, $\Delta = 0.1$, it takes about 150 iterations to converge, (but, converges to a stable transmission power level), which becomes constant over time for a particular $SINR$. For higher value of Δ though it takes fewer number of iterations to converge initially, but does not converge to a stable transmission power level. From simulations we have observed that our algorithm converges even for fading in wireless channel. This shows the robustness of our algorithm. Intuitively, one can establish the fact that, though the message passing can be erroneous due to fading, the convergence of cross layer algorithm is achieved and hence is robust.

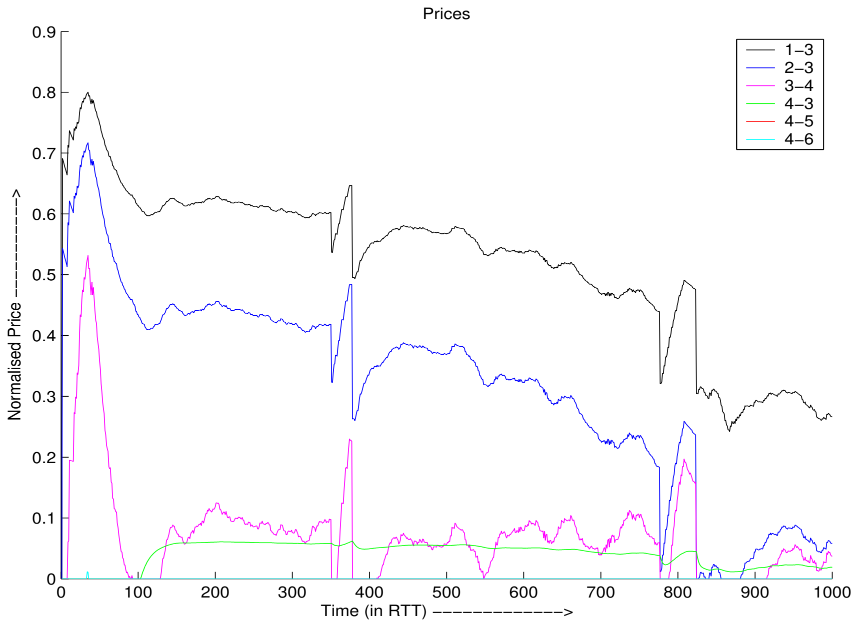

Figure 2.5: Variation of Price - with Power Control

2.10 Discussion and Conclusion

Our cross layer congestion control technique for Reno-2 converges very fast. This is due to the small step size ($\Delta = 0.1$) of Steepest Descent method Eqn. (2.22), and fixed maximum and minimum transmitting powers for each node. Our simulations verify the theoretical models we have discussed, which is a maximization problem of an utility function. As expected, the cross layer congestion control technique provides stabilized throughput at low power transmission.

But, if the channel conditions are very bad, then there would be more losses due to bad channel resulting a significant increase in λ, which in turn results in an increase in power transmission. In that case, our power control algorithm does not converge. This is a drawback of the joint power and congestion control algorithm. This algorithm holds good

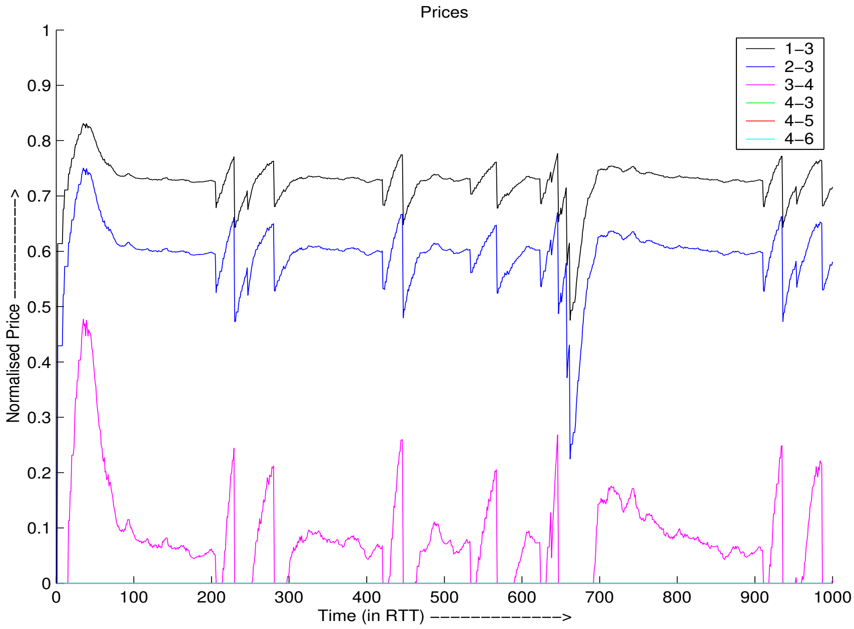

Figure 2.6: Variation of Price - without Power Control

as long as the minimum $SINR$ is maintained at the nodes.

We have considered a simple topology for our simulation. A complex topology can be used to study other issues. Also, use of joint power and congestion control algorithm in bad channel condition needs some modification in the definition of packet loss and congestion. This modification can significantly increase the throughput.

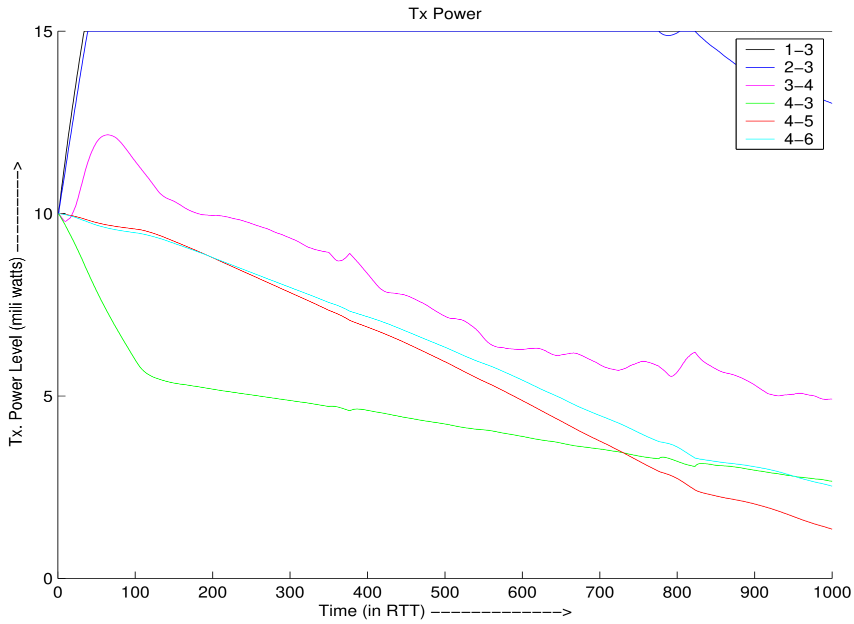

Figure 2.7: Transmission Power - with Power Control

Chapter 3

QoS based Resources Allocation in Wireless Ad hoc Networks

In this chapter, we model a QoS based resource allocation scheme as an extension of our work in the previous chapter. We suggest a joint power and resource control algorithm. We assume co-operation among users for sending traffic and our source and destination pairs are multi-hop apart. Our approach in this chapter is as follows:

1. Pricing based resource allocation by maximizing individual and social optimality condition.
2. Solving social optimality condition with constraints on signal to interference and noise ratio ($SINR$).
3. Transmission power $SINR$ as a function of both energy and congestion cost.

3.1 Model

We assume a multi-hop ad-hoc network, where the nodes are placed uniformly over a 2-d area. For simplicity, we assume the stationarity of the nodes. These nodes are identical in nature. The destination of nodes are uniformly placed over the network. A snap of this model is shown in Figure 3.1. Any node in our model can send its own traffic and can forward/relay others' traffic. All traffic are elastic in nature. We have introduced a new term called energy price along with the congestion cost/price in this chapter. Energy price is the price that is spent by the nodes (both relaying and relayed nodes) for maintaining a minimum $SINR$ in the link. The capacity of a link is a function of power transmission in that link. We discuss both energy price and congestion price in detail in the following sections. The notations in this chapter are same as those used previous chapter except some few new terms.

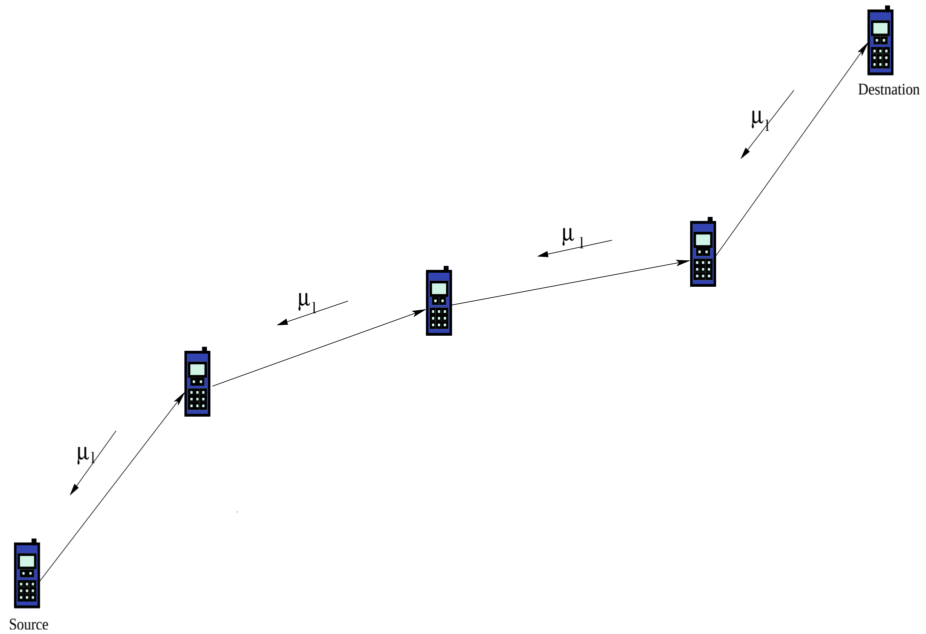

Figure 3.1: A Snap of our Ad hoc Network

3.2 Profit Maximization: Individual Optimization

3.2.1 Objective Function

Each source-sink pair i tries to find-out the maximum data rate x_i, such that it maximizes its individual profit. This is expressed as:

$$\begin{aligned} &\max_{x_i}\Big[U_i(x_i) - \sum_l R_{li}\lambda_{li}x_i - \sum_l R_{li}b_{li}x_i\Big] \\ &\qquad \Leftrightarrow \max_{x_i}\Big[U_i(x_i) - q_i x_i\Big] \\ &\qquad \text{s. t.} \quad \sum_i R_{li}x_i \leq c_l(P) \\ &\text{where} \quad q_i = \sum_l R_{li}(\lambda_l + b_l) = \sum_l R_{li}\mu_l \end{aligned} \tag{3.1}$$

The new term b_l, that we have introduced in the above equation is the energy cost of the link. One can notice that the above equation differs from the previous individual optimal equation (Eqn. (2.5)) only in b_l. In the earlier case, node was fed back by congestion cost λ_l only, whereas in this model it is fed back by "network cost" μ_l, which is a sum of both congestion and energy cost.

3.3 Congestion Cost

We use congestion cost λ_l similar to the shadow price discussed in Section 2.5. For completeness, the expression for congestion cost is expressed below:

$$\lambda_l(t) = p_l(c_l, y_l) = \begin{cases} \frac{\max\left(0,(y_l(t)-c_l(t))\right)}{y_l(t)} & \text{if } y_l > 0, \\ 0 & \text{if } y_l = 0. \end{cases} \tag{3.2}$$

where, y_l is the total aggregate flow in a link as in Eqn. (2.2). We assume a small probability of packet drop, so that the end-to-end probability of packet drop and hence the congestion cost is approximated as:

$$1 - \prod_{l \in L}(1 - R_{li}\lambda_l) \approx \sum_{l \in L} R_{li}\lambda_l \tag{3.3}$$

3.4 Energy Cost

We consider the energy cost b_l as a function of amount of energy transmitted for successfully delivering the packet at the receiving node. This is a function of the transmission power and time. We consider linear pricing and non-linear pricing model to describe the energy cost of a link. These are discussed below.

3.4.1 Linear Pricing Model

In this model, energy cost b_l is considered as a linear function of amount of energy transmitted for successfully delivering the packet at the receiving node. Though it is a simpler one, but it is not practical, as there is a minimum amount of power required for maintaining successful transmission. Also the power transmission of nodes are limited by minimum and maximum power limit. We do not use this pricing model in our problem formulation.

3.4.2 Non-Linear Pricing Model

Energy cost b_l in this model is a function of power transmission by the node, but the function in this case is a non-linear one. This is expressed as follows:

$$b_l(t) = \theta + k_1 log(k_2 P_l(t)), \tag{3.4}$$

where θ is the minimum price that needs to be paid (equivalent to saying minimum energy consumption) and k_1 and k_2 are constants, which can be chosen for a pricing model by the network operator.

3.5 Utility Function: Proportionally Fairness

A set $X = \{x_i\}, i \in I$ is said to be proportionally fair [36], if it feasible, and if for any other feasible set $X^* = \{x^*{}_i\}$, then the aggregate of proportional changes is zero or negative: $\sum_i \frac{x^*{}_i - x_i}{x_i} \leq 0$. As discussed in [5], the utility function for a proportional fair system is $U_i(x_i) = \log x_i$, which is strictly concave and double differentiable for $x_i \geq 0$ satisfying the requirement to achieve a global maxima. The nature of the utility function is known to the network. The utility functions of individual users may differ from other.

3.6 Social Optimization

Since, all users try to maximize their individual utility, a social optimality (which is also known as system optimality) as in [36] can be achieved for an optimum transmission rate vector and price vector. The social optimality function for the individual optimal equation described in Eqn. (3.1) is as follows: [1]

$$\begin{aligned} \max_{X \geq 0} \sum_i U_i(x_i); \\ \text{s. t.} \quad RX \leq C(P); \\ \text{where,} \quad X = \{x_i\}, \text{and} \quad C(P) = \{c_l(P)\} \end{aligned} \tag{3.5}$$

Using Karush-Kuhn-Tucker (KKT) condition [37] on the social optimal equation as in [20], if there exists a Lagrangian Multiplier μ_l, then a constrained global maxima can be achieved.[2] The KKT condition for the above equation is as follows:

$$\nabla \sum_i U_i(x_i) - \sum_l \mu_l \nabla (RX) = 0; \quad \forall \mu_l > 0 \tag{3.6}$$

The Lagrangian Multiplier μ_l is the "network cost" in our discussion and is a combination of congestion cost and energy cost ($\mu_l = \lambda_l + b_l$).

3.6.1 Optimal Solution

The solution to the social optimal equation described in Eqn. (3.5) is similar to the solution described in Section 2.4.1. Using KKT [37] optimality conditions to solve the complementary slackness conditions at equilibrium, we determine the stationary points of the Lagrangian as:

[1] It can be proved that if the social optimal condition satisfy the Eqn. (3.5) then, individual optimality is achieved in a selfless manner

[2] Since the utility function $U(x_i)$ is concave and increasing, KKT conditions are necessary and sufficient for a global maximum

$$\begin{aligned}\phi_{system}(X,P,\mu) &= \phi_{system}(X,P,\lambda,b)\\ &= \sum_i U_i(x_i) - \sum_l \mu_l \Big(\sum_i R_{li}x_i - c_l(P)\Big)\\ &= \sum_i U_i(x_i) - \sum_l \mu_l \sum_i R_{li}x_i + \sum_l \mu_l c_l(P)\end{aligned} \tag{3.7}$$

To maximize $\phi_{system}(X,P,\mu)$ defined above, we decompose Eqn. (3.7) by using the sum product algorithm [27] as follows:

$$\begin{aligned} \max \quad I(X,\mu) = \sum_i U_i(x_i) - \sum_l \mu_l \sum_i R_{li}x_i,\\ \max \quad I(P,\mu) = \sum_l \mu_l c_l(P),\\ \text{s.t.,} \quad X \geq 0; \quad 0 \leq P_l \leq P_{l_{Max}}, \end{aligned} \tag{3.8}$$

Each node will solve the first maximization equation involving $I(X,\mu)$ to find x_i, as it knows its own utility function $U_i(x_i)$ and the "network price" is fed back by the link. We solve the second maximization equation involving $I(P,\mu)$ by choosing appropriate transmission power of wireless nodes as discussed below. Both $I(X,\mu)$ and $I(P,\mu)$ will converge to a global maxima, as $I(X,\mu)$ is concave (since $U_i(x_i)$ is concave) and $I(P,\mu)$ is also concave function of transmission power. A formal proof of this (concavity and global convergence to a maxima) is not discussed in this chapter.

Now taking similar approximation as we did for $SINR$ in Section 2.4.1, capacity of a link c_l is written as $c_l = log(SINR_l(P))$. Hence, we write $I(P,\mu)$ as follows:

$$\begin{aligned} I(P,\mu) &= \sum_l \mu_l \log(SINR)_l(P)\\ &= \sum_l (\lambda_l + b_l)\log(SINR)_l(P) \end{aligned} \tag{3.9}$$

Now differentiating $I(P,\mu)$ with respect to P_l, we evaluate the l^{th} component of the gradient $\nabla I(P,\mu)$ as (taking $b_l = \theta + k_1 log(k_2 P_l(t))$)

$$\nabla I(P,\mu) = \frac{k_1}{P_l}(\log(SINR_l)) + \frac{\lambda_l}{P_l} - \sum_{j\neq l} \frac{\lambda_j G_{jl}}{\sum_{k\neq j} P_k G_{jk} + n_j} \tag{3.10}$$

Now, we use the Steepest Descent method as in [3] [37] to solve the maximization problem as:

$$
\begin{aligned}
P_l(t+1) &= P_l(t) + \Delta\Big(\nabla I(P,\mu)\Big) \\
&= P_l(t) + \Delta\frac{k_1}{P_l(t)}(\log(SINR_l(t))) + \Delta\left(\frac{\lambda_l(t)}{P_l(t)} - \sum_{j\neq l}\frac{\lambda_j(t)G_{jl}}{\sum_{k\neq j}P_k(t)G_{jk}+n_j}\right) \\
&= P_l(t) + \Delta\frac{k_1}{P_l(t)}(\log(SINR_l(t))) + \Delta\frac{\lambda_l(t)}{P_l(t)} - \Delta\sum_{j\neq l}m_j(t)G_{lj},
\end{aligned}
\tag{3.11}
$$

The above power transmission equation is different from that derived in Eqn. (2.22) by the term $(\Delta\frac{k_1}{P_l(t)}(\log(SINR_l(t))))$ only. Power transmission update in both equations require message passing, which is the key component of sum product algorithm. After getting the optimum power update as above, and after getting the price update in Eqn. (3.2) and Eqn. (3.4) (the optimal network cost μ_l is the dual variable), we solve $I(X,\mu)$ as follows:

$$
\begin{aligned}
\max \quad I(X,\mu) &= \max\sum_i U_i(x_i) - \sum_l \mu_l \sum_i R_{li}x_i \\
&= \max\sum_i U_i(x_i) - \sum_i\Big[\sum_l R_{li}\mu_l x_i\Big] \\
&= \max\sum_i U_i(x_i) - \sum_i q_i x_i
\end{aligned}
\tag{3.12}
$$

For a feasible optimal dual μ_l, we find a optimal primal x_i by maximizing the above equation as follows [20]:

$$
\begin{aligned}
\max \quad I(X,\mu) &= \max\sum_i U_i(x_i) - \sum_i q_i x_i \\
&= \sum_i \max\left(U_i(x_i) - q_i x_i\right)
\end{aligned}
\tag{3.13}
$$

This is equivalent to maximizing the individual optimality of each user. Since each user knows their individual utility function and the optimal dual variable "network cost", they can compute the transmission rate and transmit at the required power level P_l (fed back by the link), such that the individual profit is maximized. This is a joint power control and rate control technique for resource allocation in ad-hoc networks. We suggest an algorithm to implement this joint power control and rate control in Algorithm 2.

3.7 Discussion

The resource allocation based on "network cost" is a cross layer approach involving PHY and transport layer. The rate of transmission of this joint power and rate control algorithm

Algorithm 2 : Cross Layer Resource Allocation Algorithm

1: Set initial $x_i(t)$, $c_l(t)$ and $P_l(t) = P_{l_{Min}}$
2: Compute network cost $\mu_l(t)$ at the link ($\lambda_l(t)$ and $b_l(t)$) based on Eqn. (3.2) and Eqn. (3.4)
3: Determine $P_l(t+1)$ using Eqn. (3.11)
4: **if** $|P_l(t+1) - P_l(t)| \leq \delta$ **then**
5:     Continue transmission at $P_l(t)$
6: **else**
7:     Transmit at $\min(P_l(t+1), P_{l_{Max}})$
8: **end if**
9: Determine $b_l(t+1)$ using Eqn. (3.4) for the above $P_l(t+1)$ and $c_l(t+1)$ using Eqn. (2.16)
10: Compute $\lambda_l(t+1)$ using Eqn. (3.4)
11: Calculate the network cost $\mu_l(t)$ and hence $q_i(t)$ for the source sink pair i
12: Compute x_i for the optimum q_i computed above
13: Update $SINR_l(t+1)$ at each node and go to Step 3

can be mapped into a new TCP congestion control protocol, where along with packet drop due to congestion, energy cost will also be considered to determine *cwnd*. Fading in wireless channel can also be modeled in packet loss calculation and hence can be included in the new TCP congestion control scheme.

Chapter 4

Discussions and Future Work

4.1 The Cross Layer Approach

As discussed before, the layered architecture is good for wired network, whereas a cross layer approach can be modeled for wireless network, such that the performance can be increased. In a cross layer approach [19], the layers like PHY, MAC and transport can share the information among themselves to improve the system performance. Cross layer approaches like joint power and congestion control algorithms have already proved better performance than layered approach in a wireless network. Similarly, a power aware routing can be used [45], in which PHY and network layer can interact to find the minimum cost path in a wireless network. We will extend our QoS based resource allocation scheme in this direction in the future. Similarly, fading of wireless channel can be incorporated into a cross layer congestion control scheme, where the congestion control algorithm can distinguish packet loss due to congestion and due to bad channel condition and can take appropriate remedies to achieve better performance. Various authors like [46, 47, 48] have tried to model a correct framework to achieve this, but it is still in its infancy. In [46] the authors have proposed a cross layer approach in which the channel information is shared to upper layers, such that the upper layers can decide whether to transmit or not. The upper layer should transmit in “good” channel condition only, whereas it should stop transmission in “bad” channel condition. By doing this the mobile node can save power and can improve performance. In [49], the authors have modeled a 2-state Markov Model for the time varying channel. In this the authors have suggested to halt transmission, if the channel is bad. This approach may create problem in multi-hop network case, as temporary halting of transmission of packets can results in timeouts of TCP timers, resulting performance degradation. Such issues should be included in the cross layer approach.

4.2 Cautionary Note

Though there should not be any direct extension of the layered [2] architecture of wired network to wireless network, it is still optimal [50] to some order to use the layered network in wireless network. The cross layered architecture are mainly aimed for performance improvements. But, by doing so, we may loose the flexibility and robustness of layered architecture. Though the cross layer approach provides optimal solution and higher throughput, but in long term, the flexibility of designing a protocol at any layer is gone. So, any changes in any protocol in any layer will require to redesign all layers. In [50] the authors have critically examined the stability, robustness, system performance etc. for some recent developments in cross layer approach on adaptive rate MAC protocol and nested adaption of transmit power. The authors have issued a cautionary note on any cross-layer design, such that the modular nature and long-term benefit of layered architecture is guaranteed.

4.3 Future Work

1. Power Aware Routing (PAR) for wireless ad-hoc networks: We will model a cross layer design involving network layer and PHY layer to find a minimum cost path (in terms of congestion cost and energy cost).

2. Design of new TCP congestion control algorithm for wireless ad-hoc networks: The new TCP congestion control algorithm, we will model, will include fading of wireless links and will distinguish packet loss due to fading and congestion. This can be an Active Cost Management (ACM) technique for wireless ad-hoc networks.

3. Extension of ECN into wireless ad-hoc network.

Bibliography

[1] B.-J. Kwak, N.-O. Song, and L. E. Miller, "On the Scalability of Ad hoc Networks," *IEEE Communications Letters*, vol. 8, pp. 503–505, August 2004.

[2] V. Raghunathan and P. R. Kumar, "A Counterexample in Congestion Control of Wireless Networks." Submitted to the 8th ACM/IEEE International Symposium on Modeling (MSWIM 2005), http://black.csl.uiuc.edu/∼prkumar/ps_files/05-06-10-counterexample.pdf.

[3] M. Chiang and R. Man, "Jointly Optimal Congestion Control and Power Control in Wireless Multi-hop Networks," in *Proc. of GLOBECOMM*, December 2003.

[4] Y. Qiu and P. Marbach, "Bandwidth Allocation in Ad-hoc Networks: A Price-Based Approach," in *Proc. of IEEE INFOCOM*, March 2003.

[5] F. P. Kelly, "Charging and Rate Control for Elastic Traffic," *European Transactions on Telecommunications*, vol. 8, pp. 33–37, January 1997. (Corrected version of the original paper is available at http://www.statslab.cam.ac.uk/∼frank/elastic.ps).

[6] S. H. Low and D. E. Lapsley, "Optimization Flow Control-I: Basic Algorithm and Convergence," *IEEE/ACM Transactions on Networking*, vol. 7, November 1999.

[7] S. Ryu, C. Rump, and C. Qiao, "Advances in Internet Congestion Control," *IEEE Communications Surveys and Tutorials*, vol. 3, pp. 28–39, 2003.

[8] M. Allman, V. Paxson, and W. Stevens, "TCP Congestion Control," April 1999. RFC-2581.

[9] V. Jacobson, "Congestion Avoidance and Control," in *ACM SIGCOMM*, pp. 314–329, August 1988.

[10] R. Braden, V. Jacobson, and L. Zhang, "TCP Extensions for Highspeed Paths," vol. RFC-1185, October 1990.

[11] M. Mathis, J. Mahdavi, S. Floyd, and A. Romanow, "TCP Selective Acknowledgment Options," April 1996. RFC-2018.

[12] L. S. Brakmo, S. W. O'Malley, and L. L. Peterson, "TCP Vegas: New Techniques for Congestion Detection and Avoidance," in *ACM SIGCOMM*, pp. 24–35, 1994.

[13] U. Hengartner, J. Bolliger, and T. Gross, "TCP Vegas Revisited," in *Proc. of IEEE INFOCOM*, pp. 1546–1555, March 2000.

[14] S. Floyd, M. Handley, J. Padhye, and J. Widmer, "Equation Based Congestion Control for Unicast Applications: the Extended Version," in *ACM SIGCOMM*, March 2000.

[15] T. V. Lakshman and U. Madhow, "The Performance of TCP/IP for Networks with High Bandwidth-Delay Products and Random Loss," *IEEE/ACM Transactions on Networking*, vol. 5, pp. 336–350, June 1997.

[16] S. J. Golestani and S. Bhattacharyya, "A Class of End-to-End Congestion Control Algorithms for the Internet," in *6th International Conference on Network Protocols (ICNP)*, October 1998.

[17] R. J. Gibbens and F. P. Kelly, "Resource Pricing and Evolution of Congestion Control," *Automatica*, vol. 35, December 1999.

[18] S. Shenker, "Fundamental Design Issues for the Future Internet," *IEEE Journal on Selected Areas in Communication*, vol. 13, September 1995.

[19] S. Shakkottai, T. S. Rappaport, and P. C. Karlsson, "Cross-Layer Design for Wireless Networks," *IEEE Communications Magazine*, pp. 74–80, October 2003.

[20] S. H. Low, "A Duality Model of TCP and Queue Management Algorithms," *IEEE/ACM Transactions on Networking*, vol. 4, pp. 526–536, August 2003.

[21] S. H. Low, F. Paganini, and J. C. Doyle, "Internet Congestion Control," *IEEE Control Systems Magazine*, pp. 28–43, February 2002.

[22] F. P. Kelly, "Fairness and Stability of End-to-End Congestion Control," *European Journal of Control*, vol. 9, pp. 159–176, 2003.

[23] Y. Tian, K. Xu, and N. Ansari, "TCP in Wireless Environments: Problems and Solutions," *IEEE Radio Communications*, pp. S27–S32, March 2005.

[24] S. Kunniyur and R. Srikant, "A Time-Scale Decomposition Approach to Adaptive Explicit Congestion Notification (ECN) Marking," *IEEE Transactions on Automatic Control*, vol. 47, June 2002.

[25] M. Chiang, "To Layer or Not To Layer: Balancing Transport and Physical Layers in Wireless Multihop Networks," in *Proc. of IEEE INFOCOM*, March 2004.

[26] M. Chiang, "Balancing Transport and Physical Layers in Wireless Multi-hop Networks: Jointly Optimal Congestional Control and Power Control," *IEEE Journal on Selected Areas in Communication*, vol. 23, pp. 104–116, January 2005.

[27] M. Chiang, "Distributed Network Control through Sum Product Algorithm on Graphs," in *Proc. of GLOBECOMM*, pp. 2395–2399, November 2002.

[28] K. Ramakrishnan, S. Floyd, and D. Black, "The Addition of Explicit Congestion Notification (ECN) to IP," September 2001. RFC-3168.

[29] S. Kunniyur and R. Srikant, "End-to-End Congestion Control Schemes: Utility Functions, Random Losses and ECN Marks," *IEEE/ACM Transactions on Networking*, vol. 11, October 2003.

[30] S. Chkrabarti and A. Mishra, "QoS Issues in Ad hoc Wireless Networks," *IEEE Communication Magazine*, pp. 142–148, February 2001.

[31] D. Julian, M. Chiang, D. O'Neill, and S. Boyd, "QoS and Fairness Constrained Convex Optimization of Resource Allocation for Wireless Cellular and Ad hoc Networks," in *Proc. of IEEE INFOCOM*, June 2002.

[32] D. N. Mung Chiang, Chee Tan and D. Julian, "Power Control By Geometric Programming," (http://www.princeton.edu/∼chiangm/powergp.pdf, Last visited on 18th March, 2005).

[33] K. C. Samarth H. Shah and K. Nahrstedt, "Dynamic Bandwith Management in Single-Hop Ad-hoc Wireless Networks," *Mobile Networks and Applications*, vol. 10, pp. 199–217, February 2005.

[34] K. N. Yuan Xue, Baochun Li, "Optimal Resource Allocation in Wireless Ad-hoc Networks: A Price-based Approach," (www.eecg.toronto.edu/ bli/papers/xue-tmc.pdf, Last visited on 18th March, 2005).

[35] H. K. Rath and A. Sahoo, "Congestion Control in Wireless Networks." Technical Report, IIT-Bombay, Mumbai, January 2005, (available at: www.it.iitb.ac.in/∼sahoo/technote.pdf).

[36] C. Courcoubetis and R. Weber, *Pricing Communication Networks: Economics, Technology and Modeling.* John Wiley and Sons, 2003.

[37] D. P. Bertsekas, *Nonlinear Programming.* Athena Scientific, 2nd ed., 1999.

[38] F. Paganini, Z. Wang, J. C. Doyle, and S. H. Low, "Congestion Control for High Performance, Stability and Fairness in General Networks," *IEEE/ACM Transactions on Networking*, vol. 13, February 2005.

[39] J. Monks, V. Bharghavan, and W. Hwu, "A Power Controlled Multiple Access Protocol for Wireless Packet Networks," in *Proc. of IEEE INFOCOM*, pp. 219–228, April 2001.

[40] A. Muqattash and M. Krunz, "CDMA-based MAC Protocol for Wireless Ad hoc Networks," in *MobiHoc'03*, pp. 153–162, June 2003.

[41] A. Butala and L. Tong, "Cross Layer Design for Medium Access Control in CDMA ad-hoc Networks," *Eursip Journal on Applied Signal Processing*, December 2004.

[42] A. Muqattash and M. Krunz, "A Single-Channel Solution for Transmission Power Control in Wireless Ad-hoc Networks," in *MobiHoc'04*, pp. 210–222, May 2004.

[43] S. K. Bose, *An Introduction to Queuing Systems*. Kluwer Academic / Plenum Publishers, 2002.

[44] "Matlab version-6." http://www.mathworks.com.

[45] R. Manohar and A. Scaglione, "Power Optimal Routing in Wireless Networks," *Cornell Computer Systems Technical Report*, January 2003. (http://www.csl.cornell.edu/publications.html, Last visited on 28th June, 2005).

[46] P. P. Pham, S. Perreau, and A. Jayasuriya, "New Cross-Layer Design Approach to Ad-hoc Networks Under Rayleigh Fading," *IEEE Journal on Selected Areas in Communications*, vol. 23, January 2005.

[47] U. C. Kozat, I. Koutsopoulos, and L. Tassiulas, "A Framework for Cross-layer Design of Energy-efficient Communication with QoS Provisioning in Multi-hop Wireless Networks," in *Proc. of IEEE INFOCOM*, March 2004.

[48] J. Wang, L. Li, and S. H. L. J. C. Doyle, "Cross-Layer Optimization in TCP/IP Networks," *IEEE/ACM Transactions on Networking*, vol. 13, June 2005.

[49] C. C. Tan and N. C. Beaulieu, "On First-Order Markov Modeling for Rayleigh Fading Channel," *IEEE Transactions on Communications*, vol. 48, December 2000.

[50] V. Kawadia and P. R. Kumar, "A Cautionary Perspective on Cross-Layer Design," *IEEE Wireless Communications*, pp. 3–11, February 2005.