國 立 交 通 大 學

電機學院與資訊學院 電信學程

碩 士 論 文

在適應性無線系統提供 TCP 服務
之跨層暫存器管理技術

Cross-Layer Buffer Management for Supporting TCP Traffic
in Rate Adaptive Wireless Systems

研 究 生：李經豪

指導教授：王蒞君 教授

中 華 民 國 九 十 五 年 一 月

在適應性無線通訊系統提供 TCP 服務之跨層暫存器管理技術
Cross-Layer Buffer Management for Supporting TCP Traffic in Rate Adaptive Wireless Systems

研 究 生：李經豪　　Student：Ching-Hao Lee

指導教授：王蒞君　　Advisor：Li-Chun Wang

國 立 交 通 大 學
電機學院與資訊學院
碩士在職專班電信學程

碩 士 論 文

A Thesis

Submitted to Degree Program of Electrical Engineering and Computer Science

College of Electrical Engineering and Computer Science

National Chiao Tung University

in Partial Fulfillment of the Requirements

for the Degree of

Master of Science
in

Communication Engineering

January 2006

Hsinchu, Taiwan, Republic of China

中華民國九十五年一月

在適應性無線系統提供 TCP 服務之跨層暫存器管理技術

學生：李經豪　　　　　　　　　　　　　　　　　　　　指導教授：王蒞君 教授

國立交通大學電機學院與資訊學院 電信學程（研究所）碩士班

摘　　要

在下世代的無線通訊系統中，透過無線通道來傳送 TCP 的資料流是一個重要的議題，同時，也造成了很多挑戰。大部分的無線網路根據無線通道情形調整傳輸速率。雖然調整傳輸速率可以增加頻寬使用率，但在傳送端沒被通知無線頻寬變化的情形下，卻會也造成的一些擁塞的情形。為改善這種問題我們提出兩種方案。

在第一種方案中，我們首先分析實體層跟 TCP 暫存器容量的關係。然後我們得到一個關於暫存器容量、傳送速率與延遲的方程式。根據這個方程式，我們提出了一個 TCP 與實體層跨層的暫存器控制方法。從模擬的結果證明我們提出的暫存器管理方法可以彈性調整暫存器來權衡傳送速率與延遲的需求。

在第二種方案中，我們提出一個快速的估測可允許傳輸速率(FATRE)的方法。FATRE 應用於 TCP 層來改善應用層在適應性無線系統中的表現。使用這一個方法可以快速的通知可允許的傳輸速率給應用層進而改善頻寬使用率與擁塞資料量。在這個方案中，我們首先推導出一個關於頻寬使用率、多餘傳送位元、封包大小與通道容積的方程式。模擬結果應證推導出的方程式的正確性，並且證明在適應性無線系統中提出的 FATRE 方法可以改善頻寬使用率與擁塞資料量。

Cross-Layer Buffer Management
for Supporting TCP Traffic in Rate Adaptive Wireless Systems

student：Ching-Hao Lee Advisors：Dr. Li-Chun Wang

Degree Program of Electrical Engineering Computer Science
National Chiao Tung University

ABSTRACT

Delivering transmission control protocol (TCP) traffic through a wireless channel is crucial for the next generation wireless systems and also poses many challenges. Most wireless networks adapt the transmission rates according to radio channel conditions to improve channel utilization. However, it will cause congestion if the sender is not notified with the rate change in the wireless channel. In this thesis, we proposed two methods to overcome this issue.

In the first method, we suggest a TCP-PHY cross-layer buffer management mechanism based on a newly derived closed-form expression for the average throughput in terms of average number of buffered segments, queue delay and channel capacity. The derived analytical model considers the factors of bandwidth in the physical layer (PHY) and the queue capacity in the TCP layer. Our simulation results demonstrate that the proposed equation-based bandwidth-aware buffer management scheme can provide flexibility to performance tradeoff between throughput, delay and jitter in rate adaptive wireless systems when supporting TCP traffic.

In the second method, we propose a fast allowable transmission rate estimation (FATRE) mechanism in the TCP layer to adjust the source rate in the application layer for rate adaptive wireless systems. The proposed FATRE mechanism can quickly notify the allowable transmission rate to the source. In this scheme, we first derive closed-form expression for bandwidth utilization and the congestion data over the link capacity in terms of packet size and channel capacity. We validate analytic results by simulations and show that the FATRE mechanism can increase bandwidth utilization and decrease congested data.

Cross-Layer Buffer Management for Supporting TCP Traffic in Rate Adaptive Wireless Systems

A Master THESIS

Presented to

The Academic Faculty

By

Ching-Hao Lee

In Partial Fulfillment

of the Requirements for the Degree of

Master in Communication Engineering

Department of Communication Engineering

National Chiao-Tung University

January, 2006

Acknowledgments

I would like to thank my parents and my wife who alway give me supports and endless love. I especially would like to thank Dr. Li-Chun Wang who gave me many valuable suggestions and guidance in the research. I would not finish this work without his guidance, comments.

In addition, I am deeply grateful to my laboratory mates, Jane-Hwa, Chiung-Jang, Ming-Bing, Chih-Wen, Wei-Cheng, Chang-Lung, Chung-Wei, Lei, Wei-Jun, Yi-Cheng, Ching-Hau, Kuang-Nan ,Yun-Huai, Assane, Hung-Hsi, Cheng-Wei and Chu-Jung at WNLAB at the Department of Communications in National Chiao-Tung University. They provide me much assistance and share much happiness with me.

Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

Providing Internet services is a key issue for future wireless networks. In many Internet applications, such as HTTP, FTP, TELNET, and so on, the wildly used transport layer protocol is the transmission control protocol (TCP). However, establishing TCP connections in a wireless channel experiences two major challenges [1, 2]. First, high loss rate in a wireless channel significantly degrades the performance of the TCP connection because TCP's congestion control mechanism reduces window size at the occurrence of packet loss [3]. Second, fast and large rate adaptation in the wireless link can cause congestion or poor link utilization since TCP's flow control mechanism cannot follow the fast variation of radio channels [4–8]. In this thesis, we focus on the second problem, i.e., the performance issues of providing TCP connections in the rate adaptive wireless communication system.

The objectives of this thesis are two folds. First, we analyze the relation between the bandwidth in the physical layer and the maximum queue capacity in the TCP layer. Then we derive a closed-form expression for the average throughput in terms of the average number of buffered segments, queue delay and channel capacity. Last, based on the analytical model, we suggest a TCP-physical cross-layer equation-based buffer management mechanism, called the explicit rate change notification (ERCN). Our simulation results demonstrate that the proposed ERCN scheme can improve the performance tradeoff between throughput, delay and delay jitter for rate adaptive wireless systems when supporting TCP traffic.

Secondly, we propose a fast allowable transmission rate estimation (FATRE) scheme in the transmission control protocol (TCP) layer to adjust the source rate of the application layer. Because the radio network controller (RNC) dynamically adapts transmission data rates according to channel conditions, congestion or poor link bandwidth utilization may occur if the application layer is not aware of the rate change in the wireless channel. The proposed FATRE scheme can quickly notify the source up to the application about the link rate change in the wireless link. In this work, we first derive the closed-form expressions for bandwidth utilization and the congestion data over the link capacity in terms of packet size and channel capacity. We validate analytic results by simulations and show that the FATRE scheme can improve bandwidth utilization and have fewer congestion data than the legacy TCP-friendly rate control (TFRC) method.

1.1 Problem and Solution

1.1.1 An Equation-Based Bandwidth-Aware Buffer Management for TCP Congestion Control in Rate Adaptive Wireless Systems

In this part, we focus on the performance issues of providing TCP connections in the fast rate varying wireless channel. In the literature, the impacts of rate adaptation on the TCP performance were mostly investigated in the wireline environment. In [9], the authors proposed a random early detection (RED) mechanism to early detect the rate variation in the TCP link before the buffer reaches its capacity limit. In [3,10,11], the explicit congestion control (ECN) mechanism was proposed to distinguish congestion from random loss in the TCP link. However, because of the constraint of constant queue capacity, the RED and ECN mechanisms neither fully utilize the link capacity nor resolve the delay issue when a wireless system frequently change data rates.

To our knowledge, only few studies investigate the TCP performance over the fast varying wireless channel. In [12], the authors proposed the packet discard prevention counter (PDPC) method to adapt the queue capacity to the fast varying radio channel aiming to maximize radio link capacity. However, the impacts of queue delay and link delay jitter were not considered in [12].

In this work, we first derive a closed-form expression for the TCP throughput in terms of average number of buffered segments, queue delay and channel capacity. Next, we suggest a physical-TCP cross-layer equation-based buffer management technique, called the explicit rate change notification (ERCN) mechanism. The ERCN mechanism can explicitly notify the TCP sender about the usage of the wireless link capacity through an embedded bit in the acknowledgement (ACK) packet. Through analysis and simulations, we demonstrate that the proposed method can support the tradeoff among throughput, queue delay, and jitter.

1.1.2 A Fast Allowable Transmission Rate Estimation for Source Rate Adjustment of Application Layer in Rate Adaptive Wireless Systems

To improve the fluency of application layer, such as video conference, it is necessary to notify the source to adapt the source rate to the achievable transmission rates in the air link [13–19]. If the video source rate is lower than the air link data rates, bandwidth utilization is poor. On the contrary, if the video source rate is higher than the air link data rates, the congestion data will be buffered in the queue and increase the end-to-end delay.

The challenges to support real-time services in rate adaptive wireless systems lie in two folds [6–8, 20]. First, due to fast varying wireless link quality, a wireless system may change transmission rates much faster than the wire-line Internet. Second, the amount of rate variation in wireless systems is much larger than that in the wire-line Internet. To overcome these challenges, we propose a fast allowable transmission rate estimation scheme to notify the source in the application earlier.

In the literature, some allowable transmission rate estimation mechanisms for flow control have been discussed. For the wire-line Internet, an equation-based link bandwidth measurement was proposed for the TCP friendly Rate control (TFRC) protocol [21]. The TFRC estimates the allowable transmission rate by the congestion signals. Because the queue and link capacity in the Internet are usually contention-based, the allowable transmission rate estimation by the congestion signals require a longer time average. However, the TFRC-based rate estimation scheme does not take into account of the effect of fast and large rate changes. In the rate adaptive wireless systems, the larger measurement latency will cause poor link bandwidth utilization and have more congestion data than in the Internet.

Thus, we suggest a fast allowable transmission rate estimation (FATRE) scheme for the source rate adjustment in rate adaptive wireless systems with dedicated link method. The basic idea of the FATRE scheme is to fast estimate the allowable transmission rate by viewing congestion signals in rate adaptive wireless systems with a dedicated link as the indicator of rate change.

1.2 Thesis Outline

The rest of this thesis is organized as follows. Chapter 2 introduces the background on TCP protocol. In Chapter 3, we describe the ERCN mechanism. In Chapter 4, we present the FATRE scheme. At last, Chapter 5 gives the concluding remarks and suggestions for future works.

CHAPTER 2

Background on TCP

2.1 The Impact of Congestion Control on the CWND Variation

Figure 2.1 illustrates that the congestion window (CWND) size of the TCP connection varies according to the additive increase multiplicative decrease (AIMD) principle [22]. Specifically, the CWND size increases by one segment in the next round when receiving ACK, and decreases by half when receiving the rate change notification. In the figure, a round (T_R) is defined as the duration from the beginning of sending the first packet until the ACK is received. A TCP block period (T_B) is defined as the duration between two congestion indications. Note that a higher data rate yields a larger CWND size. As shown in the figure, for link data rate $L_h > L_l$, the CWND size (denoted by W) in the period of L_h is larger than that during the period of L_l.

Figure 2.2 shows the i-th TCP block in four rounds each of which has various CWND sizes. Let W_i be the CWND size in the last round of the i-th TCP block. According to the AIMD principle, the CWND size in the first round of the i-th TCP block is $W_{i-1}/2$. Generally, there are $(W_i - 0.5W_{i-1} + 1)$ rounds in the i-th block. Thus, the total number of segments in the i-th block becomes $(W_i - 0.5W_{i-1} + 1)(0.5W_{i-1} + W_i)/2$. Note that T_B is related to the average number of segments buffered at the queue, which is equal to the sum of the round trip time in each round. Observing the

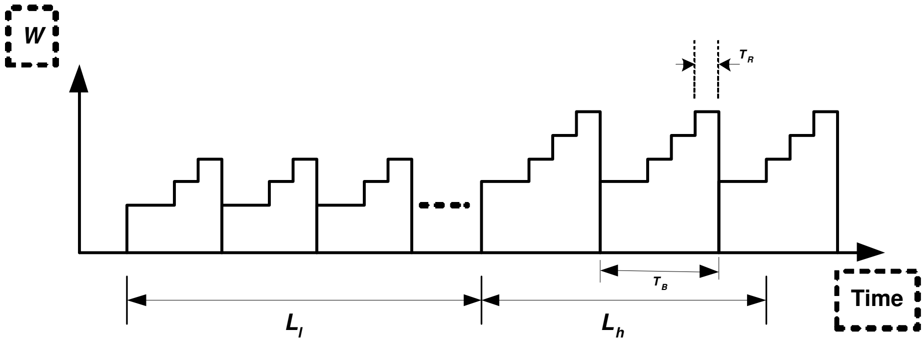

Figure 2.1: The variation of CWND size in the TCP connection, where L_l and L_h are two different data rates ($L_h > L_l$), T_R is the round period and T_B is the block period.

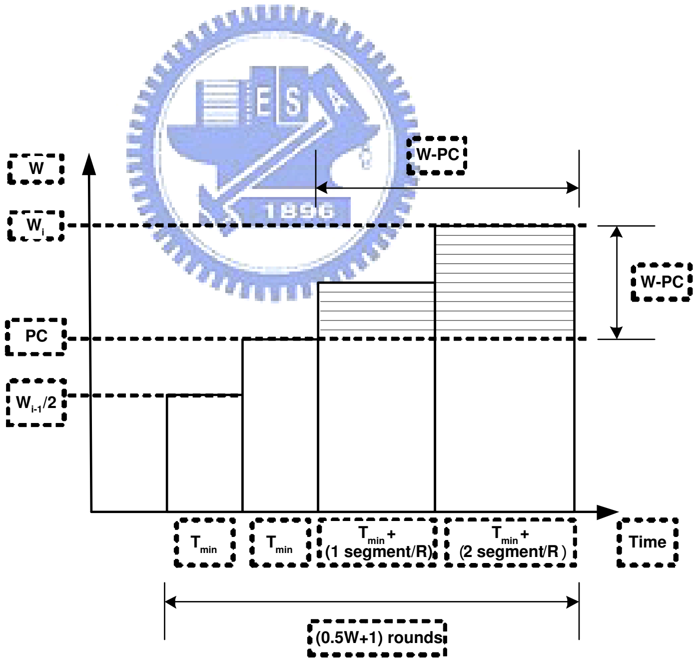

Figure 2.2: The variations of the CWND sizes and the buffered segments in a TCP block.

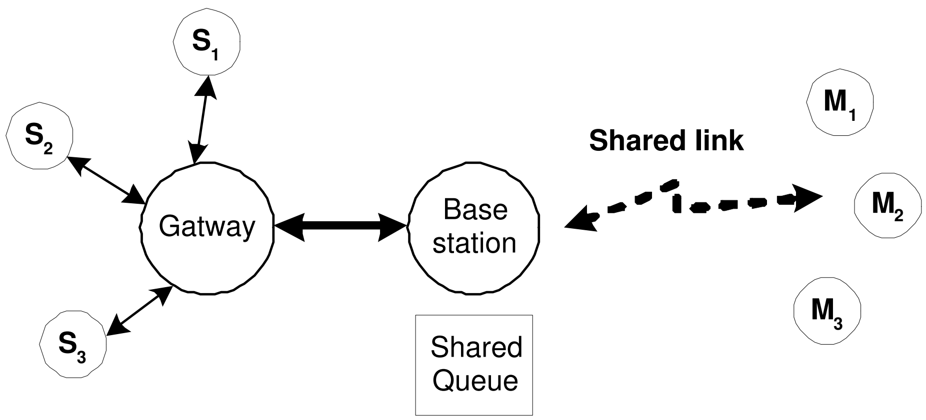

Figure 2.3: The TCP data flow in a shared (contention-based) link.

figure, in the first and second round, the CWND size is smaller than link capacity. On the contrary, in the third and forth rounds, the CWND size is larger than the link capacity. Thus, some TCP segments are buffered and lead to longer delay.

2.2 TCP Flows in Shared Channel and Dedicated Channel

Figure 2.3 shows the TCP data flow in a shared (contention-based) link, the queue capacity and link bandwidth are shared. In the figure, there are three TCP flows between (S_1, M_1), (S_2, M_2) and (S_3, M_3), respectively. Because the bottleneck link is shared by three data flows, the congestion signals may not be direcly mapped to the link bandwidth variation. Thus, the link bandwidth measurement scheme in the shared link require a longer moving average [21, 23–25].

Figure 2.4 shows the TCP flow in a rate adaptive wireless system with a dedicated link, the wideband code-division multiple-access (WCDMA) system. In the figure, there are also three TCP flows between (S_1, M_1), (S_2, M_2) and (S_3, M_3),

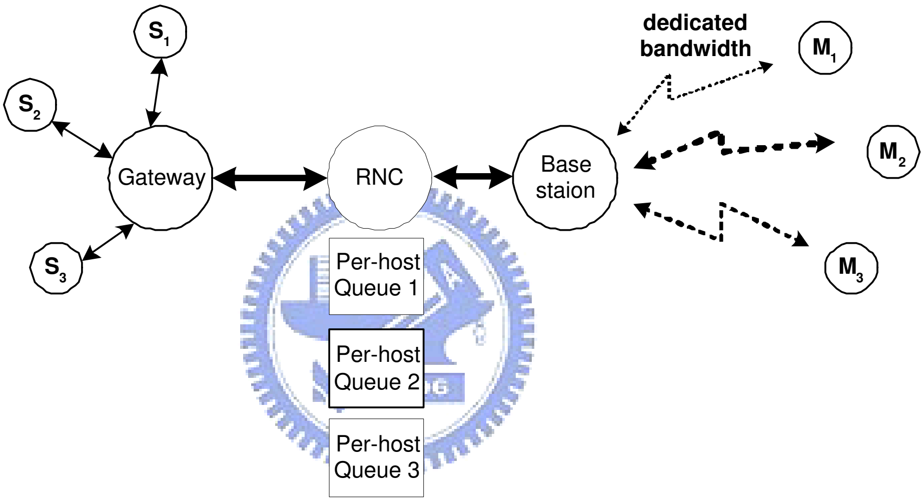

Figure 2.4: The TCP data flows with a dedicated link.

respectively. The dedicated link bandwidth between the base station and mobile terminals is the bottle-neck link. Between the RNC and the base station is the link layer traffic. In such system, the per-host queue of the RNC is dedicated to each mobile terminal [12]. Thus, the congestion signals from the per-host queue can be an indicator for the variation of the link bandwidth. Consequently, the link bandwidth can be measured without the moving average. Thus, the delay of the link bandwidth measurement is shortened.

CHAPTER 3

An Equation-Based Bandwidth-Aware Buffer Management for TCP Congestion Control in Rate Adaptive Wireless Systems

In this chapter, we develop a PHY/TCP cross-layer buffer management scheme. The proposed buffer management scheme can improve the performance tradeoff between throughput and delay. In addition, the proposed scheme also can decrease the delay jitter for rate adaptive wireless systems when supporting TCP traffic. This chapter is organized as follows. Section 3.1 introduces the motivation. Section 3.2 analyzse the impact of the queue size of the buffer management on the throughput and delay. Section 3.3 proposes the ERCN mechanism. Numerical results are shown in Section 3.4. We give our concluding remarks in Section 3.5.

3.1 Motivation

In this chapter, we consider the performance issues of TCP connections in the downlink transmission of the rate adaptive wireless system, such as the wideband code-division multiple-access (WCDMA) system. Figure 3.1 shows a TCP flow from the

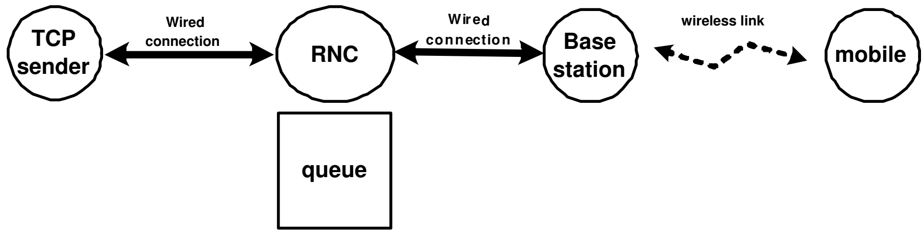

Figure 3.1: The TCP flow in a rate adaptive wireless system, such as the WCDMA system. The RNC is equipped with a queue for congestion control.

TCP sender through an intermediate radio network controller (RNC) and base station and to mobile terminals. Between the RNC and base station is the link layer traffic. Usually, the link capacity between TCP sender and the base station is higher than that of the wireless link. Note that the RNC is equipped with a queue to buffer the TCP segments for congestion control.

Note that when the CWND size is larger than link capacity, the channel is fully utilized. Nevertheless, some TCP packets are buffered at the queue of the RNC and so increase the packet delay. Thus, optimizing the performance tradeoff between throughput and queue delay of delivering TCP traffic become a challenging and important issue.

One of the major issues is to design an effective buffer management technique. To this end, instead of trial and errors, we suggest an equation-based buffer management mechanism, called the ERCN mechanism. The ERCN mechanism is based on a newly derived closed-form expression for the average throughput in terms of average number of buffered segments, queue delay and channel capacity. The developed analytical model can facilitate the design of an effective buffer management technique which can take into account of both link utilization and queue delay.

It will be shown that the ERCN mechanism can effectively improve the performance tradeoffs between throughput, queue delay and jitter in supporting TCP services under a fast rate varying wireless channel.

3.2 Analysis

3.2.1 Impact of the Average Number of Buffered Segments on Throughput and Delay

We first derive the relation between the TCP throughput and average buffered segments in the queue of the RNC for congestion control. The throughput in the T_B duration is equal to

$$TH = \frac{(0.5W_{i-1} + W_i)(W_i - 0.5W_{i-1} + 1)}{2}\frac{1}{T_B} . \tag{3.1}$$

For a constant rate, $W_{i-1} = W_i = W$. We can further simplify TH as

$$TH = \frac{\frac{3}{8}W^2 + \frac{3}{4}W}{T_B} . \tag{3.2}$$

The average round trip time T is

$$T = T_{min} + \frac{Q}{L} . \tag{3.3}$$

Denote Q the average number of buffered segments in a round and T_{min} the round trip time when no segments is buffered in the queue. For a general case with $(0.5W + 1)$ rounds in T_B, we know

$$T_B = T(0.5W + 1) . \tag{3.4}$$

Substituting (3.4) into (3.2), we obtain

$$TH = \frac{\frac{3}{8}W^2 + \frac{3}{4}W}{(0.5W + 1)(T_{min} + \frac{Q}{L})} . \tag{3.5}$$

With the knowledge of T_{min} and L, T and TH in (3.3) and (3.5) becomes a function of Q and W.

3.2.2 The Relations of Q and W

Now we derive the relation of the average number of buffered segments (Q) and the CWND size (W) during the period of T_B. When the CWND size is larger than the radio link capacity (PC), the number of buffered segments is equal to the difference of the CWND size and PC. Thus, there are $(W_i - 0.5W_{i-1} + 1)$ rounds between two congestion indications. For simplicity, we consider the case of $W_{i-1} = W_i = W$, i.e., $0.5W + 1$ rounds between two congestion indications.

In the case $PC \leq W \leq 2PC$, 1, 2, ..., $(W - PC)$ segments are buffered in the last $(W - PC)$ rounds, respectively. Thus, there are total $(1 + W - PC)(W - PC)/2$ buffered segments in $(0.5W + 1)$ rounds. Thus, it is followed that

$$Q = \frac{(1 + W - PC)(W - PC)}{2} \frac{1}{0.5W + 1} \ . \quad (3.6)$$

From (3.6), we can express W as

$$W = \frac{-A + \sqrt{A^2 - 4(PC^2 - PC - 2Q)}}{2}, \quad (3.7)$$

where

$$A = -2PC - Q + 1 \ . \quad (3.8)$$

For the case of $2PC < W$, $(0.5W - PC)$, $(0.5W - PC + 1)$, ..., $(W - PC)$ segments are buffered in the $(0.5W + 1)$ rounds, respectively. Therefore, we can obtain

$$Q = \frac{[(0.5W - PC) + (W - PC)](0.5W + 1)}{2} \frac{1}{0.5W + 1} \ , \quad (3.9)$$

Accordingly, we have

$$W = \frac{4(PC + Q)}{3} \ . \quad (3.10)$$

Based on (3.5), (3.7) and (3.10), we can plot the relation between the TCP throughput and Q in Fig. 3.2. In the figure, the size of TCP segment is 500 bytes;

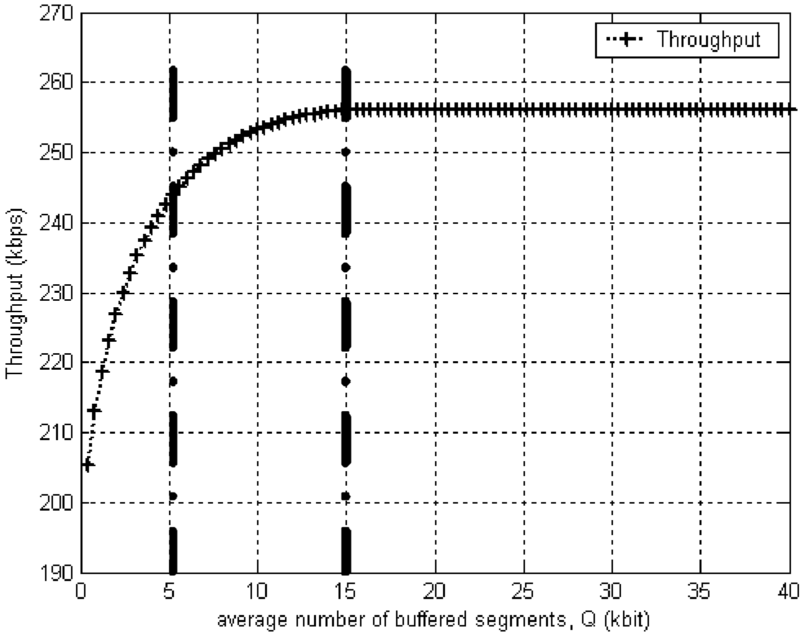

Figure 3.2: The TCP throughput is a function of Q.

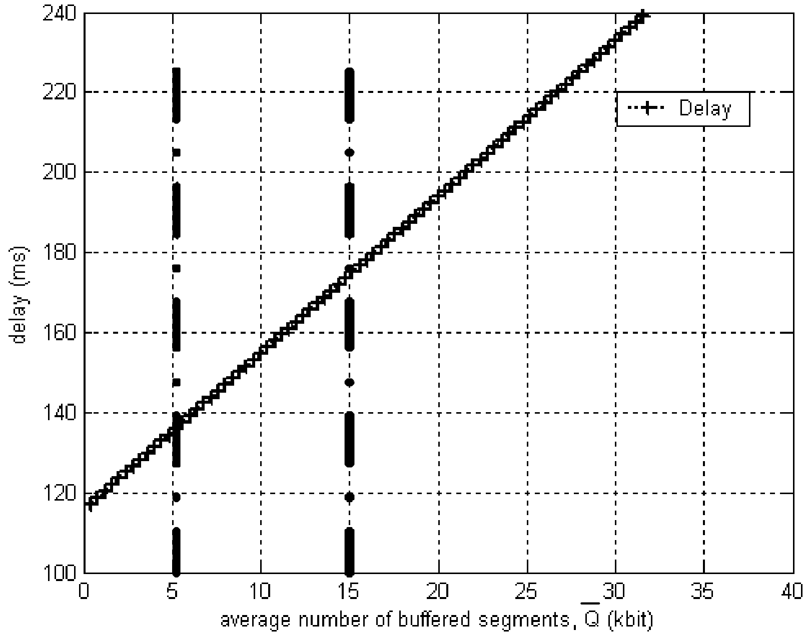

Figure 3.3: The TCP delay is a function of Q.

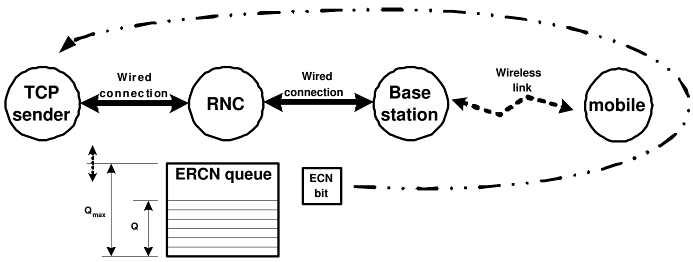

Figure 3.4: The ERCN mechanism adapts the Q_{max} in the rate adaptive wireless system.

L = 256 kbps; T_{min} = 100 msec; $PC = (T_{min} \times L)$ is equal to 25.6 kbits. When Q increases from 5 kbits to 15 kbits, the TCP throughput increases only 10 kbits, i.e. (2%). On the other hand, Fig. 3.3 shows the relation between the TCP delay and Q based on (3.3), (3.7) and (3.10). When Q increases from 5 kbits to 15 kbits, the TCP delay increases about (30%), i.e. 40 msec. Thus, compared to Q=15 and 5 kbits, one can know that it is possible to reduce TCP delay (30%) at the only cost of slightly lower throughput reduction (2%).

3.3 The Proposed ERCN Mechanism

3.3.1 The Behavior of the ERCN Mechanism

From the above analysis, we can improve the tradeoff between the link utilization and delay by controlling Q. In our work, we adapt the queue capacity (Q_{max}) of the ECN mechanism [10] to control Q in the rate adaptive wireless system. We call the proposed mechanism the ERCN mechanism.

Figure 3.4 shows that the ERCN mechanism notifies the TCP sender about the rate change by the marked ECN bit. The ERCN mechanism marks the ECN bit of the TCP header when Q is closer or larger than Q_{max}. The marked ECN bit will

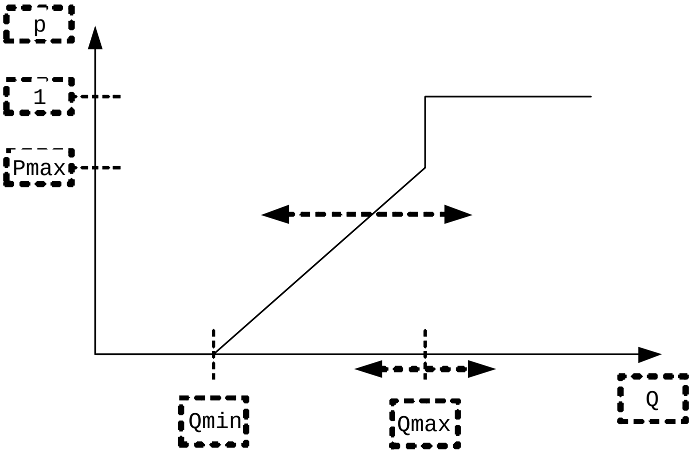

Figure 3.5: The ERCN mechanism adapts the Q_{max} dynamically to different radio link data rates.

be echoed back to the TCP sender. Thus, the TCP sender can lower the transmission rate and decrease Q.

Denote p the marking rate. We show the relation between Q, Q_{max} and p of the ERCN mechanism in (3.11).

$$p = \begin{cases} 0 & , Q < Q_{min} \\ P_{max}\frac{Q-Q_{min}}{Q_{max}-Q_{min}} & , Q_{min} \leq Q < Q_{max} \\ 1 & , Q_{max} \leq Q \end{cases} \tag{3.11}$$

According to (3.11), Figure 3.5 shows that when $Q < Q_{min}$, $p = 0$. when $Q_{min} \leq Q < Q_{max}$, p is proportional to Q. Otherwise, when $Q_{max} \leq Q$, $p = 1$. The figure also shows the adaptive Q_{max} will impact the relation between Q, and p.

3.3.2 The Relation between Q_{max}, T and TH

From Section III, we know different Q can meet different tradeoffs between throughput and delay in the rate adaptive wireless system. In the ERCN mechanism, since Q

is relative to Q_{max}, we derive an equation to show the relation between Q_{max} and Q. Then, substituting (3.3) and (3.5) into the equation, we can obtain relations between Q_{max}, T and TH which can facilitate Q_{max} design of the ERCN mechanism for different delay and throughput requirements.

Because Q is relative to p of the ERCN mechanism, we investigate the relation among Q and p. First, From [26], we know TH is also relative to p, i.e.,

$$TH = \frac{\frac{1}{p} + \sqrt{\frac{8}{3p}}}{T(\sqrt{\frac{2}{3p}} + 1)} . \tag{3.12}$$

Now we derive the relation between Q and TH. Since the average segments in the TCP link is equal to $TH \times T$, Q is equal to the difference between of $TH \times T$ and PC. Thus, Q is equal to

$$Q = TH \times T - PC . \tag{3.13}$$

Substituting (3.12) into (3.13), we can obtain

$$Q = \frac{\frac{1}{p} + \sqrt{\frac{8}{3p}}}{\sqrt{\frac{2}{3p}} + 1} - PC . \tag{3.14}$$

From (3.14), we obtain the relation between p and Q as follows:

$$p = (\frac{-\sqrt{\frac{2}{3}}(Q + PC - 2) + Y}{2(Q + PC)})^2, \tag{3.15}$$

where

$$Y = \sqrt{\frac{2}{3}(Q + PC - 2)^2 + 4(Q + PC)} . \tag{3.16}$$

Now, substituting (3.15) into (3.11), we can derive the relation between Q_{max} and Q. In (3.11), we set the Q_{min} as $Q_{max}/3$ as recommended by [9]. Thus, we obtain

$$Q_{max} = \frac{3(P_{max})(Q)}{2((\frac{-\sqrt{\frac{2}{3}}(Q+PC-2)+Y}{2(Q+PC)})^2) + P_{max}} . \tag{3.17}$$

With the knowledge of P_{max} and PC, Q_{max} is a function of Q.

Below we derive the relations between Q_{max}, T and TH. First, from (3.3), we can obtain

$$Q = L(T - T_{min}) \ . \tag{3.18}$$

Substituting (3.18) into (3.17), we can derive the relation between Q_{max} and T as

$$Q_{max} = \frac{3(P_{max})(L(T - T_{min}))}{2((\frac{-\sqrt{\frac{2}{3}}(L(T-T_{min})+L\times T_{min}-2)+M}{2(L(T-T_{min})+L\times T_{min})})^2) + P_{max}} \ . \tag{3.19}$$

where denote M as

$$M = \sqrt{\frac{2}{3}(L(T - T_{min}) + PC - 2)^2 + 4(L(T - T_{min}) + PC)} \ . \tag{3.20}$$

From (3.5), we can express $TH = f(Q)$, then obtain Q as

$$Q = f^{-1}(TH). \tag{3.21}$$

Substituting (3.21) into (3.17), we can express the relation between Q_{max} and TH as

$$Q_{max} = \frac{3(P_{max})(f^{-1}(TH))}{2((\frac{-\sqrt{\frac{2}{3}}(f^{-1}(TH)+L\times T_{min}-2)+N}{2(f^{-1}(TH)+L\times T_{min})})^2) + P_{max}} \ , \tag{3.22}$$

where N is denoted as

$$N = \sqrt{\frac{2}{3}(f^{-1}(TH) + PC - 2)^2 + 4(f^{-1}(TH) + PC)} \ . \tag{3.23}$$

3.4 Numerical Results

3.4.1 The Simulation Model

Since the WCDMA system is a well-known wireless communication system with the rate adaptation mechanism, we validate and evaluate the performance of the ERCN

Table 3.1: The connection information of the simulation model.

	to node	bandwidth ($Mbps$)	delay ($msec$)
TCP source	gateways	100	30
gateways	RNC	622	10
RNC	basestaion	622	15

mechanism in the WCDMA system. We simulate a system shown in Fig. 3.6. by applying the enhanced UMTS radio access network extensions (EURANE) [27] for ns-2 (network simulator version 2.26) [28]. In the figure, a TCP connection is established between the TCP sender and mobile terminal. Within the TCP path, the layer 2 traffic exist between the RNC, base station and in the radio link, while RNC and gateway can detect layer 3 traffic. Thus RNC is equipped a queue with the ERCN mechanism. The gateways node can transmit and receive packets from the Internet to the WCDMA system.

The link capacity between TCP sender and the base station is high ,while the capacity of the wireless link between the base station and the mobile terminal is very limited. Table. 3.1 shows the delay and bandwidth information of the simulation model. The radio link capacity is dynamically allocated by the radio resource control (RRC) function in the WCDMA system.

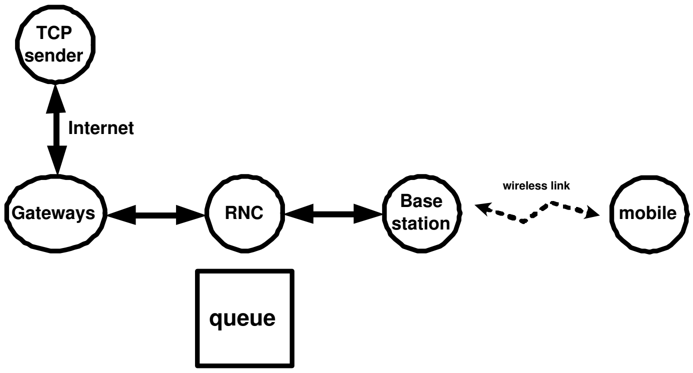

Figure 3.6: The simulation setup of the EURANE in the NS2 simulator.

3.4.2 Validation

In Fig. 3.6, let P_{max}=0.04; Q_{max}=3Q_{min} and TCP segment size is 500 bytes of the ERCN mechanism and the radio data rates vary between 128 kbps and 256 kbps every 20 seconds.

Figure 3.7 shows the relation of Q_{max} and the delay for L=128 kbps and 256 kbps. The analytical results match the simulation results well. Now consider the delay requirement is 260 ms set for example. From the figure, one can obtain Q_{max}=10 kbits and 55 kbits for L=128 kbps and 256 kbps, respectively. The achievable throughput by this option can be obtained by the next figure.

Figure 3.8 shows the relation of Q_{max} and throughput by simulations and analysis. One can find the simulation results match the analytical results well. From the figure, one can determine the expected throughput for different values of Q_{max}. For example, to meet the throughput requirement 120 kbps and 240 kbps for L=128 kbps and 256 kbps, we can set Q_{max}=10 kbits and 20 kbits, respectively.

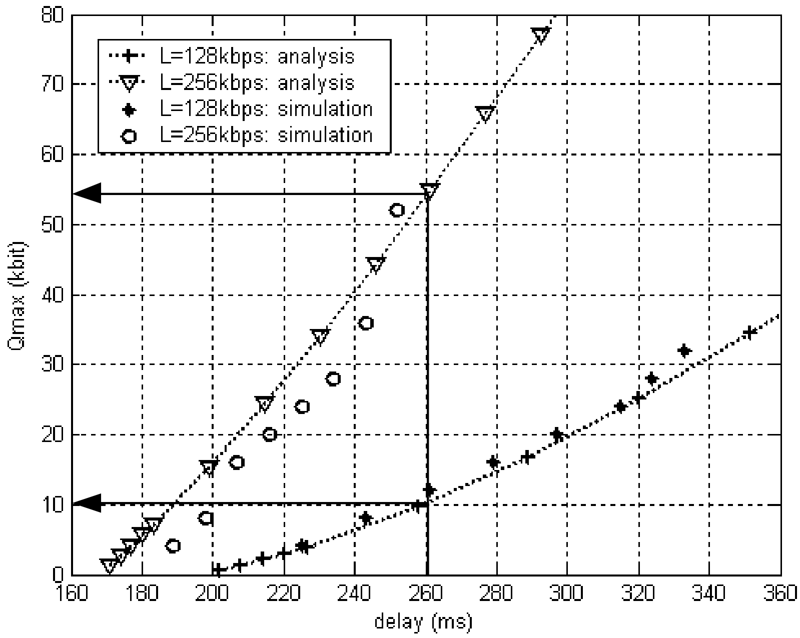

Figure 3.7: The analytical and simulation results show the relation between Q_{max} and delay when $R = 128$ kbps and $R = 256$ kbps.

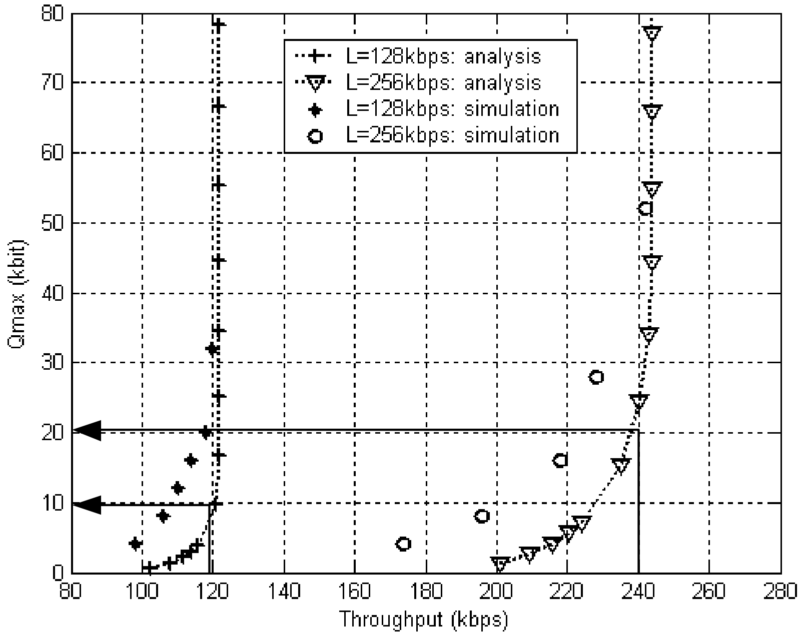

Figure 3.8: The analytical and simulation results show the relation between Q_{max} and throughput when $L = 128$ kbps and $L = 256$ kbps.

Table 3.2: The comparisons of the ECN and ERCN mechanism.

	Throughput ($kbps$)	delay ($msec$)
ECN	115	307
ERCN	119	286

3.4.3 Comparison of the ECN and ERCN Mechanism

Based on the previous simulation setup, we have an example to compare the numerical results of the ECN and ERCN mechanism in the WCDMA system. We assume the RRC function allocated 128 kbps for the wireless interface between the base station and mobile terminal in the wireless link setup state. After the wireless link setup state, for maintaining specific bit error rate, the WCDMA system can adapt the link data rates to varying wireless link conditions. For example, the data rate is 64 kbps in the poor channel condition, while the data rate becomes 256 kbps in a better link condition [20]. In this example, the ERCN mechanism dynamically adapts the Q_{max} for different data rates, but Q_{max} in the ECN mechanism is a fixed value. Table. 3.2 shows the numerical results of the ERCN and ECN mechanism. It is shown the throughput of the ERCN mechanism is improved by 4 kbps (3%) and the delay is increased by 21 msec (7%) over the ECN mechanism.

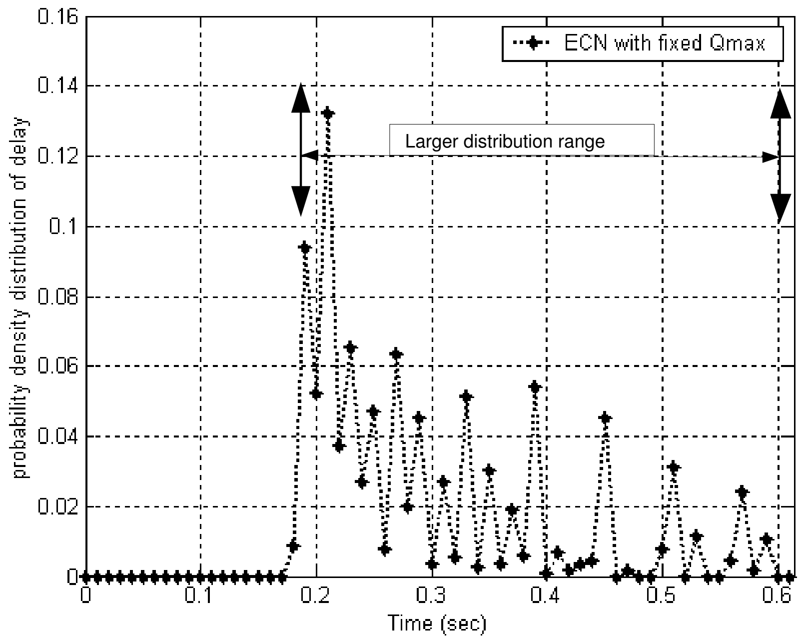

Figure 3.9: The delay distribution of the ECN mechanism.

3.4.4 The Delay Jitter of the ERCN Mechanism

Since the ERCN mechanism maintains a proper queue size for different data rates, thereby improving the delay jitter over the ECN mechanism. Figures 3.9 and 3.10 show the simulation results of the ECN and ERCN mechanism, respectively. We find that the delay of the ERCN mechanism is spread in a smaller range than the ECN mechanism. Table 3.3 compares the delay jitter between the ECN and ERCN mechanism. The delay jitter of the ERCN mechanism is improved $1.13 \times 10^4 (msec^2)$ (65%) over the ECN mechanism.

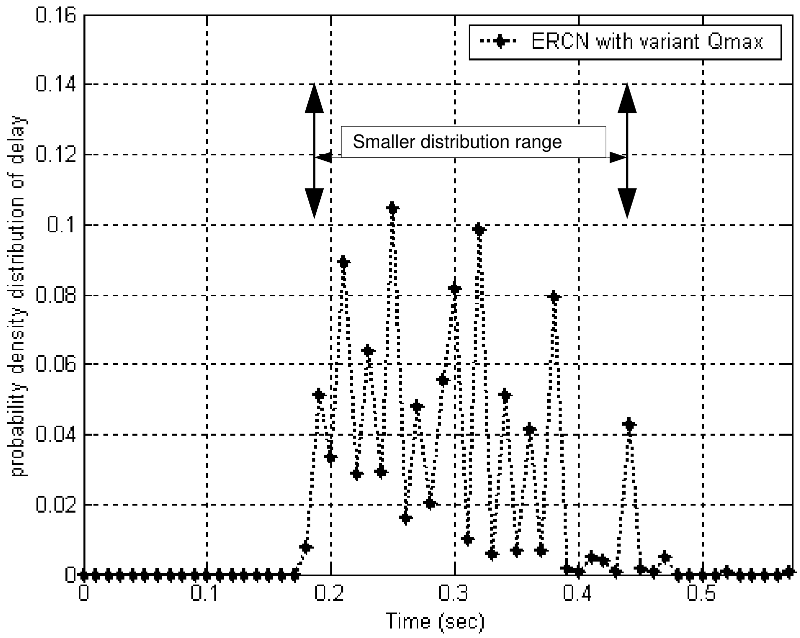

Figure 3.10: The delay distribution of the ERCN mechanism.

Table 3.3: The delay jitter comparisons of the ECN and ERCN mechanism.

	delay jitter $\times 10^4 (msec^2)$
ECN	1.73
ERCN	0.6

3.5 Conclusions

In this chapter, we have analyzed the relation among the physical layer bandwidth, the TCP layer throughput, delay and queue size when delivering the TCP traffic in rate adaptation wireless systems. The developed analytical formulas can facilitate the design of queue sizes in the buffer of the RNC for different TCP throughput and delay requirements. Specifically, we propose an explicit rate change notification (ERCN) mechanism to dynamically change the queue capacity of the buffer in the RNC based on the TCP/Physical cross-layer performance issues. Hence, when the radio link capacity in the physical layer is changed, the suitable queue capacity in the RNC can be effectively adapted and the TCP sender can accordingly change the transmit data rates, thereby both the TCP delay and throughput requirements can be met in the varying radio channel conditions. In addition, we validate the accuracy of the analysis by simulations in the WCDMA system which is one kind of the wireless communication system with the rate adaptation mechanism. Finally, we have numerical results to show the ERCN mechanism can improve the TCP throughput, delay and delay jitter when the rate adaptation mechanism in rate adaptation wireless systems.

CHAPTER 4

A Fast Allowable Transmission Rate Estimation for Source Rate Adjustment of Application Layer in Rate Adaptive Wireless Systems

In this chapter, we suggest a fast allowable transmission rate estimation (FATRE) scheme for the source rate adjustment in rate adaptive wireless systems with dedicated link method. The basic idea of the FATRE scheme is to estimate the allowable transmission rate faster by taking advantage of the congestion signals in rate adaptive wireless systems. The results of analysis and simulation show that the proposed scheme can improve the performance of link utilization and delay in the application layer. This chapter is organized as follows. Section 4.1 introduces the motivation. Section 4.2 proposes the FATRE scheme. Section 4.3 analyzes the estimation latency and the performance of the FATRE mechanism. The analysis validation and numerical results are shown in Section 4.4. We give our concluding remarks in Section 4.5.

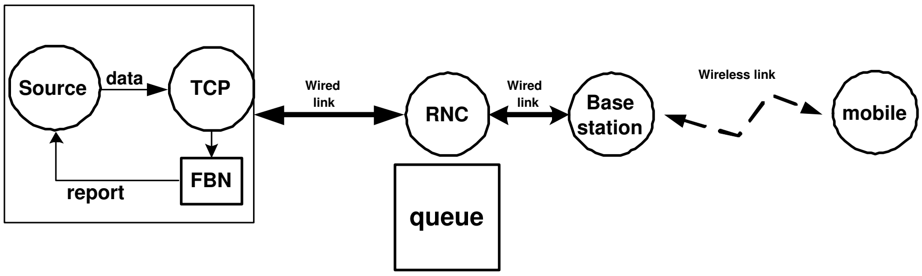

Figure 4.1: The source rate adjustment with the FATRE mechanism in the rate adaptive wireless communication system.

4.1 Motivation

In this chapter, we consider the source rate adjustment issues upon TCP layer in rate adaptive wireless systems, such as the wideband code-division multiple-access (WCDMA) system. Figure 4.1 shows a TCP flows from the application layer through an intermediate radio network controller (RNC), base station and to mobile terminals. Usually, the link capacity between TCP sender and the base station is higher than that of the wireless link and the wireless link capacity varies faster and larger than the wired-link. Between the RNC and the base station is the link layer traffic. Thus, the RNC is equipped with a queue to buffer the TCP segments for congestion control [29]. In addition, we implement the ECN mechanism in the queue of the RNC to distinguish the congested packets from lost packets in the wireless channel [3, 30]. Therefore, the queue will mark the ECN bit as the congestion signal instead of dropping a packet when the buffer packet size is larger than the pipe capacity.

In the application layer, if the source rate can not adapt to the varying air link data rates of the wireless link, poor link bandwidth utilization or larger delay may occur. It is because if the source rate is smaller than the air link data rate, the link bandwidth utilization is poor. On the contrary, when the source rate is lager than the air link data rate, some congestion data will be buffered in the queue and increase the end-to-end delay in the application layer.

In the literature, some allowable transmission rate estimation mechanisms for flow control have been discussed. For the wire-line Internet, an equation-based link bandwidth measurement was proposed in the TCP friendly Rate control (TFRC) protocol [21]. The TFRC estimates the allowable transmission rate estimation by the congestion signals. Because the queue and link capacity in the Internet are contention-based, the allowable transmission rate estimation by the congestion signals need long time average for accuracy estimation. Thus, the TFRC scheme measures the allowable transmission rate with moving average method which increases measurement latency. However, the above allowable transmission rate estimation schemes do not take into account of the effect of fast and large rate changes in the rate adaptive wireless communication system. In the rate adaptive wireless systems, the larger measurement latency will cause poorer link bandwidth utilization and larger congestion data than in the Internet. Therefore, the estimation latency in rate adaptive systems need be faster than the Internet.

Some queue management mechanisms for improving the throughput and delay of the TCP layer in rate adaptive systems have been discussed [12] and [29]. For fully utilizing the link, [12] proposed the packet discard preventive counter (PDPC) method for improve the wireless link utilization of the TCP layer in the rate adaptive wireless communication systems. In [29], the proposed early rate change notification (ERCN) can flexibly adjust the queue capacity in the RNC for the tradeoff between TCP throughput and delay. However, in [12] and [29], the focus is on the buffer management to improve the wireless link utilization and delay in the TCP layer not in the application layer. In this work, we will also evaluate the performance of the flexible queue capacity control in the application layer.

Thus, we suggest the FATRE scheme for the source rate adjustment in the rate adaptive wireless communication system. The FATRE scheme can take advantage of the congestion information from the dedicated link and implemented in the TCP layer which can quickly estimate the available allowable transmission rate and notify the source. Then, the source can adjust the transmission rate according to the report from the TCP layer. Therefore, the bandwidth utilization is improved and the congestion data over the wireless link capacity are reduced. Consequently, with the FATRE scheme, the application layer can have better link utilization and smaller delay in the rate adaptive wireless communication system .

There are three contributions in this chapter. First, for rate adaptive wireless systems with dedicated link, we proposed the FATRE scheme which can estimate the allowable transmission rate quickly. Second, we derive closed-form expressions for bandwidth utilization and the congestion data over the link capacity in terms of packet size and channel capacity in the FATRE scheme. From the analysis, we show the flexible queue capacity control can be used for the tradeoff between link utilization and congested data. Third, the analytic results are validate by simulations and show that the FATRE mechanism improves bandwidth utilization and fewer congestion data than the legacy TCP-friendly rate control (TFRC) in the rate adaptive wireless systems with dedicated link.

4.2 The Operation Principle

Figure 4.2 shows a TCP block sent in the duration between two marked explicit congestion notification (ECN) bits, where denote T_i the duration between two ECN bits, W_i the CWND size of the last round, P_i the link pipe capacity and T_{Di} the dormant period [30] of $i-th$ TCP block. In the RNC, we set the queue capacity as Q_i and Q_{i-1} while the air link data rate is L_i and L_{i-1}, respectively. We can flexibly control the Q_i and Q_{i-1} for the tradeoff between link utilization and delay.

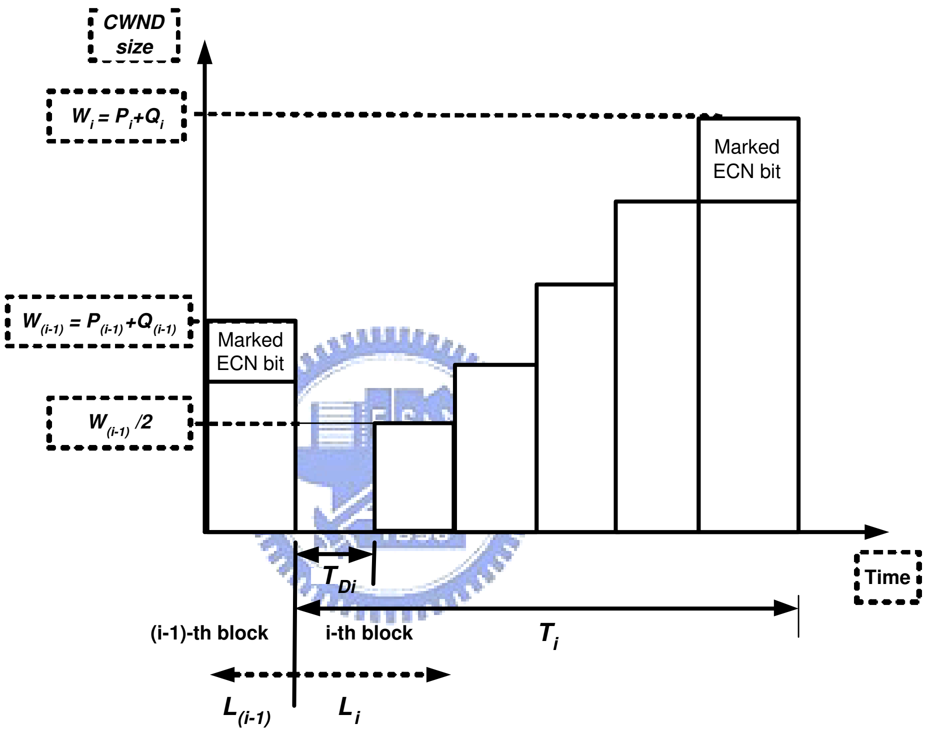

Figure 4.2: The packets with marked ECN bits in the TCP blocks in the queue of the RNC in the rate adaptive wireless communication system.

The basic ideal of the FATRE scheme is measuring the number of received ACKs and the duration between two consecutive marked ECN bits. Let N the received number of the ACKs in T_i. Then, the total transmitted bits in a TCP block is equal to $(N \times S)$. Denote S the TCP segment size in bit. Then, we can estimate the allowable transmission rate R_t during a TCP block as

$$R_t = \frac{N \times S}{T_i} . \tag{4.1}$$

Note that in this chapter, the mobile terminal echoes an ACK packet while receiving a TCP segment.

Figure 4.3 shows the flow chart for the FATRE scheme operation principle. First, whenever an ACK packet is received, N is increased by one. Then the FATRE scheme checks whether the ECN bit is marked in the received ACK packet. If so, the allowable transmission rate is estimated by (4.1). Otherwise, the TCP sender waits for the next ACK. After estimating the allowable transmission rate, the FATRE scheme resets N to zero and immediately reports the estimated allowable transmission rate to the source. Consequently, The source can adjust the source rate according to the estimated allowable transmission rate. Note that the estimation latency will impact the performance of the FATRE scheme. In the next section, we will analyze the estimation latency of the FATRE scheme in the rate adaptive wireless communication system.

4.3 Analysis

4.3.1 Measurement Latency

Figure 4.4 shows the estimation latency in the FATRE scheme while the air link data rate varies between L_l and L_h. In the figure, the raising edge represents that the

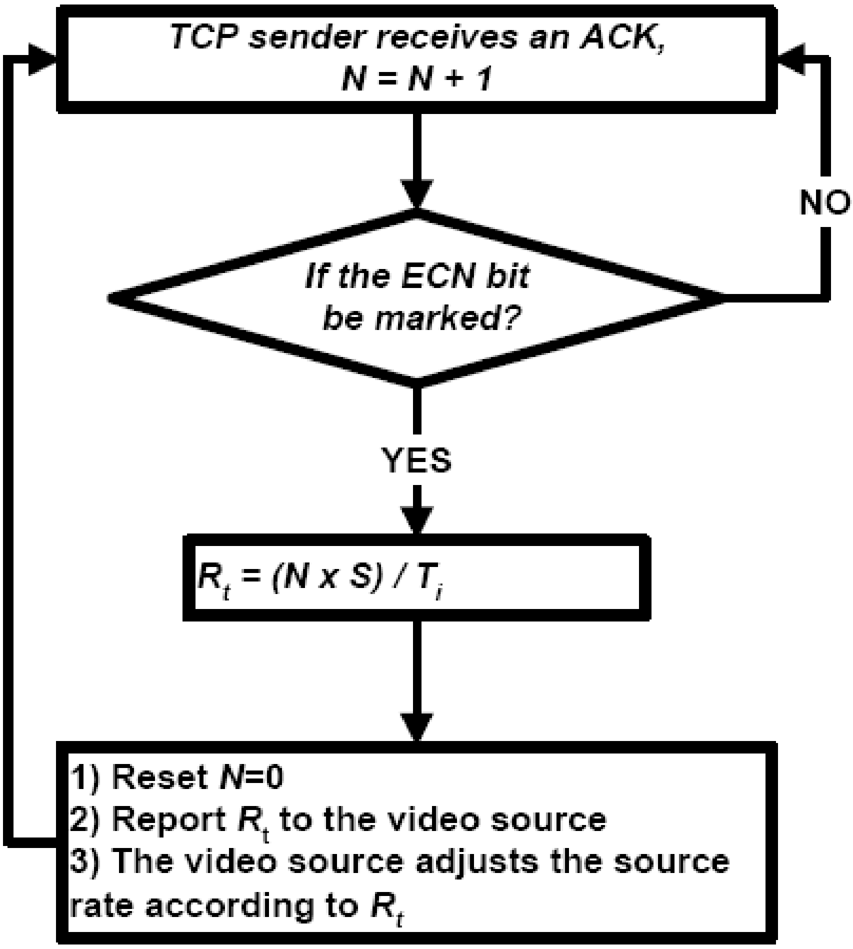

Figure 4.3: The flow chart of the FATRE behavior.

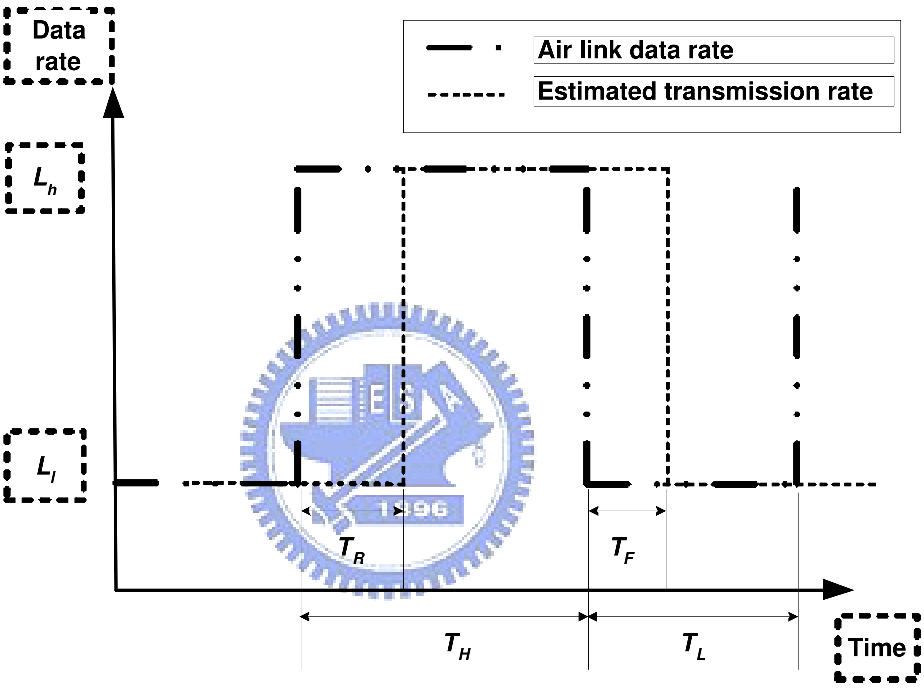

Figure 4.4: The estimation latency of the FATRE scheme while the wireless link bandwidth varies between L_l and L_h.

air link data rate varies from L_l to L_h, and the falling edge means that the air link data rate varies from L_h to L_l. Let T_R and T_F represent the estimation latency of the FATRE scheme of the rasing and falling edge, respectively. During T_R, the source rate is smaller than the available air link data rate, thereby wasting the radio resource. On the other hand, during T_F, the source rate is larger than supported rate in the wireless link, thereby causing congested data which will be buffered in the queue and increases the end-to-end delay.

Now we discuss the FATRE scheme performance in T_R. To begin with, in the $i-th$ block, let N_i be the total number of received ACKs and N_{Di} be the received number of ACKs in the dormant period T_{Di}. Note that in the dormant period, the TCP sender will not send TCP segments until receives the number of ACKs as a half of the previous CWND size. Therefore, N_{Di} is equal to a half of the previous CWND size, i.e., $(W_{i-1}/2)$. Observing Fig. 4.2, one can find that the number of received ACK packets N_i is equal to

$$N_i = (\frac{(\frac{W_{i-1}}{2} + W_i)(W_i - \frac{W_{i-1}}{2} + 1)}{2}) + N_{Di}. \quad (4.2)$$

During T_R in Fig. 4.4, the first block is the $i-th$ block. In the i-th block, let N_1 be the received number of the ACKs and T_1 be the duration of the $i-th$ block. The figure shows that

$$W_{i-1} = P_l + Q_l, \quad (4.3)$$

and

$$W_i = P_h + Q_h \ , \quad (4.4)$$

where denote Q_h and Q_l are the queue capacity when the air link data rate is L_h and L_l, respectively, and P_h and P_l are the pipe capacity when the air link data rate is L_h and L_l, respectively. Substituting (4.3) and (4.4) into (4.2), we obtain

$$N_1 = \frac{((P_l + Q_l)/2 + P_h + Q_h)(P_h + Q_h - (P_l + Q_l)/2 + 1)}{2} + \frac{P_l + Q_l}{2} \ . \quad (4.5)$$

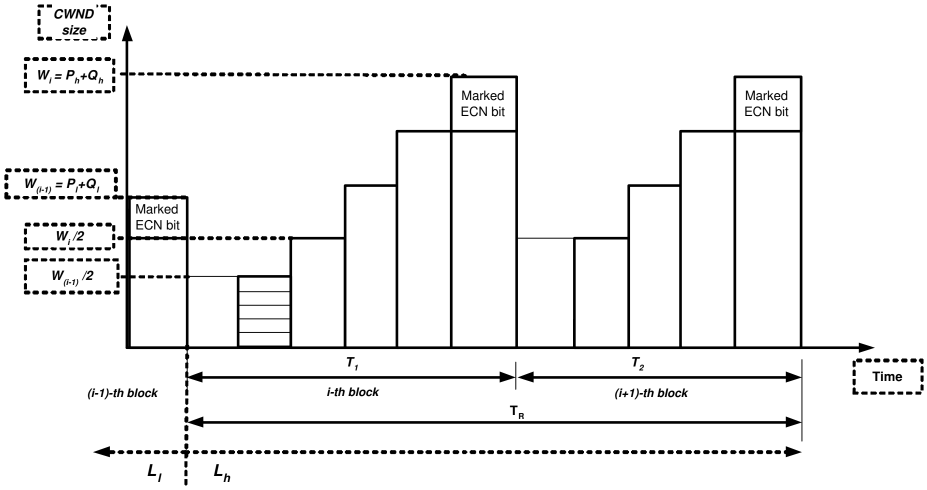

Figure 4.5: The estimation latency of the FATRE scheme while the wireless link bandwidth varies from L_l and L_h.

Note that if the CWND size is larger than the link capacity, the time interval of two received ACKs is equal to the delay of the radio link, i.e., (TCP segment size/radio link capacity). Otherwise, there will be no buffered segment and we can express the total duration of receiving the ACKs in a round is equal to T_{min}. Define T_{min} the round trip time (RTT) when no segment be buffered in the queue of the RNC. For example, in Fig. 4.5, the CWND size of the first round is smaller than the pipe capacity P_h and the duration of the first round is equal to T_{min}. Thus, in $i-th$ block, there are $(P_h - (P_l + Q_l)/2 + 1)$ rounds have CWND size smaller than the link capacity and $((P_h + Q_h) - P_h + 1)$ rounds have CWND size larger than the link capacity. Thus, we can express T_1 as

$$T_1 = A + (P_h - (P_l + Q_l)/2 + 1) \times T_{min} \ . \tag{4.6}$$

where

$$A = \frac{(P_h + (P_h + Q_h))((P_h + Q_h) - P_h + 1)}{2} \times \frac{S}{L_h}$$

$$= \frac{(2P_h + Q_h)(Q_h + 1)}{2} \times \frac{S}{L_h} \ . \quad (4.7)$$

Denote R_1 the estimated allowable transmission rate after T_1. We can express R_1 as

$$R_1 = N_1 \times S/T_1 \ . \quad (4.8)$$

In the $(i+1)-th$ block of Fig. 4.5, we have

$$W_i = W_{i+1} = P_h + Q_h \ . \quad (4.9)$$

Let N_2 be the received number of the ACKs and T_2 be the duration in the $(i+1)-th$ block. Substituting (4.9) into (4.2), we can obtain N_2 as

$$N_2 = \frac{((P_h + Q_h)/2 + P_h + Q_h)(P_h + Q_h - (P_h + Q_h)/2 + 1)}{2} + \frac{P_h + Q_h}{2} \quad (4.10)$$

There are $(P_h - (P_h + Q_h)/2 + 1)$ rounds have CWND size smaller than the link capacity. Therefore, we can express T_2 as

$$T_2 = A + (P_h - (P_h + Q_h)/2 + 1) \times T_{min} \ . \quad (4.11)$$

Denote R_2 the estimated allowable transmission rate after $(i+1)-th$ block, we can obtain R_2 as

$$R_2 = \frac{N_2 \times S}{T_2} \ . \quad (4.12)$$

In the following, we discuss the FATRE scheme performance in T_F. In Fig. 4.2 during T_F, we have

$$W_{i-1} = P_h + Q_h, \quad (4.13)$$

and

$$W_i = P_l + Q_l \ . \quad (4.14)$$

Now we derive T_F in the falling edge. Denote N_F the number of received ACKs in the $i-th$ block. In the case of $((P_h+Q_h)/2 < P_l)$, we can express N_F as

$$N_F = \frac{((P_h+Q_h)/2+P_l+Q_l)(P_l+Q_l-(P_h+Q_h)/2+1)}{2} + \frac{P_h+Q_h}{2} \quad (4.15)$$

Because there are $(P_l-(P_h+Q_h)/2+1)$ rounds have CWND size smaller than pipe capacity. Therefore, we obtain T_F as

$$T_F = D + (P_l-(P_h+Q_h)/2+1) \times T_{min}, \quad (4.16)$$

where

$$D = \frac{(P_l+(P_l+Q_l))((P_l+Q_l)-P_l+1)}{2} \times \frac{S}{L_l}$$
$$= \frac{(2P_l+Q_l)(Q_l+1)}{2} \times \frac{S}{L_l} \quad . \quad (4.17)$$

In the case of $((P_h+Q_h)/2 \geq P_l)$ as shown in Fig. 4.6, because the congestion will occur in the first round of the $i-th$ block, the TCP block has only one round and the marked ECN bit is in the $(2P_l-th)$ packet. Thus, we have

$$N_F = (P_l+Q_l)+(P_h+Q_h)/2, \quad (4.18)$$

and

$$T_F = (P_l+Q_l+(P_h+Q_h)/2) \times \frac{S}{L_l} \quad . \quad (4.19)$$

Denote R_F the estimated allowable transmission rate after T_F. We can express R_F as

$$R_F = \frac{N_F \times S}{T_F} \quad . \quad (4.20)$$

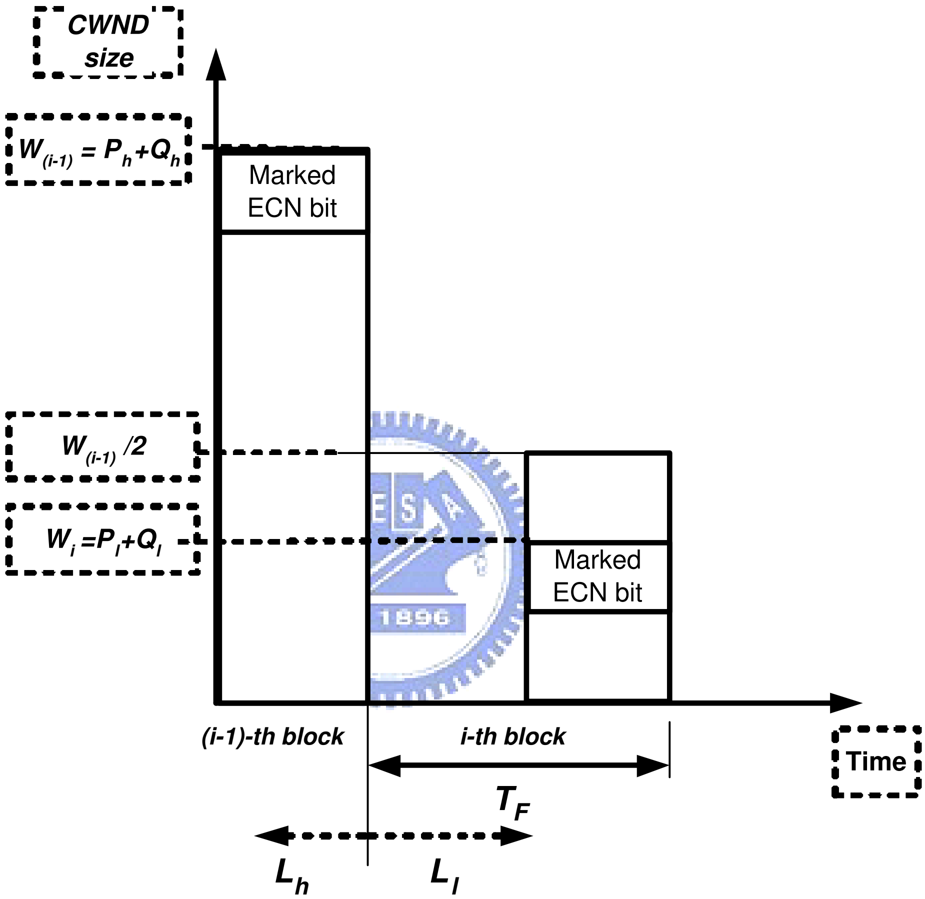

Figure 4.6: The estimation latency of the FATRE scheme while the air link data rate varies from L_h and L_l in the case of $P_h \geq 2P_l$.

4.3.2 Link Utilization and Congested Data

The source adjusts the transmission rate according to the report of the previous block by the FATRE scheme. Therefore, in Fig. 4.5, the source rate in the block (i) and (i+1) is equal to R_F and R_1, respectively. Let T_H and T_L are the duration while the air link data rate is equal to L_h and L_l, respectively. Denote W the wasted the air link capacity in a period of $T_H + T_L$. We can express Z as

$$Z = T_1(L_h - R_F) + T_2(L_h - R_1) + (T_H - T_1 - T_2)(L_h - R_2). \tag{4.21}$$

Denote B as the offered channel capacity in the T_H and T_L. We can obtain B as

$$B = L_h \times T_H + L_l \times T_L \ . \tag{4.22}$$

Denote ρ the link utilization in a period of $T_H + T_L$. Then, ρ can be written as

$$\rho \ = \ \frac{B - Z}{B} \ . \tag{4.23}$$

Next, define the congested data (C) as the the transmitted bits exceeding the link capacity. During the delay interval T_F in the falling edge, the source transmits at a rate of R_2, which is higher than the wireless link bandwidth L_l. Thus, during T_F, C is equal to

$$C \ = \ T_F \times (R_2 - L_l) \ . \tag{4.24}$$

The congested data will be buffered in the queue, thereby, increasing the end-to-end delay.

4.3.3 The Relation between Queue Capacity, ρ and C

Based on (4.23) and (4.24), we can plot the relation between queue capacity, ρ and C. Figure 4.7 shows the relation between queue capacity and link utilization. In

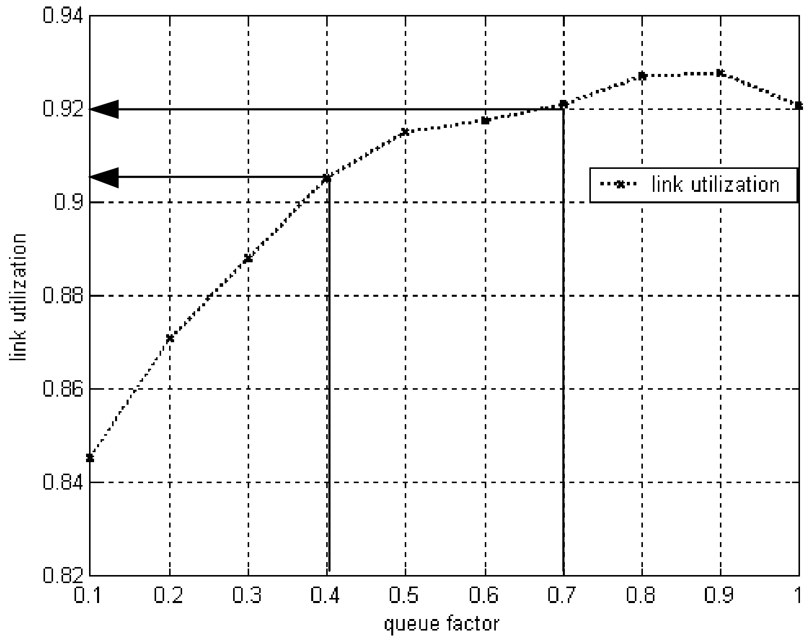

Figure 4.7: The link utilization is a function of queue capacity.

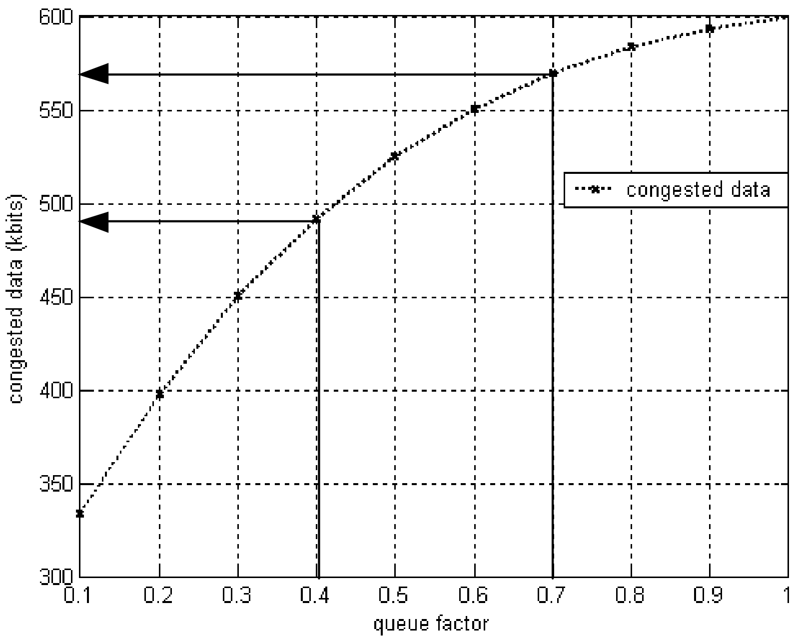

Figure 4.8: The congested data is a function of queue capacity.

the figure, the size of TCP segment is 1000 bytes; L_l = 400 kbps; L_h = 800 kbps; T_{min} = 100 msec; queue capacity = $(queue\ factor) \times (air\ link\ data\ rate) \times (T_{min})$. For example, if we choose the queue factor=0.9, the queue capacity is equal to $0.9 \times 800\ kbps \times 100\ ms = 72\ kbits$ and $0.9 \times 400\ kbps \times 100\ ms = 36\ kbits$ for the air link data rate is equal to L_h and L_l, respectively. Figure 4.7 shows when queue factor decreases from 0.7 to 0.4, the link utilization decreases only 1.5%. On the other hand, Fig. 4.8 shows the relation between the queue capacity and congested data. When queue factor decreases from 0.7 to 0.4, the congested data decreases about 80 kbits, i.e. 14%. Thus, compared to the queue factor = 0.7 and 0.4, one can know that it is possible to reduce congested data (14%) at the only cost of slightly lower link utilization reduction (1.5%).

4.4 Numerical Results

4.4.1 Analysis Validation

We first validate the analytical formulae of estimation latencies T_R and T_F by using the NS2 simulator [28]. The system is set up as in Fig. 4.1. According to performance requirements specified in a rate adaptive wireless communication system, i.e., the WCDMA system [20], the basic video phone services requires a link bandwidth of 128 kbps and the enhanced video phone services requires 384 kbps of link bandwidth. Thus, in our simulations, we set L_l=128 kbps and L_h=384 kbps. Assume that the link bandwidth varies every 10 seconds. Let the TCP segment size be 1000 bytes, the round trip delay of the wired link be 100 ms, and the delay of the radio link be 10 ms. We use the continuous bit rate (CBR) source in the NS2 simulator to model the real-time video source. The CBR can adjust the source rates according to the estimated allowable transmission rate from the transport layer. Fig. 4.9 show that

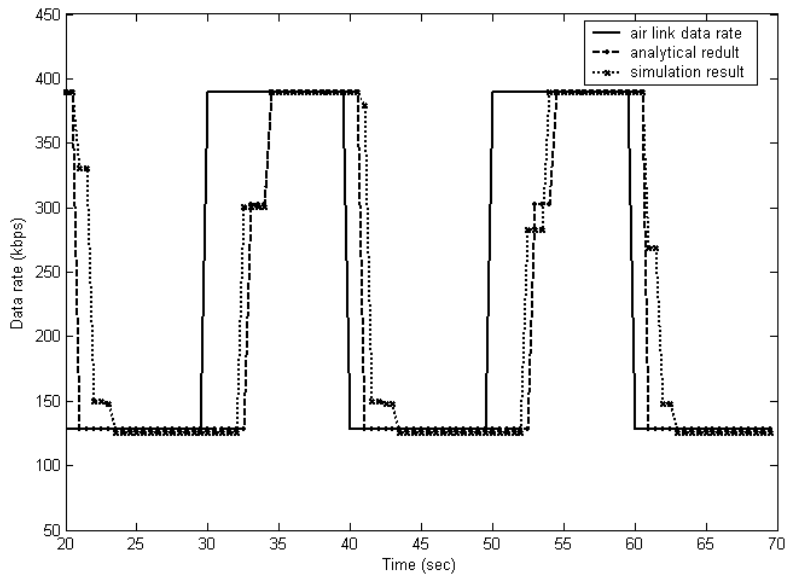

Figure 4.9: The analytic and simulation results of the FATRE scheme.

the analytical results of T_R and T_F in (7) and (11) are close to the simulation results.

4.4.2 Performance Comparison of the TFRC and FATRE Scheme

Figure 4.10 compares the simulation results using the FATRE method and the legacy TFRC method. The simulation setup is similar to Fig. 4.9. From the figure, we find that the FATRE scheme can follow the rate change of the rate adaptive wireless system much better than the TFRC method. Thus, in the raising edge, the FATRE scheme has better link utilization and in the falling edge, the FATRE scheme can reduce the congested data.

Table. 4.1 summarize performance results of the TFRC and FATRE schemes obtained in Fig. 4.10. One can see that the FATRE scheme can improve 11 % of link utilization over the TFRC method. In the meanwhile, the congested data of the FATRE method is reduced from 626 kbits to 421 kbits compared to the TFRC method. The performance improvements of the TFRC method comes from the smaller estimation latency.

Table 4.1: Comparison of link utilization and congestion data of the TFRC and FATRE scheme.

	$\rho(\%)$	$C(kbits)$
TFRC	73	626
FATRE	84	421

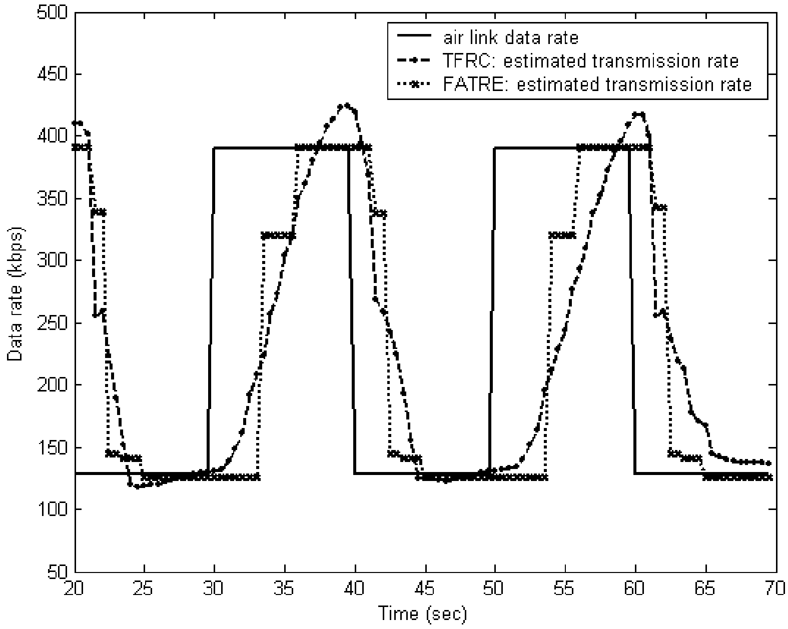

Figure 4.10: The allowable transmission rate estimation of the TFRC and FATRE scheme.

4.5 Conclusions

In this chapter, we have propose a fast allowable transmission rate estimation (FATRE) scheme for source rate adjustment in rate adaptive wireless communication systems with dedicated link method. The proposed FATRE mechanism can estimated allowable transmission rate and notify the source quickly, thereby increase the link utilization and decrease the end-to-end delay in the application layer. We have analyzed the key parameters in the FATRE scheme, including the estimation latency, link utilization, and congested data. From the analysis, we show the flexible queue capacity control can be used for the tradeoff between link utilization and congested data. By the NS-2 simulations, we demonstrate the that the FATRE scheme outperforms the legacy TCP-Friend Rate Control (TFRC) method with better link utilization and fewer congested data. The impact of loss of the ECN bit in the wireless on the proposed FATRE method is one of interesting research topics in the future.

CHAPTER 5

Concluding Remarks

There are two major contributions in the thesis. First, we suggest a TCP-physical cross-layer equation-based buffer management mechanism, called the explicit rate change notification (ERCN). The ERCN mechanism can flexibly control the queue capacity for the tradeoff between throughput and delay in the TCP layer. Second, we propose a fast allowable transmission rate estimation (FATRE) scheme for the source rate adjustment in rate adaptive wireless systems with dedicated link method. The FATRE scheme improves bandwidth utilization and has fewer congested data than the legacy TCP-friendly rate control (TFRC) in rate adaptive wireless systems. This thesis discuss the following research topics:

1. For buffer management issues, we propose a TCP-physical cross-layer equation-based buffer management mechanism, called the explicit rate change notification (ERCN).

2. Analyze the relation between the bandwidth of the physical layer and the maximum queue capacity of the TCP layer.

3. Derive a closed-form expression for the average throughput in terms of average number of buffered segments, queue delay and channel capacity.

4. For source rate adjustment issues, we suggest a fast allowable transmission rate estimation (FATRE) scheme .

5. Derive closed-form expressions for bandwidth utilization and the congestion data over the link capacity in terms of packet size and channel capacity in the FATRE mechanism.

5.1 An Equation-Based Bandwidth-Aware Buffer Management for TCP Congestion Control in Rate Adaptive Wireless Systems

In Chapter 3, We derive a closed-form expression for the average throughput in terms of average number of buffered segments, queue delay and channel capacity. The analytical model shows that we can flexibly control the queue capacity for the tradeoff between throughput and delay. Thus, in the chapter, we suggest a TCP-physical cross-layer equation-based buffer management mechanism, called the explicit rate change notification (ERCN). By the simulation results, we demonstrate that the proposed ERCN scheme can used for the tradeoff between throughput and delay. In addition, the ERCN can also improve the delay jitter of TCP layer.

5.2 A Fast Allowable Transmission Rate Estimation for Source Rate Adjustment of Application Layer in Rate Adaptive Wireless Systems

In Chapter 4, based on the difference of contention-based and dedicated link wireless system, we proposed the FATRE scheme. In the chapter, we derive closed-form expressions for bandwidth utilization and the congestion data over the link capacity in terms of packet size and channel capacity. The analytic results is validated by simulations and show that the FATRE scheme can improve bandwidth utilization and decrease congestion data in rate adaptive wireless systems with dedicated link.

5.3 Suggestion for Future Work

The following topics are suggested to extend the current work for future research:

- The impact of loss of the ECN bit in the wireless on the proposed ERCN and FATRE method.
- Derive the closed-form equation of the FATRE method while the rate change of the wireless link and the ECN-marked do not occur in the same time.

Bibliography

[1] S. Shakkottai, T. S. Rappaport, and P. C. Karlsson, "Cross-layer design for wireless networks," *IEEE Communications Magazine*, vol. 41, no. 10, pp. 74 – 80, Oct. 2003.

[2] F. Lefevre and G. Vivier, "Understanding TCP's behavior over wireless links," *Communications and Vehicular Technology*, vol. 1, pp. 123 – 130, Oct. 2000.

[3] 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, NO. 5, pp. 689 – 702, Oct. 2003.

[4] D. Dutta and Y. Zhang, "An early bandwidth notification (EBN) architecture for dynamic bandwidth environment," *IEEE International Conference on Communications*, vol. 4, pp. 2602 – 2606, Apr. 2002.

[5] E. Hossain, D. I. Kim, and V. K. Bhargava, "TCP performance under dynamic link adaptation in cellular multi-rate wcdma networks," *IEEE International Conference on Communications*, vol. 25, no. 2, pp. 1834–1839, Apr. 2002.

[6] ——, "Analysis of TCP performance under joint rate and power adaptation in cellular WCDMA networks," *IEEE Wireless Communications*, vol. 3, no. 3, pp. 865 – 879, May 2004.

[7] P. J. Ameigeiras, J. Wigard, and P. Mogensen, "Impact of TCP flow control on the radio resource management of WCDMA networks," *IEEE VTC*, vol. 2, pp. 977 – 981, May 2002.

[8] M. Yavuz and F. Khafizov, "TCP over wireless links with variable bandwidth," *IEEE VTC*, vol. 3, pp. 1322 – 1327, Sept. 2002.

[9] S. Floyd, "RED wed page," Aug. 2001. [Online]. Available: http://www.aciri.org/floyd/red.html

[10] K. K. Ramakrishnan, S. Floyd, and D. Black, *The Addition of Explicit Congestion Notification (ECN) to IP*, RFC 3168, Sept. 2001.

[11] M. Kwon and S. Fahmy, "TCP increase/decrease behavior with explicit congestion notification (ECN)," *IEEE International Conference on Communications*, vol. 4, pp. 2335 – 2340, May 2002.

[12] M. Sagfors, R. Ludwig, M. Meyer, and J. Peisa, "Buffer management for rate-varying 3G wireless links supporting TCP traffic," *IEEE VTC*, vol. 1, pp. 675 – 679, Apr. 2004.

[13] P. B. Ssesanga and M. K. Mandal, "Efficient congestion control for streaming scalable video over unreliable IP networks," *IEEE International Conference on Image Processing*, vol. 3, pp. 2059 – 2062, Oct. 2004.

[14] M. Kim and J. A. Copeland, "Bandwidth sensitive caching for video streaming application," *IEEE International Conference on Communications*, vol. 3, pp. 1557 – 1562, May. 2003.

[15] M. Chen and A. Zakhor, "Rate control for streaming video over wireless," *IEEE INFOCOM*, vol. 1, pp. 1181–1190, Mar. 2004.

[16] Q. Zhang, W. Zhu, and Y. Q. Zhang, "Network-adaptive rate control and unequal loss protection with TCP-friendly protocol for scalable video over internet," *Journal of VLSI Signal Processing Systems*, vol. 34, no. 1, pp. 67–81, May 2003.

[17] Y. S. Chan, Y. Pei, Q. Qu, and J. W. Modestino, "On cross-layer adaptivity and optimization for multimedia CDMA mobile wireless networks," *in Proceeding of The First IEEE-EURASIP International Symposium on Control, Communications, and Signal Processing*, vol. 1, pp. 579–582, Mar. 2004.

[18] T. Kim and M. H. Ammar, "Optimal quality adaptation for scalable encoded video," *IEEE Communications Magazine*, vol. 23, no. 2, pp. 344– 356, Feb. 2005.

[19] M. Baldi and Y. Ofek, "End-to-end delay analysis of videoconferencing over packetswitched networks," *IEEE/ACM Transactions on Networking*, vol. 8, pp. 479–492, Aug. 2000.

[20] H. Holma and A. Toskala, *WCDMA FOR UMTS*. England: John Wiley and Sons, 2000.

[21] M. Handley, J. Pahdye, S. Floyd, and J. Widmer, *TCP Friendly Rate Control (TFRC): Protocol Specification*, RFC 3448, Jan. 2003.

[22] R. Stevens, *TCP/IP Illustrated, vol. 1*. Addison-Wesley, 1994.

[23] S. Floyd, M. Handley, J. Padhye, and J. Widmer, "Equation-based congestion control for unicast applications," *In Proceedings of ACM SIGCOMM*, vol. 30, no. 4, pp. 43–56, Aug. 2000.

[24] Z. Wange, S. Banerjee, and S. Jamin, "Media-friendliness of a slowly-responsive congestion control protocol," *Proc. of the 14th ACM Intl Workshop on Network and Operating Systems Support for Digital Audio and Video*, vol. 1, Jun. 2004.

[25] M. Gerla, M. Y. Sanadidi, R. Wang, and A. Zanella, "TCP westwood: Congestion window control using bandwidth estimation," *Proc. of the 14th ACM Intl Workshop on Network and Operating Systems Support for Digital Audio and Video*, vol. 3, pp. 1698–1702, Nov. 2001.

[26] J. Padhye, V. Firoiu, D. F. Towsley, and J. F. Kurose, "Modeling TCP reno performance: a simple model and its empirical validation," *IEEE/ACM Transactions on Networking*, vol. 2, no. 3, pp. 133 – 145, Apr. 2000.

[27] Enhanced UMTS Radio Acces Network Extension for ns-2. [Online]. Available: http://www.ti-wmc.nl/eurane/

[28] The Network Simulator - ns-2. [Online]. Available: http://www.isi.edu/nsnam/ns/

[29] L. C. Wang and C. H. Lee, "A TCP-physical cross-layer congestion control mechanism for the multirate WCDMA system using explicit rate change notification," *IEEE International Conference on Advanced Information Networking and Applications*, vol. 2, pp. 449 – 452, Mar. 2005.

[30] S. Bae and S. Chong, "TCP-friendly wireless multimedia flow control using ECN marking," *Proceedings of IEEE Global Telecommunications Conference*, vol. 2, pp. 1794 – 1799, Nov. 2002.

Vita

Ching-Hao Lee

He was born in Taiwan, R. O. C. in 1974. He received a B.S. in Electrical Engineering from National Cheng Kung University in 1996. From July 2002 to December 2005, he worked his Master degree in the Wirelsss Internet System Engineering Lab in the Department of Communication Engineering at National Chiao-Tung University. His research interests are in the field of wireless communications.