國立交通大學

電信工程學系

碩士論文

在中繼型態WiMAX網路系統中提升傳輸量並保證服務品質的排程法

QoS Guaranteed Throughput Enhancement Scheduling in WiMAX Relay-Assisted Network

研 究 生：莊佳璇

指導教授：張仲儒 博士

中 華 民 國 九十六年六月

在中繼型態WiMAX網路系統中提升傳輸量並保證服務品質的排程法

QoS Guaranteed Throughput Enhancement Scheduling in WiMAX Relay-Assisted Network

研究生：莊佳璇　　　　Student：Chia-Shueng Chuang

指導教授：張仲儒 博士　　Advisor：Dr. Chung-Ju Chang

國立交通大學

電信工程學系

碩士論文

A Thesis

Submitted to Department of Communication Engineering

College of Electrical Engineering and Computer Engineering

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of Master

in

Communication Engineering

June 2007

Hsinchu, Taiwan, Republic of China

中華民國九十六年六月

在中繼型態WiMAX網路系統中提升傳輸量並保證服務品質的排程法

研究生：莊佳璇　　　　　指導教授：張仲儒

國立交通大學電信工程系碩士班

摘要

為了要以較低的投資成本來擴大基地台(BS)的服務範圍，以及維持良好的訊號雜訊比，使用中繼站(RS; relay station)的技術逐漸成為重要的研究範疇。此外，在現今多媒體通訊的環境下，傳輸服務品質(QoS; Quality of Service)的保證是一個很重要的課題。因此，一個能夠有效使用系統資源，同時能保證QoS的無線資源排程演算法是必要的。

在本篇論文中，我們提出了一個在有中繼型態WiMAX網路系統中，提供下行鏈路服務品質保證及系統效能提升之QoS_GTE排程法演算法。QoS_GTE由傳送時間基準(TT_based)路徑選擇演算法、服務排序基準(SO_based)資源分配演算法，以及同時傳送(TC; transmission concurrency)判別演算法所組成。TT_based路徑選擇演算法選擇傳送時間最短的路徑為傳輸路徑；SO_based資源分配演算法給予較高緊迫性的使用者較高的優先權，使不同層級的使用者可以滿足各自的QoS。在QoS 滿足的同時，又能提高系統的傳輸量。同時傳送判斷演算法則是使系統資源作有效的重複使用。由模擬結果得知，我們所提出的路徑選擇演算法確實比過去文獻中所提出的方法好，不僅系統傳輸量為最高，各層級使用者的服務品質也是最好的；資源分配演算法也在大部分的時候都能達到品質保證；同時傳送判別演算法亦可增加系統之服務量。同時，中繼站的擺放位置對系統的表現也有很大的影響，在我們設定的模擬情境中，中繼站的擺放位置以基地台服務半徑的四分之三為最佳。

QoS_Guaranteed Throughput Enhancement Scheduling in WiMAX Relay-Assisted Network

Student: Chia-Shueng Chuang **Advisor: Chung-Ju Chang**

Department of Communication Engineering
National Chiao Tung University

Abstract

In order to extend the cell coverage and maintain the SINR with low upfront investment, there has been increasing interest in the use of relay station (RS). A downlink centralized QoS guaranteed and throughput enhancement (QoS_GTE) scheduling scheme in a relay-assisted WiMAX network is proposed. The QoS_GTE consists of a transmission time based (TT_based) path selection algorithm, a service order based (SO_based) resource allocation algorithm and a transmission concurrency (TC) decision algorithm. The TT_based path selection algorithm takes path loss, shadow fading, and interference into consideration in terms of transmission time for a given transmission packet. The path with the minimal transmission time will be selected. The SO_based resource allocation algorithm gives high priority to the urgent users. The SO_based resource allocation algorithm provides throughput maximization under QoS satisfaction. The TC decision algorithm carries out flexible resource reuse by deciding which RSs can transmit concurrently using the same frequency and time slots. From the simulation results, we can conclude that the proposed QoS_GTE outperforms the other conventional schemes in terms of system throughput and the satisfaction of QoS requirements. Also, the appropriate location of RS is at three fourths of the cell radius.

誌謝

能夠完成這篇碩士論文，必須要感謝許多人。首先要感謝張仲儒教授這兩年來在專業知識上，以及待人處世方面的諄諄教誨，讓我在學業和生活上都獲益良多。此外，還要感謝特地抽空回來指導我們的芳慶學長，提供我們很多最新的資訊。感謝立峰、詠翰、文祥、志明、耀興、琴雅、俊帆學長姊們，在我遇到困難的時候，提供我寶貴的意見。當然，還要謝謝和我一起努力的世宏、建安、建興、正昕和佳泓，有你們的陪伴，研究的路途一點也不寂寞。謝謝碩一的學弟妹們，謝謝怡潔。

最後，我要感謝我的父母和姊姊，有了他們對我的支持和期待，我才有動力度過每一個難關，完成階段的目標。

莊佳璇 謹誌
民國九十六年七月

Contents

List of Figures

List of Tables

Chapter 1
Introduction

Utilization of multiple radio technologies is a key factor in the realization of the next generation communication networks. In cellular networks, there has been a concern about the limited bandwidth and weak SINR at edges of the cell due to the buildings, distance, and interference and so on. In conventional wireless cellular systems, all terminals are directly connected to the backbone infrastructure via a single hop connection. In order to extend the cell coverage and maintain the SINR with low upfront investment, ad-hoc networks and WiMAX mesh networks have been popular issues (where the traffic signal can be relayed through one or more intermediate node(s) and each node can act as a transmitter) for past few years. However, the routing (relaying node selection) is still a complicated issue in cellular multi-hop networks since each mobile station (MS) may have the chance to be a candidate relaying node for other MS that requires relaying assistance. For the reasons above, there has been increasing interest in the use of professional relay stations (RSs) to increase the capacity and the coverage of cellular networks. The RS is major in relaying traffic to MSs so that MSs do not have to be a part-time RS anymore and this idea reduces the complexity in using relay networks.

It is recognized that the relay structure combines advantages of both cellular and ad-hoc networks. The profit of using a RS is the lower complexity, the less power compared with a BS, and the absence of need for wired backhaul connection [1]. The most important cause of using relaying technique comes from the reduction in the overall path loss between a BS and a MS. Another benefit of relaying is the path diversity gain which can be achieved by selecting the most favorable path in shadowed environment. Several research studies have evaluated the multi-hop approach and demonstrated that it is a viable solution for extending 3G networks to get throughput improvement and good area coverage in use of multi-hop relaying.

Choosing an optimal RS can have effective impacts on the overall performance improvement. In [4]-[5], effects of two-hop RS selection strategies in system performance had been studied. The authors proposed several different schemes for RS selection: selection based on distance, selection based on path loss, and then random selection. These schemes can improve the SINR and the order from best to worst is: selection based on path loss, random selection, selection based on physical distance. From the path loss based schemes, least maximum path loss relaying node selection (LMP) can choose good channels with the percentage of 98%. In [6], the distance based selection is adopted under the assumption that path loss exponent from the BS to the MS is larger than path loss exponent from the BS to the RS. If the distance from the BS to the MS is greater than a certain distance, the BS will use the relay path for its transmission. However, the path loss assumption is not always right in real life. Also, if the path loss exponent from RS to MS is large enough, the total performance of relay path may not be good as it is expected.

Substantial throughput improvement can be additionally obtained by operating the concurrent relaying transmission. Therefore the spatial reuse is very crucial to realize the full potential of relay networks [3], [7]-[9]. In [3] two types of relaying with concurrent transmission were proposed. One is to allow concurrency among the different hops along the path to the same target MS. The other is to allow concurrency among the hops on the different downstream paths to multiple target MSs. The decision of which concurrent transmission method will be adapted is according to the link qualities from the BS to RS and RS to MS for multiple users. In [7], the authors improved the system throughput by finding a set of active links that can maximize the total transmission rate. Note that, although concurrent transmission increases the system throughput, it also raises the interference. Thus a concurrent transmission set, which depends on minimum number of interference link, were proposed in [8]-[9]. However, the computations in these mechanisms are a complex and time consuming work.

In cellular networks, there is a tradeoff between system throughput and the quality of service (QoS). The problem of downlink scheduling scheme, that exploited multi-user diversity to improve the shared channel utilization and the downlink capacity of the cell by reducing the intra cell interference in cellular relay networks, has been studied in [10]. Challa and Cam took the amount of intra cell interference, the instantaneous data rate and the average throughput of the users into account while picking which user and which path (direct or 2-hop path) the BS would transmit to. In their simulation case, so as in [5], there is only one traffic type hence the quality of service (QoS) for different traffic type is ignored. However, Cho and Hass [3] had shown that in the multi-hop cellular networks, BS can choose to utilize the multi-hop relaying instead of the single hop direct transmission. Such a hybrid operation can be exploited to mitigate the unfairness in QoS among the users. Therefore, QoS as well as system throughput in cellular network can be improved through the use of the relaying.

The previous schemes about using relaying have not been implemented in

WiMAX environment but it is proved that the system throughput could be improved. In the thesis, we propose a downlink centralized QoS guaranteed and throughput enhancement scheduling (QoS_GTE scheduling) scheme in WiMAX relay-assisted network. It is proven that the two-hop path is most efficient in multi-hop relaying, [2]-[3], thus the relay paths considered in the paper are all 2-hop relay path.

The QoS_GTE scheduling scheme includes three algorithms. First, the transmission time based (TT_based) path selection algorithm finds a best path for each MS based on the minimum transmission time. For QoS_guaranteed, a service order parameter is defined, which takes both priority and urgency into account as the foundation of service order. Based on service order parameter, the service order based (SO_based) resource allocation algorithm then performs resource allocation and avoids resources occupied by high priority traffics, which leads to lower performance. Finally, the transmission concurrency decision algorithm will perform resource re-allocation to allow several RSs transmit at the same time to take best utilization of the resources.

In QoS_GTE scheduling scheme, TT_based path selection algorithm makes a deliberate judgment about using relaying or not and which RS is responsible by taking the pathloss, the shadow fading and interference into account and these three components decide the transmission time. The SO_based resource allocation algorithm will give priority to MSs whose traffic constraint is going to be violated, which would result in QoS_guaranteed. Later, MSs with best channel condition will get the resource to raise the system throughput. Transmission concurrency decision algorithm uses the RSs location concept to determine which RSs can transmit at the same time without difficult computation in link point of view.

The rest of the thesis is organized as follows. Chapter 2 describes the system model and derives the available data rate. The proposed QoS_GTE scheduling scheme is presented in chapter 3. The simulation results are shown in chapter 4. Finally, the chapter 5 makes the conclusion of the thesis. From the simulation results, we can summarize that the proposed algorithm indeed outperforms other conventional scheme, such as LMP relaying node selection scheme. SO_based resource allocation can all about guarantee the QoS requirements. The system throughput improvement of using transmission concurrency decision algorithm is about 33% in maximum. Also, the suitable location of RS is three fourth of the cell radius.

Chapter 2
System Model

2.1 Environment

Consider a WiMAX relay_assisted cellular network in the urban area, which consists of one base station (BS) at the center of each cell and wireless fixed RSs surrounding the BS uniformly. Also, all RS positions are also on the intersections. As shown in Figure 2.1, a regular Manhattan type of grid in each cell is also considered, where solid line represents the street and the intersections are the crossroads. MSs are uniformly distributed on the street and snap-shot method is used. Assume that all the proposed algorithms are done in the BS, therefore all what RSs have to do is to merely forward the information which is received from the BS. Each RS can only communicate with BS and MSs in the same cell, while, the communication between any two RSs in the same cell is not allowed. Moreover, transmitting and receiving actions do not occur at the same time in both BS and RSs.

We concentrate on the two level interference range wrapping around with nineteen cells, which can imitates the condition with infinite interference from the cells. The frequency reuse factor in the wrapping model is four which is also shown in Figure 2.1.

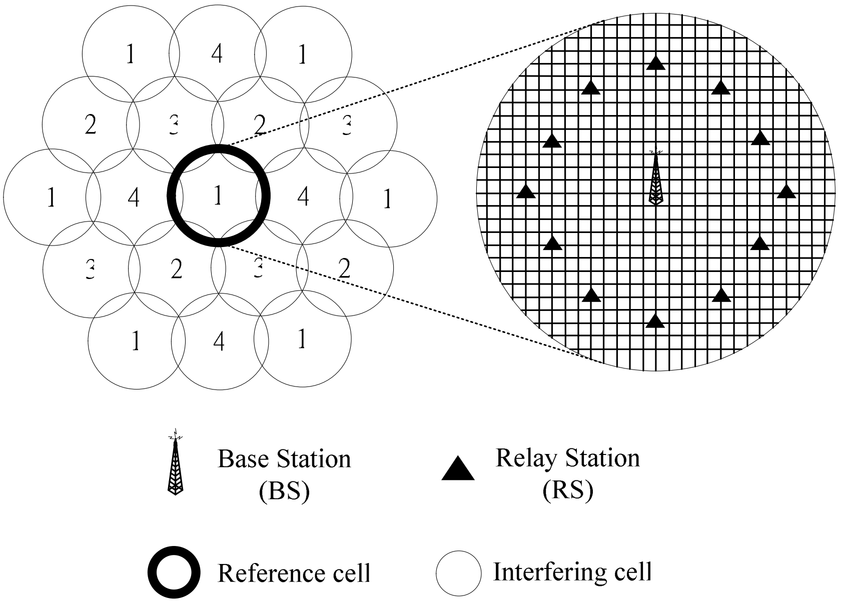


Figure 2.1. The system model

2.2 WiMAX Technology

The WiMAX system is an OFDMA/TDD system. It’s sub-carriers in a sub-channel can be adjacent, as illustrated in Figure 2.2(a), or spread out, as illustrated in Figure 2.2(b). Here the spread out case is adopted, thus the channel condition of each sub-channel can be regarded the same, and the frequency selective condition does not happen. Also, we choose equal power control for all sub-channels. It is because some literatures [11]-[12] have proved that the resource allocation algorithm with equal power distribution over all sub-carrier can perform as well as the optimal power and sub-channel allocation.

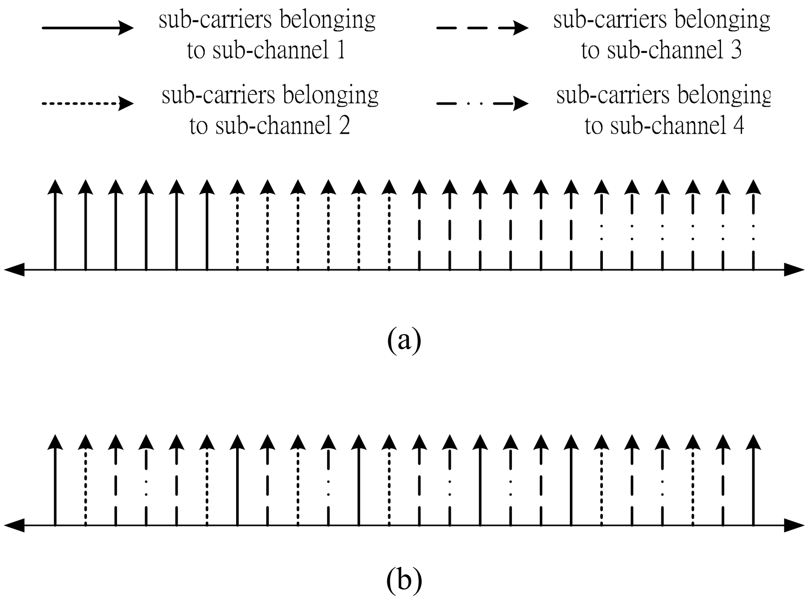


Figure 2.2. (a) adjacent sub-carriers (b) spread sub-carriers

In the WiMAX relay-assisted network, the downlink transmission may involve a direct path or a relay path. A direct-path transmission is that a BS transmits directly to a MS without any intermediate node. A relay-path transmission is that a BS transmits to a RS and the RS forwards the received traffic to the MS. The WiMAX network supports relay with two-hop only because it has been concluded in [2][3] that the throughput enhancement in multi-hop relaying comes mostly from two hop relaying.

Figure 2.3 is a downlink transmission frame in WiMAX relay_assisted system. BS and RSs share the same spectrum through a centralized scheduling for channel access in TDMA. At the start of a downlink frame, BS transmits first and the transmission time for BS is TB. Later, RSs transmit for the rest of the time which is denoted as TR. All RSs share TR in TDMA mode, too. TB and TR are variable in each downlink frame. Flexible boundary between TB and TR is adopted in order to achieve good resource usage.

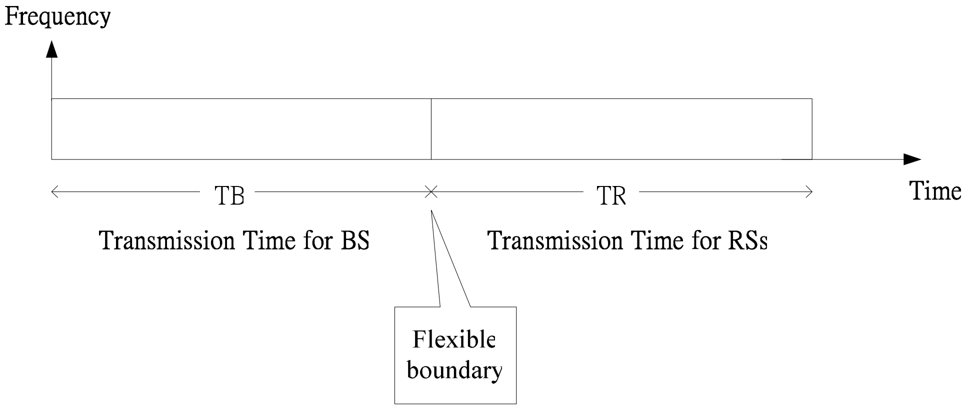


Figure 2.3. A downlink frame in WiMAX relay_assisted system.

2.3 Services

The BS provides one individual finite queue for each MS. The arrival process of packets at each queue for different users is independent from one frame to the next frame, and the packets are served in the first-in first-out manner.

There are three service classes in WiMAX system. Differentiated QoS parameters are considered for different service classes. One service class is the real-time service; its QoS parameters are bit-error-rate (BER), maximum delay tolerance, and maximum dropping ratio. Another class is the non-real-time service; its QoS parameters are BER and minimum required transmission rate. The other class is the best-effort service; its QoS parameter is BER only.

We consider four kinds of traffic types. They are voice traffic of real-time service, streaming video traffic of real-time service, HTTP traffic type of non-real-time service, and FTP traffic of best-effort service. Assume that one user has only one traffic type for his demand. The packets of real-time services will be dropped if the delay of packet exceeds the maximum delay tolerance. On the other hand, packets of non-real-time service are allowed to be queued in stead of being dropped if the occupancy of the buffer is not overflowed.

The voice traffic of real time service [14] is modeled as a classic ON-OFF model as shown in Figure 2.4. This assumption is based on the observation that the human speech is decomposable in talkspurt (ON) and silence (OFF) period, where 1/α and 1/ß are the mean of talkspurt and silence periods, respectively. The ON period and OFF period follow an exponential distribution. Packets are generated during ON period with constant rate and none during OFF periods.

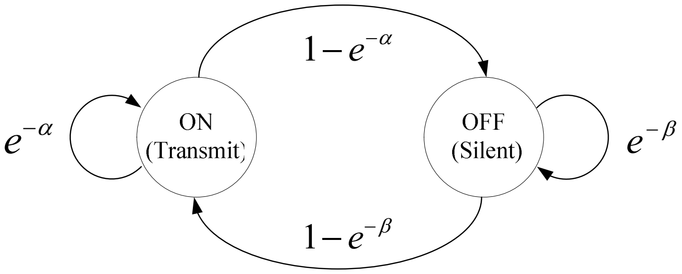


Figure 2.4. ON_OFF model

The second traffic type is the streaming video traffic of real-time service [15].Video traffic is divided into frames. As shown in Figure 2.5, each frame arrives at a regular interval T, and each frame contains a fixed number of packets *N*. The size of these packets, denoted by Y, is distributed as a truncated Pareto, with parameters $\alpha = \alpha_s$, $k = k_s$, and the maximum $m = M_s$. The probability density function (pdf) of Y is given by:

$$f_Y(y) = \begin{cases} \dfrac{\alpha k^{\alpha}}{y^{\alpha+1}}, & k \le y < m, \\ \left(\dfrac{k}{m}\right)^{\alpha}, & y = m. \end{cases}$$

The video encoder introduces delay jitter between the packets of a given frame. These intervals are also modeled by a truncated Pareto distribution, with the parameters $\alpha = \alpha_i$, $k = k_i$, and the maximum $m = M_i$.

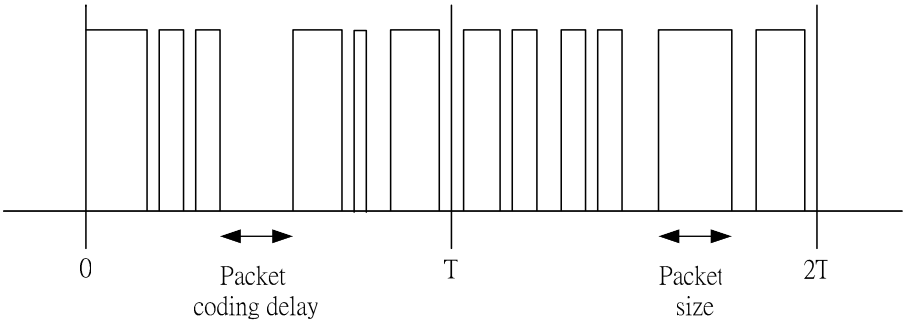


Figure 2.5. Video streaming traffic model

The third one is HTTP traffic type [15] of non real-time service, where the behavior of web browsing is modeled. Figure 2.6 shows the packet trace of typical HTTP traffic, which is a sequence of page downloads. Each page is composed of a main object and several embedded objects and each object is divided into several packets. The main object size is modeled as a truncated lognormal random variable, denoted by *X*, with pdf given by:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma x}\exp\left[\frac{-(\ln x - \mu)^2}{2\sigma^2}\right], \quad m \le x \le M,$$

where the parameters $\sigma = \sigma_m$, $\mu = \mu_m$, $m = m_s$, $M = M_s$. The size of the embedded objects can also be modeled as a truncated lognormal variable, with the parameters $\sigma = \sigma_e$, $\mu = \mu_e$, $m = m_e$, $M = M_e$. The number of embedded objects per main page, denoted by N_e, is truncated Pareto variable with the parameters $\alpha = \alpha_{Ne}$, $k = k_{Ne}$, and $m = M_{Ne}$. Both the reading time and the parsing time can be modeled as exponentially distributed random variable, denoted by Z, with pdf given by:

$$f_Z(z) = \lambda e^{-\lambda z}, \quad z \geq 0.$$

For the reading time, that arrival rate is $\lambda = \lambda_r$, while for the parsing time, the arrival rate is $\lambda = \lambda_p$. Note that the reading time is a heavily user-dependent parameter, and it can vary significantly. The HTTP packet size can be modeled as fixed size, *L*.

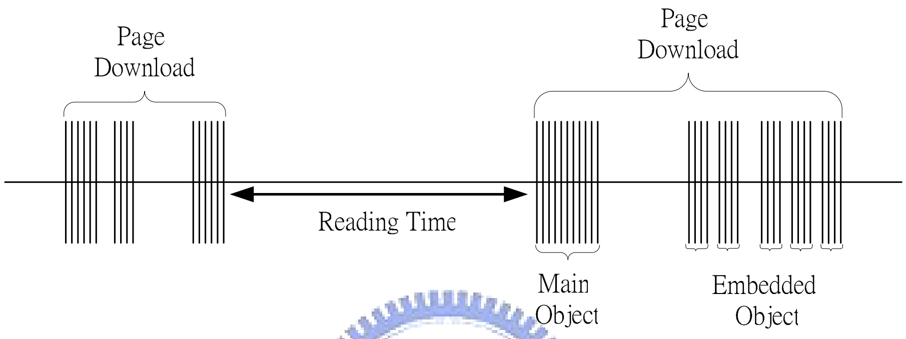


Figure 2.6. Packet trace in typical HTTP traffic

The last one is FTP traffic type [15]of best effort service. Figure 2.7 shows the packet trace of typical FTP traffic, where each FTP user data is modeled as a sequence of file downloads. The size of a file is distributed as a truncated lognormal distribution with parameters $\sigma = \sigma_f, \mu = \mu_f$, and the maximum size M_f. The reading time is distributed in exponential distribution with the parameter $\lambda = \lambda_r$.

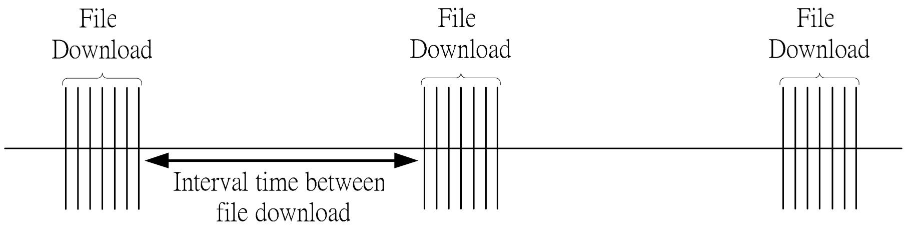


Figure 2.7. Packet trace in a typical FTP traffic

2.4 Channel Model

Assume each BS has complete transmission information at the start of each time frame. The information consists of (i) the transmission rates on all links during the frame that will be scheduled, (ii) the channel state information which includes path-loss and shadow fading, and (iii) QoS requirements of each MS. The transmission is supposed to be with equal power, and the channel state information is assumed constant in a frame but various frame by frame.

The channel is affected by thermal noise, slow fading which includes path-loss and shadow fading. Due to the different transmitter and receiver pair, there are three kinds of channel model. As in Figure 2.8, one is the channel model between BS and MS, another is that between BS and RS, the other is the one between RS and MS [16]. The first two channel models have two factors. One is path-loss which is a function of the distance from BS to the receiver, given by,

$$Pathloss = L_1 + L_2 \log_{10}(d) + 20\log_{10}(f_c / F), \quad (dB), \tag{1}$$

where d is the distance (meters) between the BS and the receiver, f_c is the carrier frequency and L_1, L_2, L_3, F are coefficients. The other factor is shadow fading which is log-normal random variable with zero mean and standard deviation σ. As for the channel model between the RS and MS, the path-loss, denoted by *Pathloss_RM*, is given by:

$$Pathloss\_RM = p_{LOS} \times Pathloss\_LOS + p_{NLOS} \times Pathloss\_NLOS, \tag{2}$$

where p_{LOS} is the probability of line of sight (LOS) condition and p_{NLOS} is the probability of non-line of sight (NLOS) condition. The formula of $Pathloss\_LOS$ is as (1). $Pathloss\_NLOS$ is defined as:

$$Pathloss\_NLOS = N_1 d_1 + N_2 + 20\log_{10}(f_c / F) + (N_3 - N_4 d_1)\log_{10}(d_2), \qquad (3)$$

which is a function of distance along the main street (d_1) and distance along the perpendicular street (d_2) as in Figure 2.8 and N_1, N_2, N_3, N_4 are constant weights. The shadow fading is log-normal random variable with zero mean and standard deviation σ_{LOS} when the link is LOS. The shadow fading is log-normal random variable with zero mean and standard deviation σ_{NLOS} when the link is NLOS.

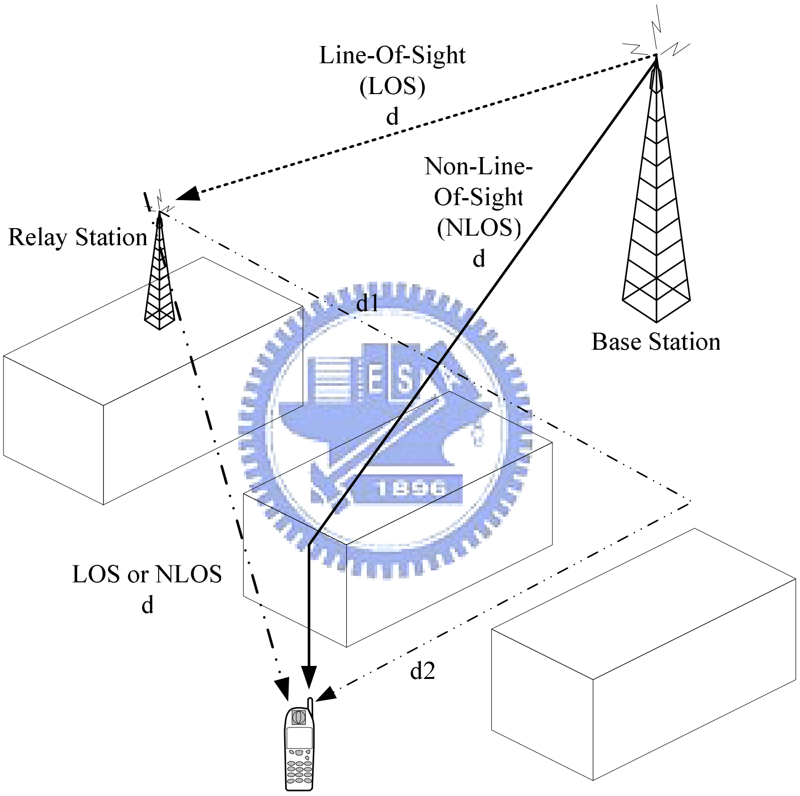


Figure 2.8. The channel classification for the relay system

2.5 Available Data Rate

For three kinds of transmitter and receiver pairs mentioned before, their SINR computation can be given by:

- The SINR from BS to MS *m*

$$SINR_{m,0} = \frac{L_{m,0}P_0}{N_0 + (\sum_{b \in B} L_b P_0) + (\sum_{r \in R,} L_r P_r)}, \tag{4}$$

- The SINR from BS to RS *i*

$$SINR'_{m,i} = \frac{L'_{m,i}P_0}{N_0 + (\sum_{b \in B} L_b P_0) + (\sum_{r \in R,} L_r P_r)}, \tag{5}$$

- The SINR from RS *i* to MS *m*

$$SINR''_{m,i} = \frac{L''_{m,i}P_r}{N_0 + (\sum_{b \in B} L_b P_0) + (\sum_{r \in R} L_r P_r) + (\sum_{c \in C, c \neq i} L_c P_r)}, \tag{6}$$

where P_0 is BS transmit power and P_r is RS transmit power, $L_{m,0}, L'_{m,i}, L''_{m,i}$ are channel gain from BS to target MS, from BS to target RS, and from RS to target MS in the reference cell, respectively, which are determined by shadow fading ×pathloss, N_0 is the thermal noise density, *B* is the set of the BSs in neighbor cells that use the same spectrum as the reference BS does, *R* is the set of active RSs in neighbor cells that use the same spectrum as the reference BS does, *C* is the concurrency set of RSs in the reference cell, and L_b, L_r, L_c are channel gain from other cell BS to reference cell target MS, from other cell RS to reference cell target MS, and from other transmit concurrent RS to target MS in the same reference cell, respectively.

After SINR is determined, the modulation order M_j for each transmission path can be decided [17] as follows:

$$m = K \times SINR + 1, \; with \; K = \frac{-1.5}{\ln(5 \times \varepsilon)}, \tag{7}$$

where ε is the bit error rate requirement. For a given value *m*, we can find its corresponding M_j from the table below.

m	$0 \le m < 4$	$4 \le m < 16$	$16 \le m < 64$	$64 \le m$
M_j	0	4	16	64

Thus available data rate is then defined as:

$$R = \frac{1}{M_j}. \tag{8}$$

Chapter 3
QoS_GTE Scheduling Scheme

As shown in Figure, the proposed QoS-guaranteed and throughput enhancement (QoS_GTE) scheduling scheme contains three parts: the transmission time based (TT_based) path selection algorithm, service order based (SO_based) resource allocation algorithm, and transmission concurrency (TC) decision algorithm.

First, TT_based path selection takes channel state information of all MSs as its input to compute the minimum transmission time for all possible paths and pick up the optimal path for each MS. Under the given path, SO_based resource allocation assigns the transmission time and bandwidth to the MSs according to the either QoS requirements or channel condition of MSs. Eventually, TC decision checks if RSs' transmissions are able to be combined together. If the concurrency occurs, the occupied resource will be released. This released resource can be assigned to other MSs by SO_based resource allocation algorithm again. The loop between the SO_based resource allocation algorithm and TC decision algorithm will stop when there is no released resource anymore.

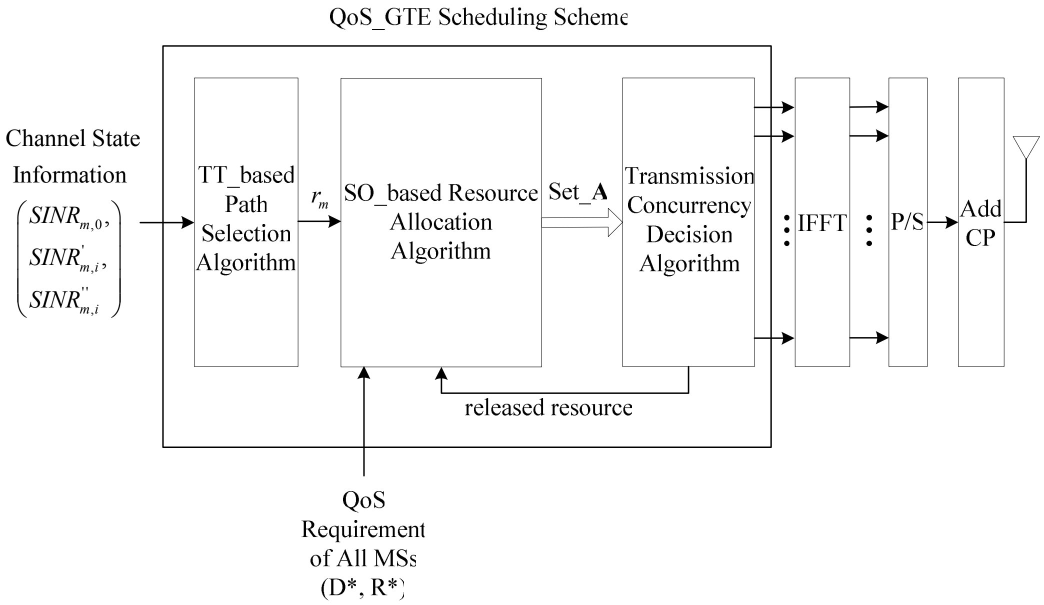


Figure 3.1. The block diagram of QoS_GTE Scheduling Scheme

3.1 TT_based Path Selection Algorithm

Relaying technique can increase the received SINR, however, it occupies two transmission time to transmit the same traffic. BS first transmits the traffic to a certain RS then the RS forwards what it receives to the destination MS. The resource using efficiency for relaying is 1/2. From this point of view, we propose transmission time based (TT_based) path selection algorithm both to decide when the system uses relay assistant and which RS is responsible for the transmission.

Different from other path selection methods based on either distance or path-loss [4]-[5], the TT_based path selection algorithm is based on the total transmission time that a path takes, which takes the distance, the path loss, the shadow fading and the interference all into consideration. To make the system throughput high, it is necessary to transmit information within as minimal transmission time as possible. TT_based path selection algorithm intends to find the minimum transmission time path for a MS_m from all possible paths at the start of each frame with the objective to increase the system performance.

TT_based path selection algorithm first computes the transmission time for each possible path. Denote $R'_{m,i}$ the available data rate for MS_m from BS to RS_i. $R''_{m,i}$ is the available data rate for MS_m from RS_i to MS_m. Then the total transmission time from BS through RS_i to MS_m $t_{m,i}$ is

$$t_{m,i} = \frac{1}{R'_{m,i}} + \frac{1}{R''_{m,i}}, i \neq 0 . \tag{9}$$

Denote $R_{m,0}$ the available data rate through direct path, the transmission from BS to MS_m directly, so the transmission time from BS to MS_m $t_{m,o}$ can be

$$t_{m,0} = \frac{1}{R_{m,0}} . \tag{10}$$

After transmission time calculating, TT_based path selection algorithm will select a suitable path which transmission time is the minimum for MS_m.

Figure shows the detail TT_based path selection algorithm process for M MSs. There are $I_m \in \{0,...,i\}$ possible transmitters for MS_m, including *i* RSs and a BS. After getting (*i*+1) different path transmission times, the path with the minimum transmission time is to be chosen. r_m represents the inter node index, that is, remarking which RS is selected for the transmission to MS_m. For instance, $r_m = i$ means BS will transmit to RS_*i* then RS_i is going to forward the information to the MS_m. If $r_m = 0$, it means the BS will transmit to MS directly through no RSs.

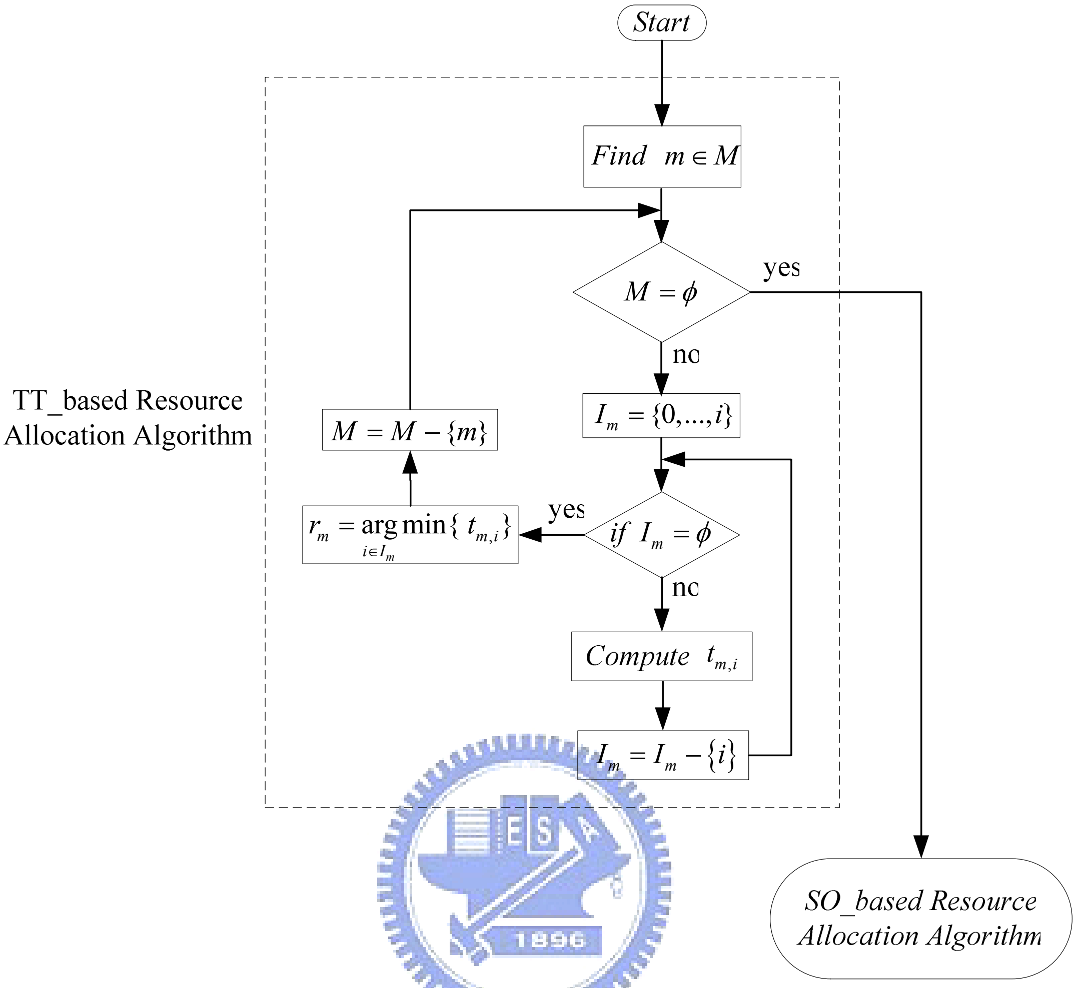


Figure 3.2. The flow chart of TT_based path selection algorithm

3.2 SO_based Resource Allocation Algorithm

Since there is a tradeoff between QoS guarantee and system throughput, the proposed service order based (SO_based) resource allocation first serve MSs which are not QoS satisfied to guarantee the QoS; later, MSs with good channel conditions will be served to maximize the system throughput. The resource allocation is start after path selection.

3.2.1 Definition of S_m

In order to maintain the QoS requirement, we define a service order parameter to combine the priority and urgency together for the purpose of lowering the probability that the traffic will violate the QoS constraints. The service order parameter of MS_m at the beginning of the x^{th}-frame, denoted by $S_m(x)$ is defined as,

$$S_m(x) = \frac{p(x)}{p^*} + \frac{u(x)}{u^*}, \tag{11}$$

where $p(x)$ is the priority value of the packet to MS_m at the beginning of the x^{th}-frame, $u(x)$ is the urgency value of the packet to MS_m at the at the beginning of the x^{th}-frame, p^* is the maximum value of $p(x)$ and u^* is the maximum value of $u(x)$.

Why we consider both priority and urgency is that if only priority value is considered, low priority traffics are backlogged as long as there are traffics with priority higher than them even though these low priority traffics have waited for a long time or are going to violate the constraints. In order to avoid the unfair condition above, we advocate the low priority traffic deserves being served when it is in an emergency condition. We add normalized priority value and normalized urgency value together to decide which traffic should first be served impartially.

It has been mentioned that there are four traffic types considered. Each traffic type is given an integer priority value $p(x)$. The voice traffic is assigned with the highest priority and its priority value is four. The second higher one is the video streaming traffic and its priority value is three. The third one is the HTTP traffic and its priority value is two. The last one is FTP of best effort traffic and its priority value is one. Assume that each MS can request only one type of traffic service at each time.

However, it would be different among MSs, which have experienced various delay, even these MSs have same priority value. The urgency value *u(x)* is brought out. The higher urgency value *u(x)* of a MS implies that this MS's packet is having higher possibility to violate the QoS constraints. The *u(x)* plays an important role in QoS_guaranteed issue and *u (x)* is defined as,

$$u(x) = l(x) + t(x), \tag{12}$$

The *l(x)* is the urgency value for a packet contributed by the remaining bits which are not transmitted yet at the beginning of the x^{th} –frame, which is definition as:

$$l(x) = \frac{L - C_{x-1}}{L}, \tag{13}$$

The *L* is the size of the packet and C_{x-1} is the total bits that have been transmitted at the end of the $(x-1)^{th}$–frame. The *t(x)* is the urgency value for a packet contributed by the delay time at the beginning of the x^{th} –frame.

In the following, the *t(x)* of real-time traffic is designed. D^* is QoS maximum delay tolerance requirement in the unit for frame. We separate the D^* into U^* levels. *t(x)* is set to be one when the delay of the packet at the x^{th} –frame, *d(x)*, is not greater than $D^*/2$, which implies this packet is the less urgent one. Then we divide the other $D^*/2$ range equally into (U^*-1) parts and the value of *t(x)* is set to be two, three… U^*, respectively. The packet with *t(x)*=U^* is approaching the delay upper bound D^* and is so urgent to be served immediately. Notice that the method above only works when the delay constraint D^* is larger than the level of *t (x)*, denoted as U^*. This concept can be expressed in mathematic modeling as below.

When $D^* > U^*$

$$t(x)=\begin{cases} U^*, & \text{if } D^*-\left(\frac{1}{U^*-1}\times\frac{D^*}{2}\right)<d(x)\le D^* \\ U^*-1, & \text{if } D^*-\left(\frac{2}{U^*-1}\times\frac{D^*}{2}\right)<d(x)\le D^*-\left(\frac{1}{U^*-1}\times\frac{D^*}{2}\right) \\ U^*-2, & \text{if } D^*-\left(\frac{3}{U^*-1}\times\frac{D^*}{2}\right)<d(x)\le D^*-\left(\frac{2}{U^*-1}\times\frac{D^*}{2}\right) \\ \vdots & \\ 2, & \text{if } D^*-\left(\frac{U^*-1}{U^*-1}\times\frac{D^*}{2}\right)<d(x)\le D^*-\left(\frac{U^*-2}{U^*-1}\times\frac{D^*}{2}\right) \\ 1, & \text{if } d(x)\le\frac{D^*}{2}. \end{cases} \tag{15}$$

If we use the same mechanism to divide the urgency levels when the delay constraint D^* is not larger than U^*, it happens that the coverage of some level is a fraction of a frame. Since our minimum unit in scheduling is one frame, a fraction of a frame will never occur. For this reason we modified our mechanism when D^* is not larger than U^*. This concept can be expressed in mathematic modeling as below.

When $D^* \le U^*$

$$t(x)=\begin{cases} U^*, & \text{if } d(x)=D^* \\ U^*-1, & \text{if } d(x)=D^*-1 \\ \quad \text{M} & \\ U^*-D^*, & \text{if } d(x)=0. \end{cases} \tag{16}$$

As decreasing D^* by one frame, $t(x)$ also decreases one, $t(x)=U^*-1$, and so on till x = 0. There are total D^*+1 levels in $t(x)$ and the difference of delay constraint between any two adjacent levels is one frame, the minimum scheduling unit.

For non real-time traffic, the QoS constraint is the minimum required transmission rate R^*. If a packet with size L is transmitted at the rate R^*, the transmission time for that packet is,

$$D^*=\frac{L}{R^*}\times N \quad \text{(frames)}, \tag{17}$$

where N is the number of frame per second. In other words, D^* is the maximum transmission time for a packet with size L. After having D^*, all the process of defining $t(x)$ is the same as real-time traffic.

3.2.2 Resource Allocation

For the purpose of QoS guarantee, the MSs can be served first when their $S_m(x)$ are larger than service order threshold S_{th}. Also, the service order of these high $S_m(x)$ MSs is according to the $S_m(x)$ in descending way, which means that a MS with the higher $S_m(x)$ is able to get what it needs earlier than the one with lower $S_m(x)$. So S_{th} is a service order lower bound to make sure that the QoS is maintained as possibly as it can. If there is remaining resource after the first step, best first resource allocation is adopted, instead of using $S_m(x)$ to decide the next served MS. Residual free resource is allocated to other MSs who can get the maximum transmitted bits in that channel to enhance the total throughput.

As in Figure, SO_based resource allocation algorithm computes $S_m(x)$ for MS_m in user_set M, denote M is the set of MSs to be served. Therefore the MS which $S_m(x)$ is greater than S_{th} will be served at first. Later, if there is remaining resource, best first allocation makes use of the remaining resource as well as possible.

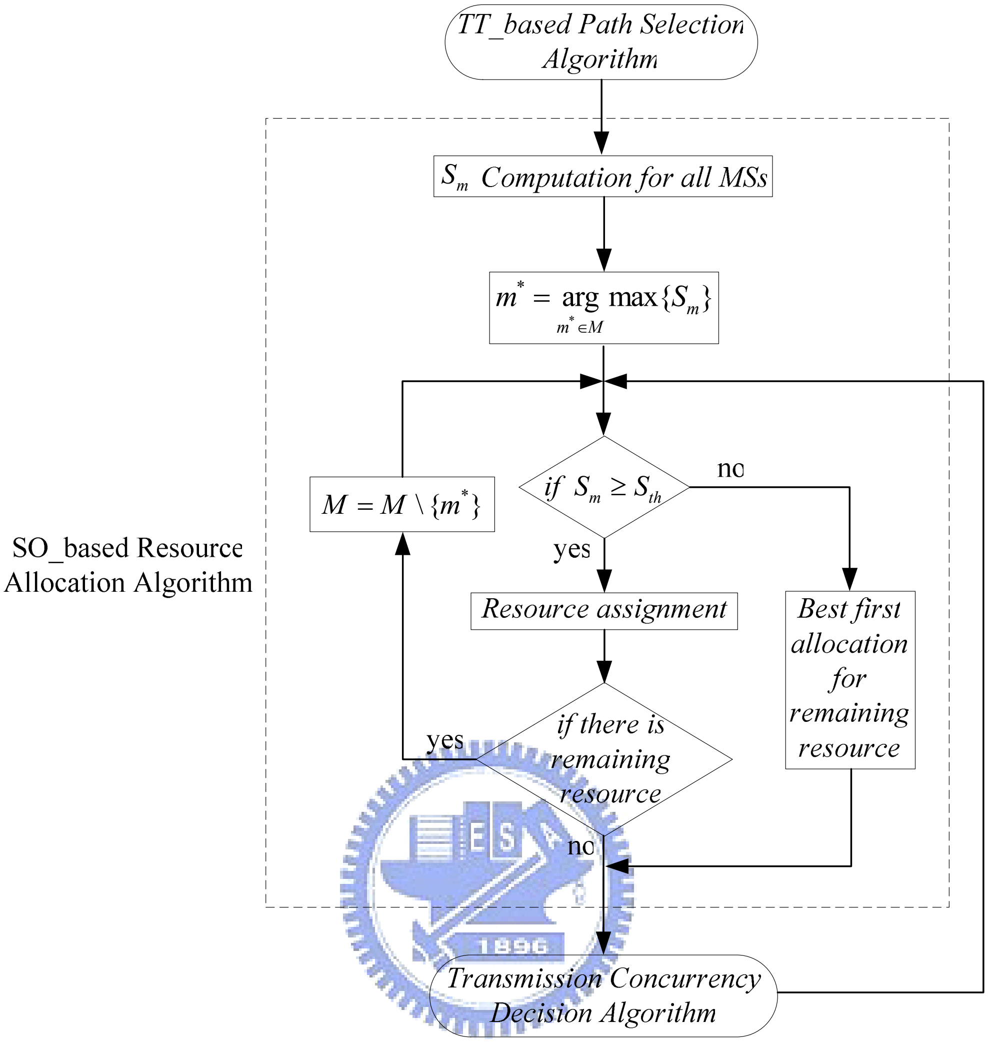


Figure 3.3. The flow chart of SO_based resource allocation algorithm.

3.3 TC Decision Algorithm

Since using relaying may cause resource inefficiency, how to reuses the resource is becoming a critical topic to enhance the system throughput. The TC decision algorithm aims to find which RSs can transmit at the same time using the same resource. Therefore, the original occupied resource can be reserved to other transmitters so that using resource efficiency can be improved, which compensates the drawback of using relaying.

After the SO_based resource allocation algorithm, which RSs are active in the frame has already been decided. Denote $I=\{1,\ldots,i\}$ is the set of active RSs. i is the index of RS. We have to build a concurrency candidate set A_i directing at each active RSi. Any active RSa whose angle separated from RSi is not less than 120° is included into A_i. In other words, RSa is the concurrency candidate of RSi. For an example, in Figure(a), assume all RSs are active so the concurrency candidates for RS1 are RS5, RS6, RS7, RS8 and RS9, that is, $A_1=\{5,6,7,8,9\}$. Notice that there are at most five concurrency candidates for RSi.

After candidate selection, picking up the exact concurrency RSs from the candidate set is followed. If A_i includes the two candidate RSs which are 120° separated from RSi, the set α_{120} is includes these two RSs and these two RSs are chosen as concurrency RSs to transmit with RSi at the same time. If $\alpha_{120} = \phi$, which means the two 120° RSs are not active at the same time, there is only one candidate RSa in A_i with maximum traffic load Δ_a selected as concurrency RS a_i for RSi. As the example illustrated above, RS5 and RS9 are both in A_i so they are concurrency RSs for RS1. The transmit time T for the transmit concurrency RSs is decided by the maximum transmission time of transmit concurrency RSs.

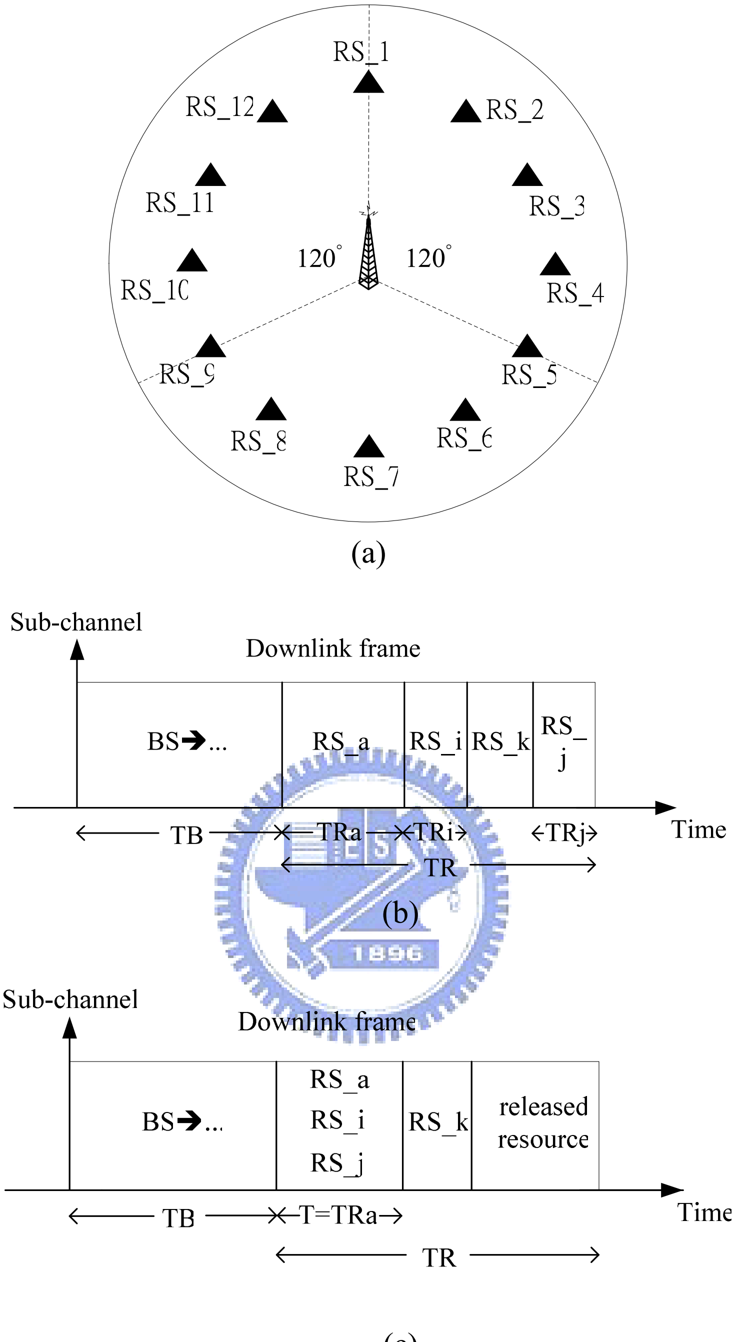


Figure 3.4.Transmission concurrency example. (a) candidate selection (b) an OFDMA frame before concurrency (c) an OFDMA frame after concurrency

Figure(b) shows an OFDMA frame before concurrency and suppose RS*a*, RS*i*, and RS*j* can transmit concurrently. After transmission concurrency, as Figure(c), RS*a*, RS*i*, and RS*j* are combined together to transmit, which makes original-occupied resources released. The transmission time T is set to be TRa since TRa is the maximum transmission time among TRa, TRi and TRj. The released resources can be utilized again and the assignment of these resources is controlled by SO_based resource allocation algorithm. The detail flow chart of TC decision algorithm is described in Figure. The loop between the SO_based resource allocation algorithm and the TC decision algorithm will never stop until there is no resource released after the TC.

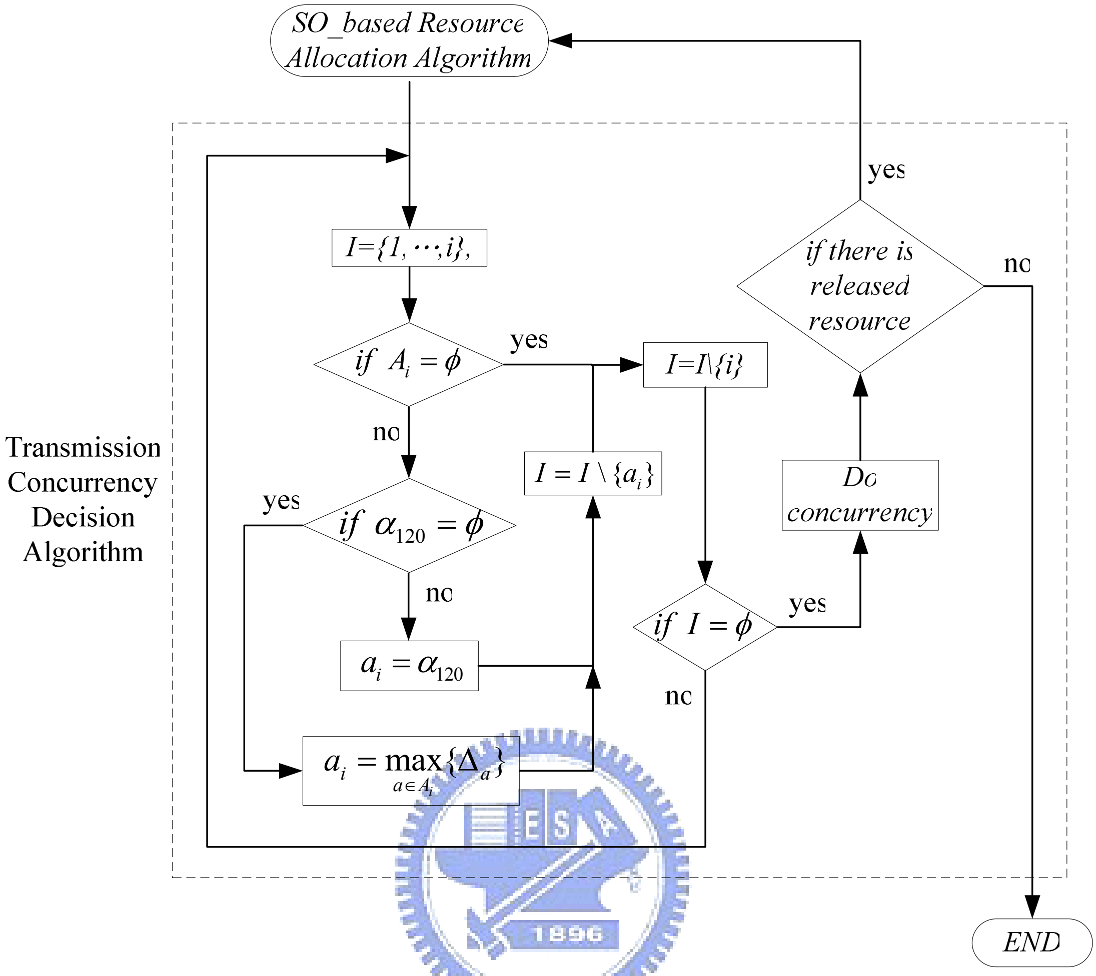


Figure 3.5. The flow chart of TC decision algorithm.

Chapter 4

Simulation Results and Discussions

In the simulations, the WiMAX relay-assisted network is set to be compatible with IEEE 802.16 standard. System parameters are listed in Table 4.1, and scalable parameters in physical layer are modified according to the suggested values in [18]. The OFDMA system is based on 5 MHz bandwidth and the frame duration is 5 ms. The number of sub-carrier is equal to the FFT size, 512, but only 384 sub-carriers are used for data transmission, while the others are used for pilot channel or guard channel. The cell size is set to be 1500 m and the RS coverage is 500 m. BS power is 25 W and each RS power is 5 W and equal power is supposed for each sub-channel.

Table 4.1. System Parameters

Parameters	**Value**
BS cell size	1500 m
RS cell size	500 m
Number of antenna at BS	1
Frame duration	5 ms
System bandwidth	5 MHz
Carrier frequency (f_c)	3.5 GHz
FFT size	512
Sampling frequency spacing	11.16 kHz
OFDMA symbol duration	100.8 μs
Useful symbol time	89.6 μs
Guard time	11.2 μs
Number of data subcarriers	384
Number of subchannels	24
Number of data subcarriers per subchannel	16
Number of OFDMA symbol for downlink transmission per frame	24
Thermal noise density (N_0)	-174 dBm/Hz
Power of BS	25 W
Power of RS	5 W

4.1 Manhattan Street Like Environment

A Manhattan street scenario is considered. The width of each street is 30 meters. The separation of two adjacent parallel streets is 200 meters. The probability of being LOS and NLOS are given as:

$$\begin{cases} P_{LOS}(d) = 1 & \text{if } d \le 15 \text{ m} \\ P_{LOS}(d) = 1 - \left(1 - \left(1.56 - 0.48\log_{10}(d)\right)^3\right)^{\frac{1}{3}} & \text{if } d > 15 \text{ m} \\ P_{NLOS} = 1 - P_{LOS}\ . & \end{cases} \qquad (18)$$

The models of path loss and shadowing are defined differently with respect to the transmitter and receiver pair, BS to MS, BS to RS, and RS to MS, which has already mentioned in (1), (2), (3) in Section 2.4. The corresponding path loss parameters for BS to MS channel, BS to RS channel, and RS to MS channel corresponding path loss parameters are listed in Table 4.2 and Table 4.3. The MS position is distributed uniformly and randomly. By snap-shot method, the MS position will be re-generated after every 10 seconds so that channels also vary after each 10 seconds. Channels are assumed to be fixed within 10 seconds.

Table 4.2. Path-loss and shadow fading parameters

Path loss Parameters	**Path loss of the channel from BS to MS**	**Path loss of the channel from BS to RS**
L_1	38.4	36.5
L_2	35	23.5
F	5	2.5
σ	8 dB	3.4 dB

Table 4.3. Path-loss parameters of channel from RS to MS

Path-loss Parameters	**Value**	**Path-loss Parameters**	**Value**
L_1	41	N_4	0.024
L_2	22.7	F	5
N_1	0.096	σ_{LOS}	2.3 dB
N_2	65	σ_{NLOS}	3.1 dB
N_3	28		

4.2 Traffic Models and QoS Requirements

The QoS requirements and detailed parameters for the four traffic models are given in this section. Each voice user is modeled as an ON-OFF model, in which the length of ON period and OFF period follows the exponential distribution with mean 1 second and 1.35 seconds, respectively [14]. During ON period, the mean data rate is 12.2 kbps. The video streaming traffic consists of a sequence of video frames which are emitted regularly with interval 100 ms. the video traffic model parameters are listed in Table 4.4.4, where the source data rate is 64 kbps.

Table 4.4. Video streaming traffic model parameters [15]

Parameters	**Value**	**Parameters**	**Value**
T	100 ms	M_s	250 bytes
N	8	α_i	1.2
α_s	1.2	K_i	2.5 ms
K_s	40 bytes	M_i	12.5 ms

Parameters for HTTP traffic model are given in Table 4.5 [15]. Note that each HTTP packet shall be smaller than 1500 bytes. As to FTP traffic, the size of each file is modeled as truncated lognormal distribution with mean 2 MB, standard deviation 0.722 MB, and a maximum value 5 MB. The arrival interval between two sequential files is exponentially distributed with mean 180 seconds. The parameters of FTP are defined in Table 4.6 [15].

Table 4.5. HTTP traffic model parameters

Parameters	**Value**	**Parameters**	**Value**
σ_m	25032 bytes	M_e	2M bytes
μ_m	10710 bytes	α_{Ne}	1.1
m_s	50 bytes	K_{Ne}	2
M_s	2M bytes	M_{Ne}	53
σ_e	126168 bytes	λ_r	0.033
μ_e	7758 bytes	λ_P	7.69
m_e	50 bytes	L	1500

Table 4.6. FTP traffic model parameters

Parameters	**Value**
σ_f	0.722M bytes
μ_f	2M bytes
M_f	5M bytes
λ_r	0.006

Table 4.7 lists the QoS requirements of each traffic types and the he transmission time for a frame is 5 ms. In addition, the minimum required transmission rate of HTTP users (in non-real-time services) is slightly larger than the arrival rate of HTTP traffic in page download. This means transmission rate for download of a web page is guaranteed and hence the response time of a web browser is small.

Table 4.7. The QoS requirements of each traffic type

Traffic Type	**Requirement**	**Value**
voice (real-time)	required BER	10^{-3}
	maximum delay tolerance	40 ms (8 frames)
	maximum allowable dropping ratio	1%
video (real-time)	required BER	10^{-4}
	maximum delay tolerance	10 ms (2 frames)
	maximum allowable dropping ratio	1%
HTTP (non-real-time)	required BER	10^{-6}
	minimum required transmission rate	100 kbps (500 bits per frame)
FTP (best-effort)	required BER	10^{-6}

4.3 Performance Evaluation

In two-hop relay network, it is proven that the least maximum path loss (LMP) relaying node selection scheme [5] has the highest probability in selecting the good channel, which is based on a route that has the lowest bottleneck (in terms of path loss). Let PL_{n1}, PL_{n2} denote the path loss in dB associated with the first and the second hop, respectively, along the nth route. Then the selected route, r_s is determined as follows:

$$r_s = \arg\min_{n\in RSs}\left[\max\left(\text{pathloss of the first hop , pathloss of the second hop}\right)\right].$$

We make some modification on LMP scheme in the simulations. That is, after choosing a two hop route r_s, it has to compare its bottleneck path loss again with the direct path pathloss. The route with lower bottleneck path loss is finally chosen.

In the simulations, the number of user is composed of four service traffic types with equal percentage. We define the traffic load as the ratio of the total average rate of all users over the system maximum transmission rate. The maximum transmission rate is achieved when Q users are multiplexed for each sub-channel and the highest modulation order is used for all users. It is equal to 11.74 Mbps in the simulation environment of this thesis. Note that the average arrival rate of voice, video, HTTP, and FTP users is equal to 5.2 Kbps, 64 Kbps, 14.5 Kbps, and 88.9 Kbps, respectively. Thus, the traffic load varies from 0.15 to 0.9 as the number of users varies from 40 to 240.

For fair comparison, the scheduling algorithm of each relay scheme is SO_based resource allocation algorithm. The level of $t(x)$, denoted as U^*, is equal to 6. The threshold value S_{th} is set to be the average of maximum and minimum S_m of HTTP traffic (S_{th}=1.0714). Although HTTP is the third priority traffic, HTTP should also own a service right to guarantee the QoS as possibly as it can when HTTP is facing some kind of time emergency. This is why we set S_{th} as the average of maximum and minimum S_m of HTTP traffic.

4.3.1 Performance Evaluation on Relay schemes

In this simulation, twelve RSs are surrounding the BS uniformly in a circle and RSs are at the cell radius of three fourths. While the desired site location of three-fourth cell radius is not exactly at an intersection, the RS would be placed on the nearest intersection from the desired site. In the following discussions, we compare the proposed QoS_GTE with LMP, LMP with TC, and without relay (w/o Relay), respectively.

Figure shows the modulation order distribution on radio links for TT_based path selection algorithm, LMP path selection scheme, and w/o Relay technique. We can observe that the probabilities of QPSK for TT and LMP are both much lower than w/o Relay and the probabilities of 64_QAM in these two relay schemes are greater than w/o Relay. Moreover, the relay 1_hop links and 2_hop links are almost used in 64_QAM. These is because channels from BS to RSs are usually LOS which causes lower path-loss and RSs re-transmit the information received from BS which can strengthen the transmitted signal power. Therefore, relaying SINR at the receiver would be larger than the no relaying SINR and the modulation order of relay schemes is intuitively higher than the without relay case.

Figure is the average modulation order on system. The overall path modulation order and transmission efficiency are considered to obtain the average modulation order on system. The average overall modulation order of a relay path is the average of its 1_hop link and 2_hop link. However, as using a relay path, it takes two transmission times, which are transmission time from BS to RS and the time from RS to MS, to send the same information. The re-transmission reduces the system efficiency half. Therefore, the average modulation order on system is half of the average overall modulation order when using relay path. Since paths in w/o relay are all direct paths, the average links modulation order for w/o Relay is the same as the average modulation order on system. From Figure we can observe that LMP average system modulation order is smaller than w/o Relay case, which is the contrary of Figure. It has been mentioned that the efficiency of using relay is half of without relay due to the re-transmission in relay path. From this point of view, it can be said that using relay improperly is not good for the system performance. Compare the average system modulation order of two relaying schemes. TT is more than LMP. It is because that TT decides the path by transmission time, which consists of whole path of path loss, shadowing, and interference, rather than according to the path loss of one hop only. TT selects a path more soundly and accurate in system than LMP does.

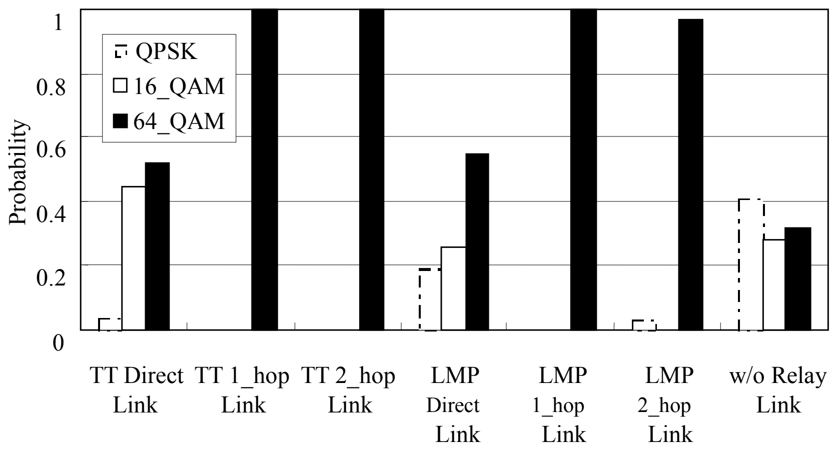


Figure 4.1. Modulation Order Distribution on Radio Links

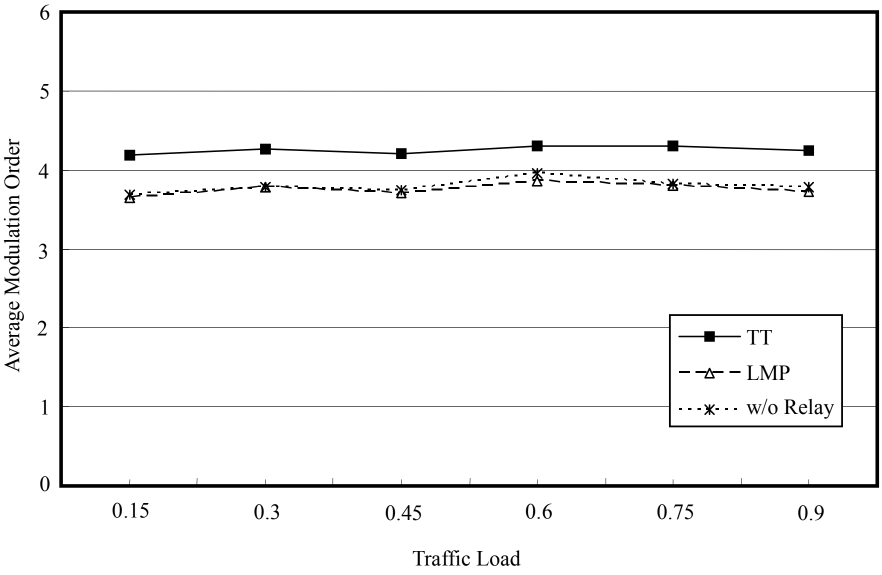


Figure 4.2. Average Modulation Order on System

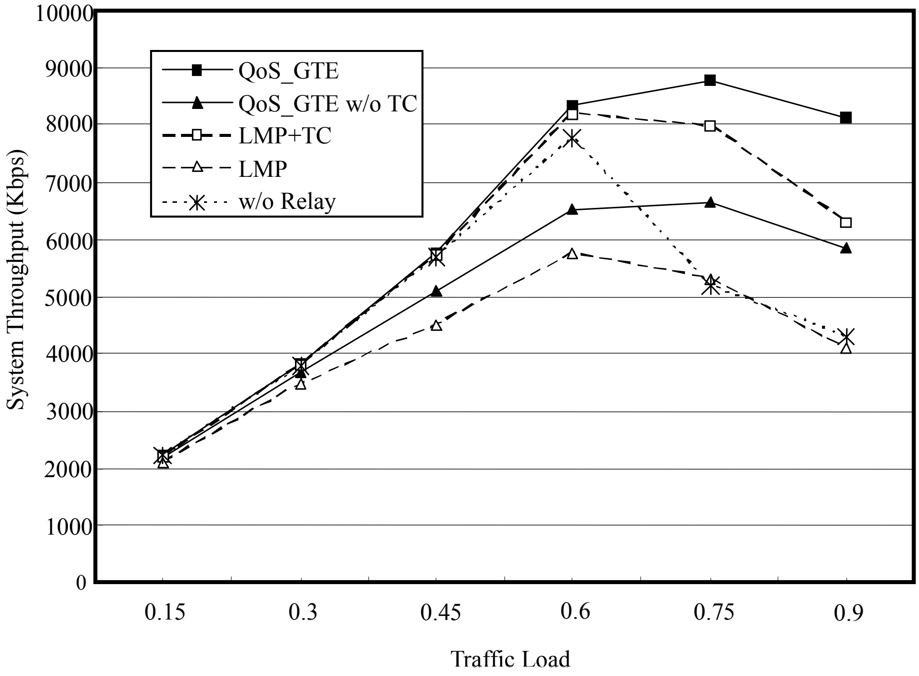


Figure 4.3. System Throughput

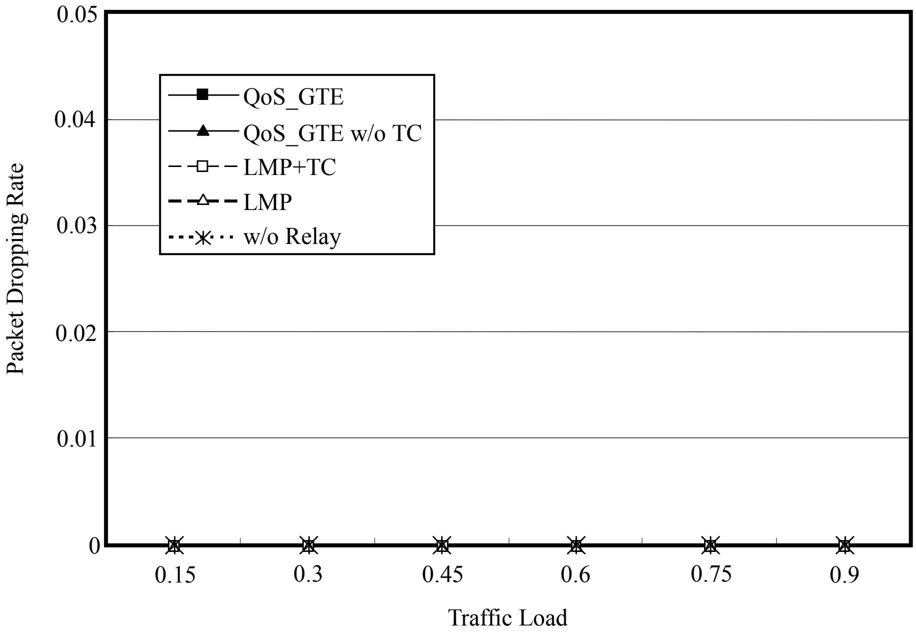
Figure shows the system throughput versus the traffic load of the QoS_GTE, QoS_GTE w/o TC decision algorithm, LMP with TC decision algorithm (LMP+TC), LMP only, and without relay (w/o Relay) scheme. QoS_GTE outperforms the other four schemes. The TT chooses the most efficient and good channel condition path. By means of choosing the minimum transmission time path, the system resource using efficiency is taken into account. No matter the direct path or the relay path, it is the most suitable path for system not only due to its higher modulation order. The TC decision algorithm make the several RSs transmit concurrently so that the resource can be fully used. Since the SO_based resource allocation is used, it not only guarantees QoS but also maximizes the throughput. Since the LMP scheme selects the path only based on the bottleneck path-loss, it does not consider the overall path situation and neglects the effects on system while using relay path. This is why the performance of TT scheme is better than LMP scheme.

The throughput of LMP and TT are less than w/o Relay when the traffic load is below 0.75. This is because that twelve RSs share the resource in TDMA mode. In TDMA mode, each RS transmits the information in different symbol time in spite of the usage of sub-channels. If the sub-channels in a symbol time for a certain RS cannot be used entirely, the spare resource is not allowed to be used by other RSs, which results in resource consuming. This is why the relay schemes with good channel quality but poor throughput performance. Therefore, TC is undoubtedly the main factor to increase the system throughput in using relay. By several RSs transmitting concurrently, TC advances the system efficiency and improves the system throughput.

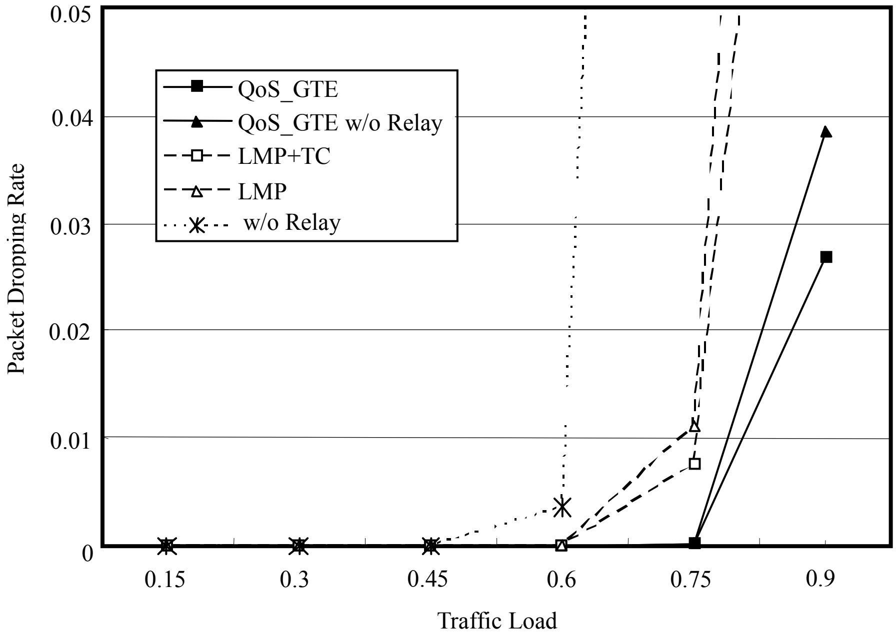
As the traffic load keeps increasing, the throughputs of these five algorithms start to decrease. LMP, LMP+TC, w/o Relay begin to fall down when traffic load is 0.6 and QoS_GTE and TT decline when the traffic load is 0.75. The reason of throughput reducing is that resource is used for QoS guarantee. When the traffic becomes heavy, more urgent packets are waiting to be transmitted. Using SO_based resource allocation, the urgent packets whose S_m is greater than S_{th} will be served no matter how its channel condition is. Consequently, the system throughput begins to reduce because the packets with good channel condition may possibly not be served. Since TT scheme can find the most suitable path by considering the resource using efficiency in terms of transmission time, it can fully utilize the resource so that the system is able to served more users which delay the dropping point of the throughput performance.

Figure(a) and (b) depict the voice and video packet dropping rate, respectively. The voice packet dropping rates are almost zero despite the varying of traffic load. As the voice packets are urgent, they have the highest S_m and SO_based algorithm will first serve these voice packets as soon as possible. In other words, the resource belongs to voice packets users prior. On the other hand, the video packet dropping rates are almost zero for light traffic load. The video packet dropping rate of QoS_GTE and QoS_GTE w/o TC start to increase when the traffic load is 0.75 and exceed the maximum allowable dropping rate (1%) at traffic load with 0.9. For LMP and LMP+TC, their video packet dropping rates increase for the traffic load higher than 0.6 and exceed the maximum allowable dropping rate at traffic load with 0.9. For the case of w/o Relay, its video packet dropping rate increases for traffic load higher than 0.45 and exceeds the maximum allowable dropping rate at traffic load with 0.6. Since a voice packet has higher priority than video one, S_m of a voice packet is higher than the one of a video packet when they are in the same urgent situation. By SO_based algorithm, voice packet is served first due to higher S_m. On the contrary, video packet has to wait until there is remaining resource after voice packet being served. As traffic load rises, and the remaining resource becomes less so that video packet dropping rate grows rapidly. Again, TT can choose a good path as more accurately as others do. It can improve the link quality and the system throughput much better than LMP so the video packet dropping rates of LMP and the case of w/o Relay are much higher than the one of TT.

Figure illustrates the non-guaranteed ratio for HTTP traffic. Unlike the real time traffic, packets of HTTP users will not be dropped but still waiting for service when they cannot reach the minimum transmission rate. In Figure, the guaranteed ratios for HTTP traffic are almost the same when the traffic load is light. As the traffic load becomes higher, the non-guaranteed ratio of w/o Relay rises steeply and those of TT, TT+TC keep low. Since we set S_{th} equal to the average of HTTP maximum and minimum S_m, about half of the HTTP packets with S_m lower than S_{th} is probably not be served priory. Only when HTTP packet is very urgent and its S_m is higher than $S_{th,}$ the HTTP packet may have a chance to be served at the first stage of SO_based algorithm. Moreover, S_m of voice packets and video packets are usually greater than HTTP packets'. Urgent HTTP packets also have to be waiting for service until real time services are almost done. This is why the non-guaranteed ratio maintains in 2% at light traffic load. As traffic load grows, the non-guaranteed ratio increases severely because the resource is allocated to the real time service to avoid larger dropping rate.



(a) Voice Packet Dropping Rate



(b) Video Packet Dropping Rate

Figure 4.4. Packet Dropping Rate of Real Time Services

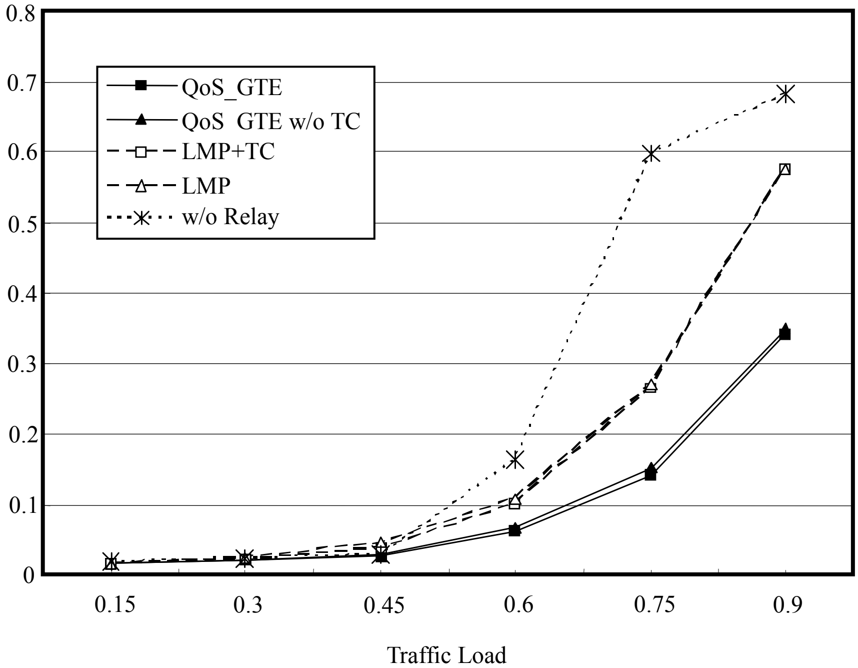


Figure 4.5. Non-guaranteed Ratio of HTTP User

Figure shows the average transmission rate of FTP. Since the S_m of FTP packets are all set to be 0.25 and this value is lower than S_{th}, FTP packets will be transmitted only in best effort manner. The FTP transmission rate is nearly the same at light traffic load. As the traffic load increases, the average transmission rate increases until a certain traffic load then begins to reduce. w/o Relay, LMP, and LMP+TC start to reduce when traffic load is 0.6 and QoS_GTE and QoS_GTE w/o TC fall down at 0.75. As FTP packet being transmitted, it indicates that the channel condition of the FTP packet is the best compared with other un-served packets because S_m of FTP is the lowest and FTP packets are served only in best first allocation. More FTP transmission represents that more resource is used to offer users whose channel condition is good. As a result, the system throughput can be enlarged by these high modulation order FTP users. Besides, the FTP traffic arrival rate is almost half of the total traffic arrival rate in our simulation which may cause FTP dominates the throughput as long as the FTP packets are served. This is why the trend of Figure is the same as Figure.

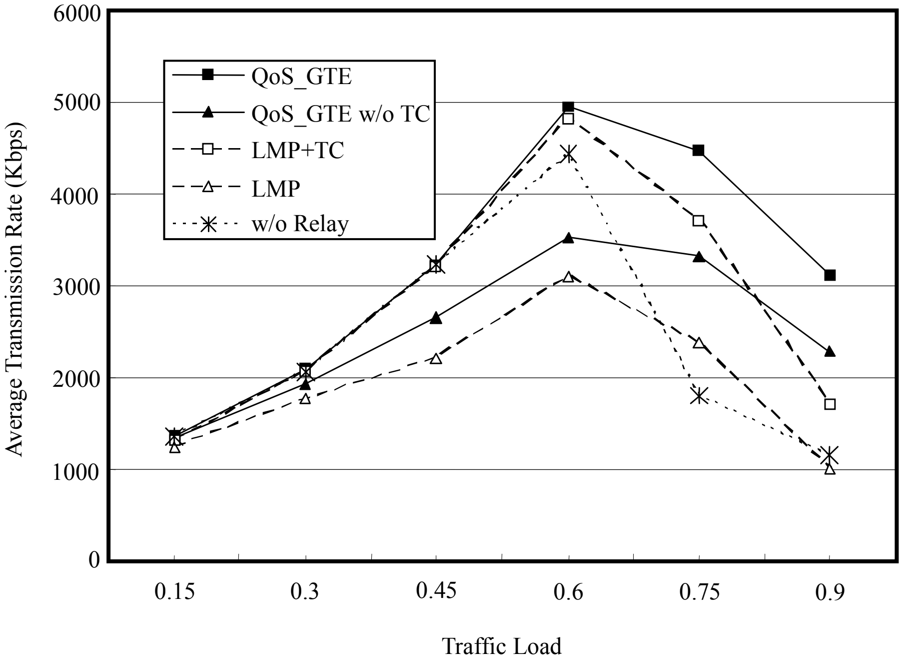
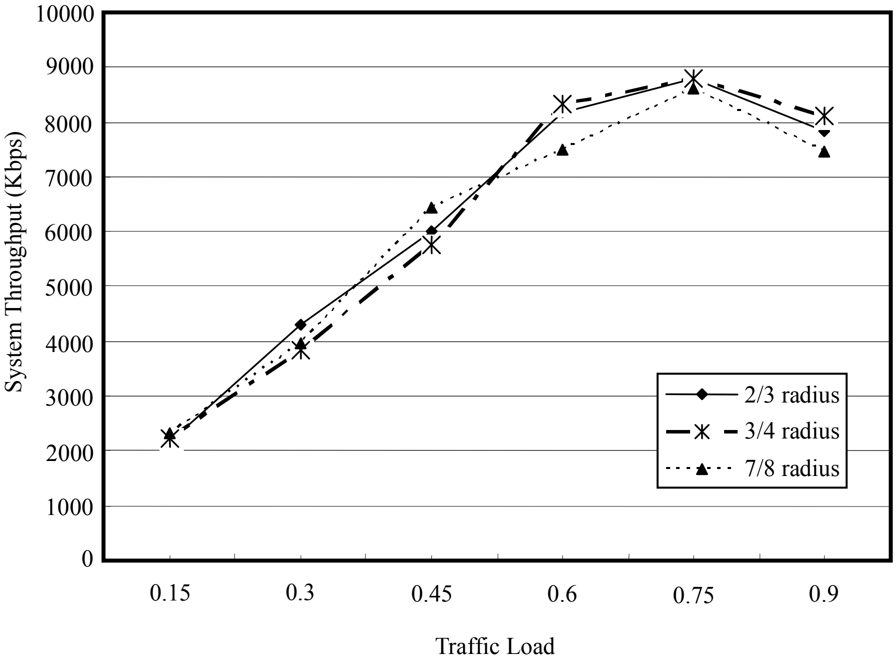


Figure 4.6. Average Transmission Rate of FTP

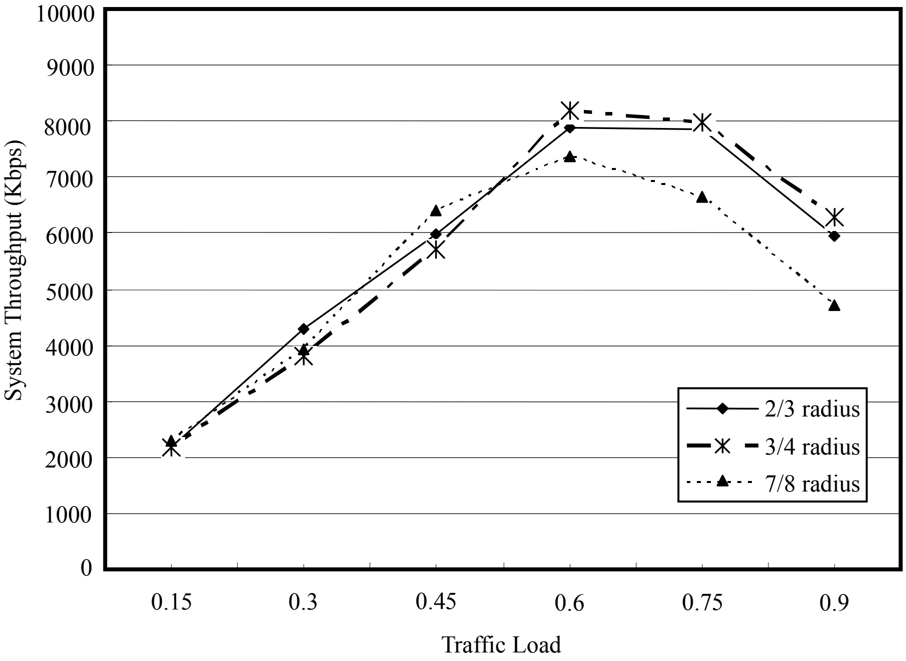
4.3.2 Station Performance Evaluation on Location of Relay

To evaluate the impact of the RS location, we compare the performances for three RS locations: one is at two thirds of the cell radius, another is at three fourths of the cell radius, and the other is at the seven eighths of the cell radius. Also, if the desired site is not located also an intersection, the RS is placed on the nearest intersection from the desired site location. Both QoS_GTE and LMP+TC are used in the simulation. SO_based resource allocation algorithm is adopted as the scheduling for LMP+TC.

Figure depicts the system throughput for three distinct RSs locations in QoS_GTE and in LMP+TC, respectively. At the light traffic load, the performance order of three cases has no regularity. When the traffic load is at 0.3, the case of 2/3 is the best. When the traffic load is 0.45, the case of 7/8 is the best. This is because there is enough resource for low traffic load; therefore, the location of RS does not affect the system performance. On the contrary, as the traffic load increases, the throughput performance from best to worst is the case of 3/4, the case of 2/3 and the case of 7/8, which is clearer in LMP+TC scheme. For the case of 2/3 radius, RSs are so close to each other that each RS coverage is overlapped and this results in reducing the service area of RSs in the cell. Besides, RSs are too close to BS to replace the transmission from BS to MSs and the probability of using relay is low. On the contrary, when RSs is at the 7/8 radius, there may be a gap between any two RSs where no RS can serves this area. Further, half of the RS service area is almost out of the cell when RSs are on 7/8 radius which also make the RS service area decrease. As the RS coverage decreasing, the probability of replacing direct path is attenuating. As a result, RSs are idle and the relay-assisted system acts like without RS system. As to the case of RSs at 3/4 radius, it can get the largest RS service area and relay path replaces the direct transmission as efficiently as it can hence its QoS_guaranteed and system throughput enhancement are both done in the best way.



(a) System Throughput of QoS_GTE



(b) System Throughput of LMP+TC

Figure 4.7. System Throughput of Tow Relay Schemes

Figure is the voice packet dropping rate in QoS_GTE and also in LMP+TC. The voice packet dropping rate is all zero for each case of RS location. Nevertheless, video packet dropping rates for different RS locations in QoS_GTE and LMP+TC are obviously distinguished in Figure. The case of 3/4 radiuses has the lowest video packet dropping rate and the case of 7/8 radius is the worst. Also, Figure reveals the non-guaranteed ratio of HTTP users in QoS_GTE and LMP+TC respectively. The case of 7/8 radiuses has the largest HTTP non-guaranteed rate of all undoubtedly.

From Figure to Figure, we can summarize that the system performance is affected severely by the location of RSs. A proper RS location makes the RS service coverage maxima and makes RSs supply BS as long as the direct path is not good enough. The location too near or too far from BS leads the advantage of setting RSs in vain.

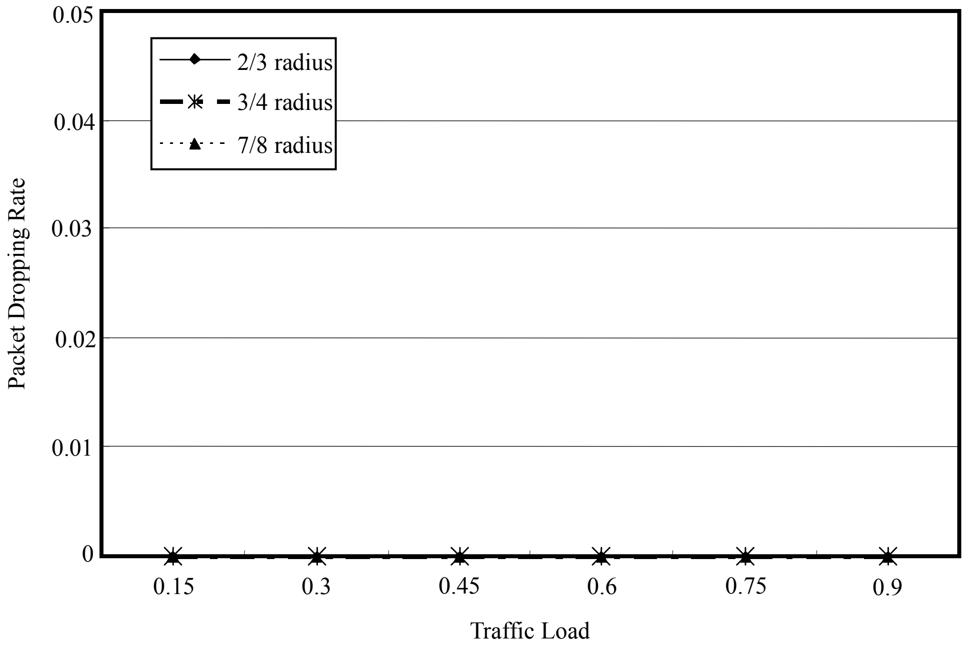
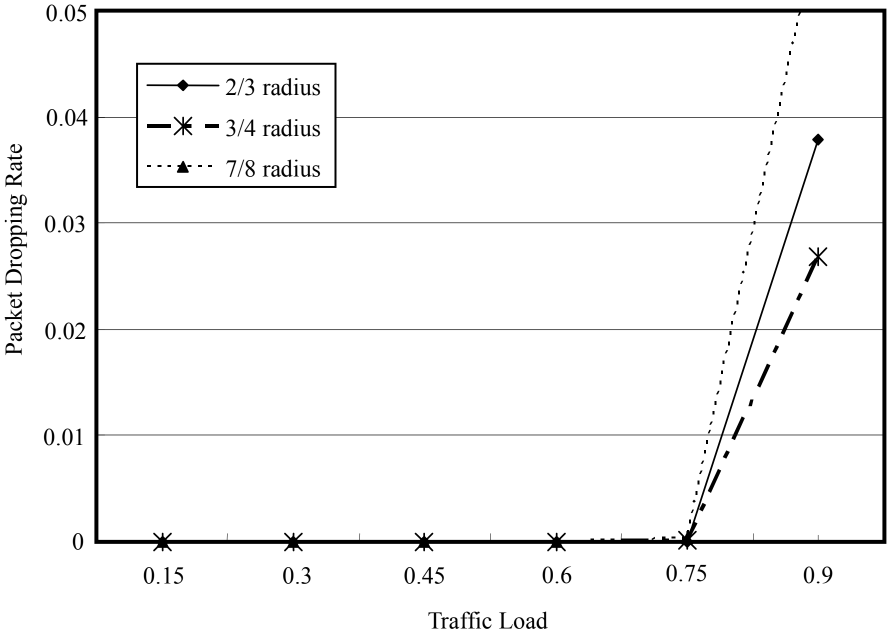
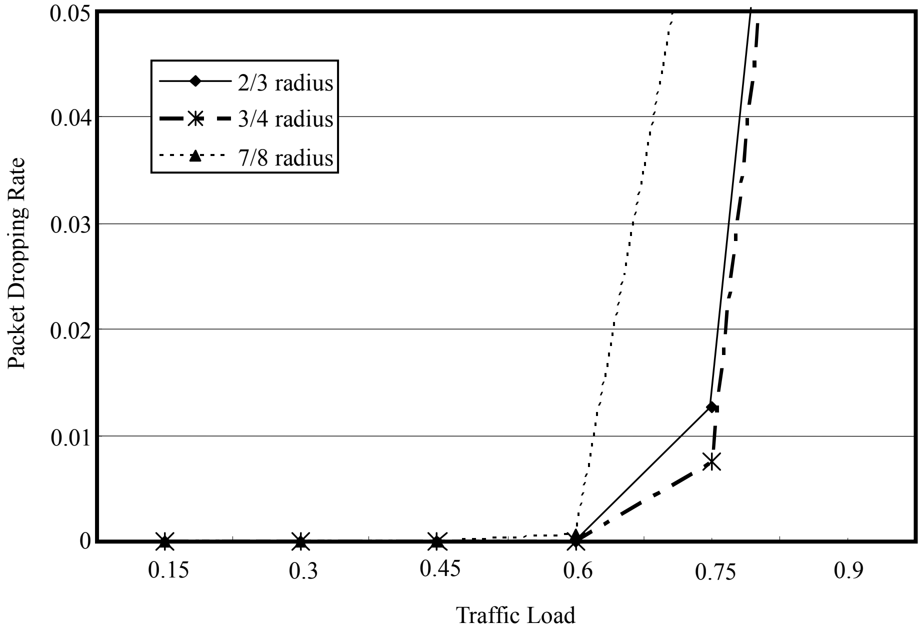


Figure 4.8. Voice Packet Dropping Rate in QoS_GTE and LMP+TC

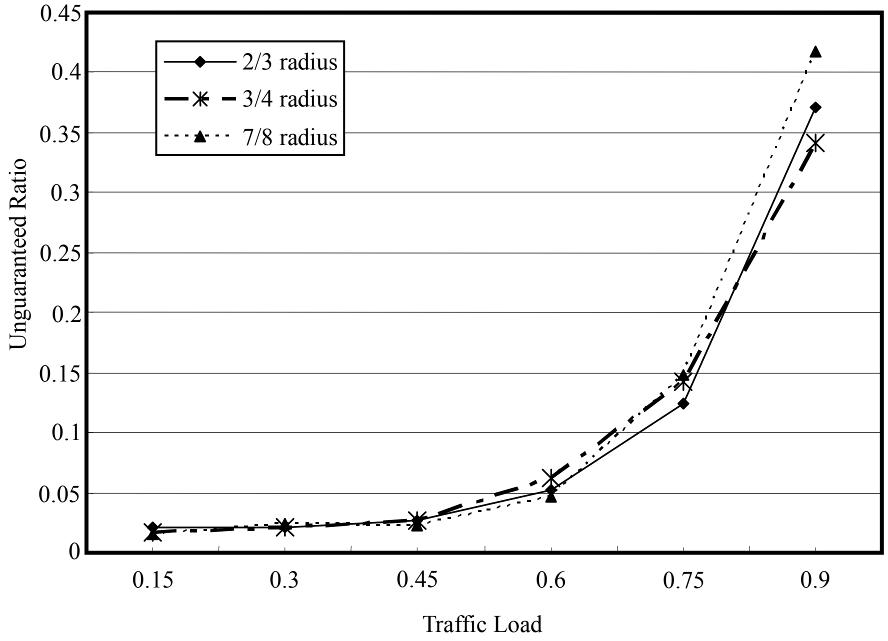


(a) Video Packet Dropping Rate of QoS_GTE

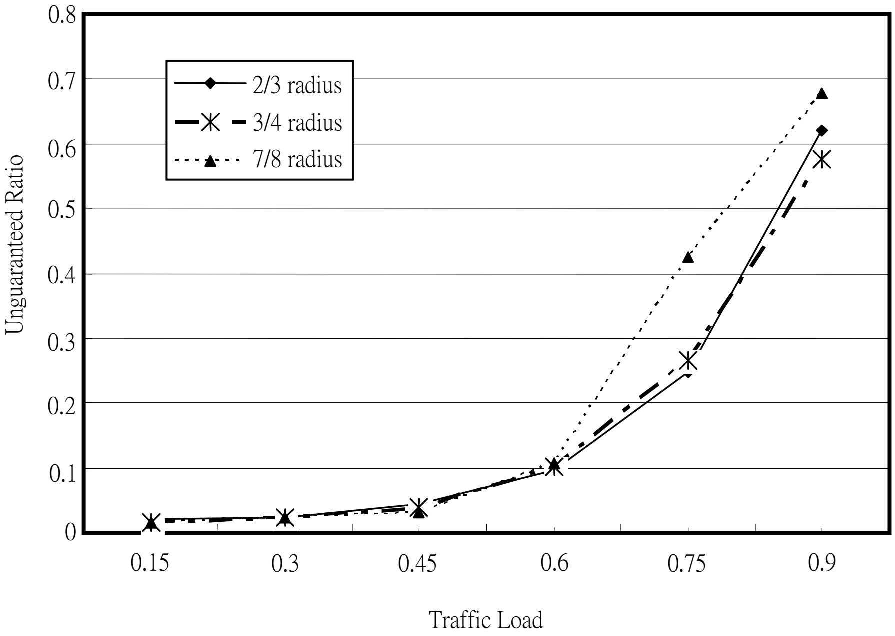


(b) Video Packet Dropping Rate of LMP+TC

Figure 4.9. Video Packet Dropping Rate



(a) HTTP traffic Non-guaranteed Ratio of QoS_GTE



(b) HTTP traffic Non-guaranteed Ratio of LMP+TC

Figure 4.10. Non-guaranteed Ratio of HTTP traffic

Chapter 5
Conclusions

In the thesis, a downlink centralized QoS guaranteed and throughput enhancement (QoS_GTE) scheduling scheme for WiMAX relay-assisted network is proposed, where two hop relay system is considered. The QoS_GTE scheduling scheme consists of a transmission time based (TT_based) path selection algorithm, a service order based (SO_based) resource allocation algorithm and a transmission concurrency (TC) decision algorithm.

The TT_based path selection algorithm takes the overall path of path-loss, shadow fading, and interference into consideration in terms of transmission time. It chooses the path with the minimal total transmission time as the transmission path. It not only considers the link quality but also the transmission efficiency, which react directly upon the transmission time. The TT_based algorithm helps to find a path by which the link quality is better than that of the algorithm without relay. Also, the system efficiency is improved compared with other path selection algorithms in relay network. This is because the conventional scheme (LMP) selects the path only based on the bottleneck path loss, which can improve the link quality but the system throughput.

The SO_based resource allocation algorithm gives high priority to the urgent users according to service order S_m at the current frame and dynamically adjusts the S_m of users frame by frame. The goals of the SO_based resource allocation algorithm are for QoS satisfaction and throughput maximization. In addition, multiple service classes, which include real time, non real time, and best effort services, are considered. Since using relay may cause resource consuming, The TC decision algorithm carries out resource reuse by deciding which RSs can transmit concurrently using the same frequency and time slots.

Simulation results show that the QoS_GTE is compared with LMP+TC and without relay case. From the results, we can conclude that QoS_GTE outperforms LMP+TC and without relay case in terms of system throughput and the satisfaction extent of QoS requirements. The TT_based path selection algorithm is better than LMP scheme both in link quality and system performance. The SO_based resource allocation algorithm can make sure that voice packets will be satisfied with the requirement all the time. The video packet cans also be satisfied until the traffic load is high. The TC decision algorithm indeed raises the throughput pretty much. Besides, the location of RS plays an important effect on the system performance. It is concluded that the proper location of RS in our scenario is at three fourth cell radius.

Bibliography

[1] B. H. Walke, R. Pabst, D. Schultz, P. Herhold, H. Yanikomeroglu, S. Mukherjee, H. Viswanathan, M. Lott, W. Zirwas, M. Dohler, H. Aghvami, D. Falconer, and G. P. Fettweis, "Relay-based Deployment Concept for Wireless and Mobile Broadband Radio," *IEEE* Commun. *Mag.*, vol. 42, no.9, pp. 80-89, 2004.

[2] H. Wu, Ch. Qiao, S. De and O. Tonguz, "Integrated Cellular and Ad Hoc Relaying System: iCAR," *IEEE Journal on Selected Area in Communications*, vol. 19, no. 10, pp. 2105-2115, October 2001.

[3] J. Cho and Z. J. Hass, "On the Throughput Enhancement of the downstream Chnnel in Cellular Radio Networks Through Multihop Relaying," *IEEE Journal on Selected Areas in Communication*. Vol. 22, no. 7, pp 1206-1219, September. 2004.

[4] T. Hui, G. Xuelin, and Z. Ping, "The Impact of Relaying Strategies on the Performance in Cellular System," *Proc. ISCIT*, pp. 1357-1360, 2005.

[5] V. Sreng, H. Yanikomeroglu, and D. D. Falconer, "Relayer Selection Strategies in Cellular Networks with Peer-to-Peer Relaying," *Proc. IEEE VTC* Fall pp. 1949-1953, 2003.J. Bae, R. Berry and M. L. Honig, "Power Allocation, Rate, and Coverage for Relay-Assisted Downlink Data Transmission," *Proc. IEEE ICC* June, 2006.

[6] J. Bae, R. Berry and M. L. Honig, "Power Allocation, Rate, and Coverage for Relay-Assisted Downlink Data Transmission," *Proc. IEEE ICC* June, 2006.

[7] H. Viswanathan and S. Mukherjee, "Performance of Cellular Network with Relays and Centralized Scheduling," *IEEE Transactions on Wireless Communications.* vol. 4, no. 5, pp 2318-2328, September 2005.

[8] H. Yu, S. Ganguly, R. Izmailov and Z. J. Hass, "Interference-Aware IEEE802.16

WIMAX Mesh Networks," *Proc. VTC 2005 Spring*, pp 3102-3106, 2005.

[9] L. Fu, Z. Cao, and P. Fan, "Spatial Reuse in IEEE 802.16 Based Wireless Mesh Networks," *Proc. ISCIT*, pp 1311-1314, 2005.

[10] N. Challa, H. Cam, "Cost-Aware Downlink Scheduling of Shared Channel for Cellular Network with Relays," *Performance, Computing, and Communication, 2004 IEEE International Conference*, pp 793-798, 2004.

[11] J. Jang and K. B. Lee, "Transmit Power Adaptive for Multiuser OFDM Systems," *IEEE J. Select. Areas Commu.*, vol. 21, pp 171-178, February 2003.

[12] W. Rhee and J. M. Cioffi, "Increase in Capacity of Multiuser OFDM System Using Dynamic Subchannel Allocation," in *Proc. IEEE VTC, Tokyo, Japan*, May 2000, pp 1085-1089.

[13] X. Zhang, E. Zhou, R. Zhu, S. Liu, and W. Wang, "Adaptive Multiuser Radio Resource Allocation for OFDMA System," *IEEE Globecom 2005*, pp 3846-3850.

[14] Universal Mobile Telecommunication System, Selection procedures for the choice of radio transmission technology of the UMTS, UMTS Std. 30.03, 1998.

[15] 3GPP TR 25.892, "Feasibility study for OFDM for UTRAN enhancement," 3rd Generation Partnership Project, Tech. Rep., 2004-06.

[16] IST-2003-507581 winner d5.4 v.1. on line available – http://www.ist-winner.org/.

[17] A. J. Goldsmith and S. G. Chua, "Variable-Rate Variable-Power MQAM for Fading Channels," *IEEE Transaction on Communication*, vol. 45, no. 10, Oct. 1997, pp1218-1230.

[18] H. Yaghoobi, "Scalable OFDMA physical layer in IEEE 802.16 WirelessMAN," *Intel Technology Journal*, vol. 8, issue 3, 2004.

Vita

Chia-Shueng Chuang was born on 1983 in Taitong, Taiwan. She received the B.E. degree in electrical engineering from National Cheng-Kung University, Tainan, Taiwan, in 2005, and the M.E. degree in the department of communication engineering, college of electrical and computer engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2007. Her research interests include radio resource management and wireless communication.