行政院國家科學委員會專題研究計畫　成果報告

子計劃一：異質多接取網路之資源管理技術(I)(電信科技合作案)

計畫類別：整合型計畫
計畫編號：NSC91-2219-E-009-013-
執行期間：91年08月01日至92年07月31日
執行單位：國立交通大學電信工程學系

計畫主持人：張仲儒

報告類型：完整報告

處理方式：本計畫可公開查詢

中　華　民　國 92年10月30日

行政院國家科學委員會補助專題研究計畫 ■ 成 果 報 告
□期中進度報告

（計畫名稱）

B3G 無線接取網路之無線資源管理技術(1/3)

Radio Resource Management Technologies for B3G Wireless Access Network (1/3)

計畫類別：□ 個別型計畫 ■整合型計畫
計畫編號：NSC 91－2219－E－009－012
執行期間： 91 年 08 月 01 日至 92 年 07 月 31 日

計畫主持人：張仲儒教授
共同主持人：沈文和教授、廖維國教授、王蒞君教授
計畫參與人員：
成果報告類型(依經費核定清單規定繳交)：□精簡報告 ■完整報告

本成果報告包括以下應繳交之附件：
□赴國外出差或研習心得報告一份
□赴大陸地區出差或研習心得報告一份
□出席國際學術會議心得報告及發表之論文各一份
□國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢
□涉及專利或其他智慧財產權，□一年□二年後可公開查詢

執行單位：國立交通大學電信工程學系

中 華 民 國 92 年 10 月 28 日

目錄

中文摘要

B3G異質性多接取網路是一整合性系統，結合了個人區域網路、無線區域網路、蜂巢式行動通訊網路的優點，並且可依照使用者的地點、移動速率、傳輸速率與要求品質的不同，提供無所不在的服務。然而頻率資源極其珍貴，必須有效管理系統資源，包括：頻譜、傳輸功率、 無線接取控制等，讓個人化寬頻無線多媒體服務，也能符合經濟效益。 因此，本研究計畫探討WCDMA/WLAN與巨細胞/微細胞異質多接取網路之無線資源管理技術。

針對WCDMA/WLAN異質多接取網路，我們首先分析異質多接取網路的系統效能。我們發展一無線資源參數子（RRI）來預估WCDMA系統中新連線所需之資源量（Part 1），該無線資源參數子將連線之訊務參數和服務品質需求轉換為統一的資源量度。此外，我們也發展一Q學習演算法式多速率傳輸控制器（Q-MRTC），用於控制WCDMA網路資源（Part 2）。我們將多速率傳輸控制問題化約為一半馬可夫決策鍊問題，並成功的利用一及時性強制型學習機制，Q學習演算法，來準確預估多速率傳輸控制的成本函數。至於WLAN網路，我們探討差異式服務之加權式傳輸公平問題（Part 3）。我們提出一馬可夫鍊佇列模型來分析此一問題，並獲得一代數解。該分析模型可精確敘述接取機率與競爭時槽的關係。模擬結果驗證了此一分析模型的正確性。

對於巨細胞/微細胞異質多接取網路，我們提出一結合功率與速率配置（JPRA）之軟性換手演算法（Part 4）。和現有單點傳輸機制相比，該演算法可具有功率負載均衡的優點。模擬結果顯示JPRA演算法可達成較低的強制斷線機率，以及較高之系統流通量。

Keywords: 異質性多接取網路、B3G、WCDMA、WLAN、無線資源管理、軟性換手機制

Abstract

Future B3G heterogeneous access network is an integrated network, which consists of PAN, WLAN, and cellular network. The B3G network can provide users always-on and ubiquitous services regardless of their geographical location, moving speed, service rate, and quality of service. The wireless spectrum is scarce so the radio resource management schemes (including: spectrum efficiency, throughput, call admission control and etc.) have to be carefully designed. Hence, the project focuses on developing radio resource management technologies on WCDMA/WLAN and macro-cell/micro-ell heterogeneous access networks.

For the WCDMA/WLAN heterogeneous access network, the system performance of heterogeneous access network is firstly analyzed. We firstly develop a radio resource index (RRI) to estimate the radio resource required for a call connection in WCDMA cellular systems (Part 1). The RRI can transform traffic parameters and quality-of-service (QoS) requirements of the call connection into a measure of resource in a unified metric, while keeping QoS requirements of existing calls guaranteed. Also, a Q-learning-based multi-rate transmission control scheme (Q-MRTC) for radio resource control (RRC) in WCDMA systems is proposed (Part 2). Here, the multi-rate transmission control problem is modeled as a semi-Markov decision process (SMDP). And we successfully apply a real-time reinforcement learning algorithm, named Q-learning, to accurately estimate the transmission cost for the multi-rate transmission control. As to the WLAN, the weighted fairness is investigated for providing differentiated services (Part 3). A Markov chain queuing model is proposed and the closed form solution is derived. The proposed model exactly describes the relationship between access probability and contention window. The simulation results justify the validity of the analytic model.

For the macro-cell/micro-cell heterogeneous access network, a joint power and rate assignment (JPRA) algorithm for multirate soft handoff is proposed (Part 4). It can achieve power balancing between cells for soft handoff better than the conventional site-selection diversity transmission (SSDT) scheme. Simulation results show that the JPRA algorithm can achieve lower forced termination probability and higher system throughput.

Keywords: heterogeneous access network, Beyond 3rd access network (B3G), WCDMA, WLAN, radio resource management, call admission control, soft handoff.

The Radio Resource Index for Real-Time Service in WCDMA Cellular Systems

Scott Shen and Chung-Ju Chang
Department of Communication Engineering
National Chiao Tung University
Hsinchu 300, Taiwan ROC
E-mail: cjchang@cc.nctu.edu.tw
Tel. No.: 886-3-5731923 Fax No.: 886-3-5710116

Abstract

In this paper, a *radio resource index* (RRI) is derived to estimate the radio resource required for a call connection in WCDMA cellular systems. An analytical model is proposed and large deviation techniques are adopted to obtain the RRI. The RRI can transform traffic parameters and quality-of-service (QoS) requirements of the call connection into a measure of resource in a unified metric, while keeping QoS requirements of existing calls guaranteed.

Keywords

radio resource index, large deviation, WCDMA.

Part of this work was published in IEEE VTC 2002-Fall, Vol.1, pp.: 299-303. The journal version is submitted to *IEEE Communication Letter*.

I. Introduction

The amount of radio resource required by a call connection in WCDMA is generally determined by its traffic parameters and QoS requirements. If a transformation that maps these parameters and requirements into a *radio resource index* (RRI) exists, it will be useful to radio resource management in WCDMA cellular systems [1].

The concept of RRI for WCDMA cellular systems is similar to that of effective bandwidth for ATM networks. However, in the derivation of RRI, the interference from both home cell and adjacent cells, the required power for each connection, and the random access behaviour of MAC layer on the interference are to be considered. Moreover, the radio resource is constrained by a packet dropping ratio in order to guarantee the call-level requirement, while the bandwidth of an ATM node is a pre-defined hard capacity.

The algorithms to allocate radio resource has been studied in many literature [2], [3]. These algorithms just allocated power to each user for fulfilling the required BER. However, from the QoS architecture in WCDMA [4], there are still some requirements other than the BER requirement, such as the packet error ratio and the delay, etc. If the allocated power considers only the BER, some requirements may not be satisfied to the time varying interference level. The characteristics of the interference process are influenced by the number and the characteristics of the active connections. In order to fulfill the BER and other requirements, a longer time scale interference process should be considered.

In this paper, we study the radio resource allocation of real-time connection in WCDMA cellular systems. We here define BER as the packet-level requirement and the packet dropping ratio and the tolerable delay as the call-level requirements. The radio resource allocation algorithm considers both the packet-level requirement and the call-level requirements. The RRI is derived for uplink call connection as an indication for base station to estimate the amount of radio resource used by the user at receiver side. We first define analytical elements in a model to equivalently describe the behaviour of the system. Some properties are developed to derive the input and output relationship of the analytical elements. We then form the failure process consisted of all the dropped packets due to either excess delay in the transmitter or the channel error. In order to fulfill the BER and dropping ratio requirement, we can find that the required radio resource, converted into the unit of power, is in terms of the SIR, source traffic characteristics and call-level requirements. Based on the concept of multiplexing different users on the shared channel, we finally calculate the equivalent radio resource index consumed by each user at the base station side. The RRI performs as a function mapping from a parametric space, which is constituted by traffic parameters and QoS requirements, to the metric space, which is of unified metric to each different connections, while keep QoS requirements of all existing calls guaranteed. We describe the overall operations and the model of the system in section II. The *Radio Resource Indicator* is derived In section III. Results and conclusions are discussed in section IV.

II. System Model

We assume the connection with bursts, of which packets of a burst arrive in a batch fashion. The head-of-line packet of a burst is firstly sent as a request for resource reservation, where the request permission probability $\tilde{r}_p$ is determined and broadcasted by the radio network controller. If the first packet is not permitted to transmit in this frame or its transmission is corrupted in the air interface, it will retry. If the first packet fails to transmit successfully or be acknowledged before the maximum tolerable dalay, the whole burst will be dropped. Once the first packet is successfully acknowledged, the remaining packets of burst can be sequentially sent without any further request. The packet sent out but corrupted in the air interface would be discarded by the receiver. We set a SIR threshold, SIR_0, to determine if the packet is successfully received or not. We can set the maximum system load I_{th}, and the required received power to transmit information in a basic rate with required SIR_0 is P_i^0. And we assume that the call-level QoS requirements of user i are the packet dropping ratio, $R_{D,i}^*$, and the maximum tolerable delay, M_D^*.

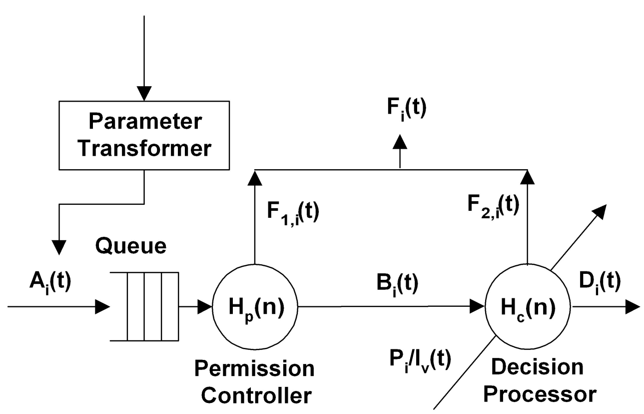


Fig. 1. The analytical model of the uplink connection i in WCDMA.

There are N_v connections in the uplink of WCDMA cellular systems. The analytical model for uplink connection i is shown in Fig. 1, where the capital letter X is to represent the cumulative process of a process.The $A_i(t)$ denotes the *arrival process* of user i. The $H_p(n)$ is the *access process* for arrivals in permission controller, where packets are transmission permitted or dropped. The dropped packets are directed toward a *failure process* denoted by $F_{1,i}(t)$; the permitted packets form an *output process* denoted by $B_i(t)$. Subsequently, the process $B_i(t)$ enters into the *decision process* denoted by $H_c(n)$. The decision processor determines the way the packet goes in the light of the level of *aggregated interference process* denoted by $I_v(t)$, which is the summation of the adjacent-cell interference $I_a(t)$ and the home-cell interference $I_h(t)$. The transmission power P_i of the transmission packet of connection i forms a process $I_i(t)$ on the air interface, which indicates the power component contributed by connection i onto the transmission channel. If $P_i/I_v(t)$ is less than SIR_0, the packet will be directed to a *failure process* $F_{2,i}(t)$, otherwise, the packet will be successfully received and forms a *departure process*, denoted by $D_i(t)$. Note that for convenience, $I_h(t) = \sum_{j=1}^{N_v} I_j(t)$ including that of user i is assumed, which would be the upper-bounded interference for each user.

III. Analysis

The arrival process $A_i(t)$ can be decomposed into a point process $A_i^\tau(t)$ and a burst length process $A_i^l(n)$. The compound process has the property described in lemma 1.

Lemma 1:

The arrival process $A_i(t)$, consisting of a point process $A_i^\tau(t)$ and a length process of the n-th burst $A_i^l(n)$, has the *Gärtner-Ellis Limit*, denoted by $\Lambda_{A_i}(\theta)$, which can be expressed as

$$\Lambda_{A_i}(\theta) = \Lambda_{A_i^\tau}(\Lambda_{A_i^l}(\theta)). \tag{1}$$

□

Lemma 2:

Consider the point process $A_i^\tau(t)$ through the permission controller with access process$H_p(n)$. Let $\Lambda_{H_p}(\theta)$ be the *Gärtner-Ellis Limits* of the $H_p(n)$. The point process output from the permission controller for $A_i^\tau(t)$, denoted by $B_i^\tau(t)$, has the *Gärtner-Ellis Limits* $\Lambda_{B_i^\tau}(\theta)$ obtained by

$$\Lambda_{B_i^\tau}(\theta) = \Lambda_{A_i^\tau}(\Lambda_{H_p}(\theta)). \tag{2}$$

□

The burst length process $B_i^l(n)$ of the output process $B_i(t)$ is preserved the same as $A_i^l(n)$. The *Gärtner-Ellis Limits* of $B_i(t)$, denoted by $\Lambda_{B_i}(\theta)$, can be directly written as

$$\begin{aligned}\Lambda_{B_i}(\theta) &= \Lambda_{B_i^\tau}(\Lambda_{B_i^l}(\theta)) \\ &= \Lambda_{B_i^\tau}(\Lambda_{A_i^l}(\theta)). \end{aligned} \tag{3}$$

Lemma 3:

The arrival process $A_i(t)$ is divided into two sub-flows: $B_i(t)$ and $F_{1,i}(t)$. And the *Gärtner-Ellis Limit* of $F_{1,i}(t)$ can be expressed as

$$\Lambda_{F_{1,i}}(\theta) = \Lambda_{A_i}(\theta) - \Lambda_{B_i}(\theta). \tag{4}$$

□

Lemma 4:

The outage probability of connection i, denoted by R_{otg}, related to the aggregated interference can be obtained by

$$R_{otg} = \lim_{N_v \to \infty} \mathbf{Pr}[I_v(t) \in G_i] \approx e^{-\Lambda^*_{I_v}(G_i)}, \tag{5}$$

where $G_i = \{I_v | P_i/I_v < SIR_0\}$, $\Lambda^*_{I_v}(G_i)$ is the rate function of *interference process*, and N_v is the total number of calls in WCDMA.

□

Lemma 5:

The *Gärtner-Ellis Limits* of the decision process $H_c(n)$, denoted by $\Lambda_{H_c}(\theta)$, can be derived as

$$\Lambda_{H_c}(\theta) = log(e^{\theta - \Lambda^*_{I_v}(G_i)} + 1 - e^{-\Lambda^*_{I_v}(G_i)}). \tag{6}$$

□

The failure process $\Lambda_{F_{2,i}}(\theta)$ from the decision processor can be gotten by

$$\Lambda_{F_{2,i}}(\theta) = \Lambda_{B_i}(\Lambda_{H_c}(\theta)). \tag{7}$$

Similar to Eq. (4), $\Lambda_{D_i}(\theta)$ can be yielded as

$$\Lambda_{D_i}(\theta) = \Lambda_{B_i}(\theta) - \Lambda_{F_{2,i}}(\theta). \tag{8}$$

Based on these results derived before, we can predict the packet dropping ratio for each connection. For a given packet dropping ratio requirement $R^*_{D,i}$, we can obtain a channel condition constraint by the following lemma.

Lemma 6:

For the QoS requirement $R^*_{D,i}$, the aggregated interference process $I_v(t)$ on connection i has the constraint given by

$$\Lambda^*_{I_v}(G_i) \geq -log(1 - \frac{1 - R^*_{D,i}}{\tilde{r}_s}), \tag{9}$$

where $\tilde{r}_s$ is the successful probability of a burst from the permission controller.

$< pf >$

The *Gärtner-Ellis Limits* for a general process $Z(t)$, denoted by $\Lambda_Z(\theta)$, has two properties. As $\theta \to 0$, it equals to zero and its derivative equals to the mean of the process μ_z.

The packet dropping ratio of connection i can be derived as

$$\begin{aligned} \hat{R}_D(i) &= \lim_{t\to\infty} \frac{1}{\mu_{A_i}} \Lambda'_{F_{1,i}(t)}(\theta)\Big|_{\theta=0} \\ &+ \lim_{t\to\infty} \frac{1}{\mu_{A_i}} \Lambda'_{F_{2,i}(t)}(\theta)\Big|_{\theta=0}. \end{aligned} \tag{10}$$

For $\Lambda'_{F_{1,i}(t)}(\theta)$ in the first term of (10), it can be obtained by

$$\begin{aligned} & \Lambda'_{F_{1,i}(t)}(\theta)\Big|_{\theta=0} \\ = & \frac{d\Lambda_{A_i}(\theta)}{d\theta}\Big|_{\theta=0} - \frac{d\Lambda_{A_{\tau,i}}(\Lambda_{H_p}(\Lambda_{A_i^l}(\theta)))}{d\theta}\Big|_{\theta=0} \\ = & \frac{d\Lambda_{A_i}(\theta)}{d\theta}\Big|_{\theta=0} - \frac{d\Lambda_{A_{\tau,i}}(\theta_1)}{d\theta_1}\Big|_{\theta_1=\Lambda_{H_p}(\Lambda_{A_i^l}(\theta))} \cdot \\ & \frac{d\Lambda_{H_p}(\theta_2)}{d\theta_2}\Big|_{\theta_2=\Lambda_{A_i^l}(\theta)} \cdot \frac{d\Lambda_{A_i^l}(\theta)}{d\theta}\Big|_{\theta=0} \\ = & \frac{d\Lambda_{A_i}(\theta)}{d\theta}\Big|_{\theta=0} - \frac{d\Lambda_{A_{\tau,i}}(\theta_1)}{d\theta_1}\Big|_{\theta_1=0} \cdot \frac{d\Lambda_{H_p}(\theta_2)}{d\theta_2}\Big|_{\theta_2=0} \cdot \\ & \frac{d\Lambda_{A_i^l}(\theta)}{d\theta}\Big|_{\theta=0} \\ = & \mu_{A_i} - \mu_{A_{\tau,i}} \cdot \tilde{r}_s \cdot \mu_{A_i^l}, \end{aligned} \tag{11}$$

and $\tilde{r}_s$ is given in Eq. (**??**). For $\Lambda'_{F_{2,i}(t)}(\theta)$ in the second term of (10), it can be derived as

$$\Lambda'_{F_{2,i}(t)}(\theta)\Big|_{\theta=0}$$

$$
\begin{aligned}
&= \left.\frac{d\Lambda_{B_i}(\Lambda_{H_c}(\theta))}{d\theta}\right|_{\theta=0} \\
&= \left.\frac{d\Lambda_{B_i}(\theta_1)}{d\theta_1}\right|_{\theta_1=0} \cdot \left.\frac{d\Lambda_{H_c}(\theta)}{d\theta}\right|_{\theta=0} \\
&= \left.\frac{d\Lambda_{B_i}(\theta_1)}{d\theta_1}\right|_{\theta_1=0} \cdot \lim_{n\to\infty}\frac{1}{n}\cdot\left.\frac{E\left[H_c(t)e^{\theta H_c(n)}\right]}{E\left[e^{\theta H_c(n)}\right]}\right|_{\theta=0} \\
&= \left.\frac{d\Lambda_{B_i}(\theta_1)}{d\theta_1}\right|_{\theta_1=0} \cdot \lim_{n\to\infty} E\left[\frac{H_c(n)}{n}\right] \\
&= \left.\frac{d\Lambda_{B_i}(\theta_1)}{d\theta_1}\right|_{\theta_1=0} \cdot \lim_{n\to\infty} E\left[\frac{\sum_{k=1}^{n} 1_{G_i}(I_v(\tau(n)))}{n}\right] \\
&= \left.\frac{d\Lambda_{B_i}(\theta_1)}{d\theta_1}\right|_{\theta_1=0} \cdot E\left[I_v(\tau(n))\right] \\
&= \mu_{B_i}\cdot exp\{-\Lambda^*_{I_v}(G_i)\}. \qquad (12)
\end{aligned}
$$

Therefore, $\hat{R}_D(i)$ can be rewritten as

$$
\begin{aligned}
\hat{R}_D(i) &= \Big(\mu_{A_i} - \mu_{A_{\tau,i}}\cdot\tilde{r}_s\cdot\mu_{A_i^l} + \\
&\quad \mu_{B_i}\cdot exp\{-\Lambda^*_{I_v}(G_i)\}\Big)/\mu_{A_i}. \qquad (13)
\end{aligned}
$$

Consequently, given the QoS constraint $\hat{R}_D(i) \le R^*_{D,i}$, $\Lambda^*_{I_v}(G_i)$ can be reformatted and obtained by

$$
\Lambda^*_{I_v}(G_i) \ge -log(1 - \frac{1-R^*_{D,i}}{\tilde{r}_s}). \qquad (14)
$$

This result shows us that the function $\Lambda^*_{I_v}(G_i)$ measuring the set G_i should satisfy condition such that the outage probability faced by this user can fulfill the packet dropping ratio. As $1 - R^*_{D,i}$ approaches $\tilde{r}_s$, that is the packet dropping due to channel loss, the required power $\Lambda^*_{I_v}(G_i)$ should be increased.

□

Lemma 7: The Required Power for Fulfilling Dropping Ratio

For a given *Gärtner-Ellis Limits* $\Lambda_{I_v}(\theta)$ with constraint $\Lambda^*_{I_v}(G_i) \ge -log(1-\frac{1-R^*_{D,i}}{\tilde{r}_s})$, and given that θ^* is a pre-set value. The required power for fulfilling the packet dropping ratio can be set as

$$
P_i \ge \frac{-log(1-\frac{1-R^*_{D,i}}{\tilde{r}_s}) + \Lambda_{I_v}(\theta^*)}{\theta^*}\cdot SIR_0
$$

$< pf >$

We can see that

$$\begin{aligned}\Lambda^*_{I_v}(G_i) &= \inf_{\alpha \in G_i} \Big(\Lambda^*_{I_v}(\alpha)\Big) \\ &= \sup_{\theta} \left\{ \frac{P_i}{SIR_0} \cdot \theta - \Lambda_{I_v}(\theta) \right\} \\ &\geq -log(1 - \frac{1 - R^*_{D,i}}{\tilde{r}_s}). \end{aligned} \tag{15}$$

For any $\theta^* > 0$,

$$\begin{aligned} & \frac{P_i}{SIR_0} \cdot \theta^* - \Lambda_{I_v}(\theta^*) \geq -log(1 - \frac{1 - R^*_{D,i}}{\tilde{r}_s}) \\ \Rightarrow \quad & P_i \geq \frac{-log(1 - \frac{1 - R^*_{D,i}}{\tilde{r}_s}) + \Lambda_{I_v}(\theta^*)}{\theta^*} \cdot SIR_0. \end{aligned} \tag{16}$$

If we select the power P_i for this user, the constraint in *Lemma 6* will be satisfied.

□

The result shows us that the required power increment consists of two elements: the call level requirement related quantity $\frac{-log(1-\frac{1-R^*_{D,i}}{\tilde{r}_s})}{\theta^*}$ and the equivalent system load $\frac{\Lambda_{I_v}(\theta^*)}{\theta^*}$. As the packet dropping ratio becomes stricter, the more power increment is needed. And if the expected interference is increased, the required power is also increased.

Lemma 8: The RRI of Connection i

The RRI of connection i can be obtained by

$$RRI_i = \frac{\Lambda_{A_i}(\Lambda_{H_p}(P_i \cdot \theta^*))}{\theta^*}. \tag{17}$$

$< pf >$

We define the total radio resource index as the maximum load received at the base station, i.e. the total RRI is set to be $\frac{\Lambda_{I_v}(\theta^*)}{\theta^*} = I_{th}$. From Lemma 7, the required power for user i to transmit in basic rate is select as

$$P_i \geq (I_{th} - \frac{-log(1 - \frac{1-R^*_{D,i}}{\tilde{r}_s})}{\theta^*}) \cdot SIR_0 \tag{18}$$

According to the selection of required power for connection i, the equivalent system load $\frac{\Lambda_{I_v}(\theta^*)}{\theta^*}$ should operate below the value I_{th}, otherwise the QoS of connection i will be violated. From the multiplexing property of $\frac{\Lambda_{I_v}(\theta^*)}{\theta^*}$, the $\frac{\Lambda_{I_v}(\theta^*)}{\theta^*}$ is additive and can be decomposed into the summation of $\frac{\Lambda_{I_i}(\theta^*)}{\theta^*}$. The RRI is therefore the equivalent load contributed from connection i and is expressed by

$$\begin{aligned} RRI_i &= \frac{\Lambda_{I_i}(\theta^*)}{\theta^*} \\ &= \frac{\Lambda_{B_i}(P_i\theta^*)}{\theta^*} \\ &= \frac{\Lambda_{A_i}(\Lambda_{H_p}(P_i \cdot \theta^*))}{\theta^*} \end{aligned} \tag{19}$$

□

IV. Results and Conclusions

To verify the effectiveness of the proposed RRI, we examine if the packet dropping requirement of each connection is satisfied under the way the RRI is allocated at light and heavy load conditions. The heavier the load is (i.e. the summation of RRI of each connection is closer to the maximum value), the closer the measured packet dropping ratio to the requirement is. In the worst case, that is, the summation of RRI of each connection is equal to the maximum value, we expect that the measured packet dropping ratio of each connection will approximate it's requirement. Two scenarios are presented here. In the first scenario, the only traffic is 12.2k conversational service with $R^*_{D,i} = 0.02$. In the second scenario, there are three types of traffic: 12.2k conversational service with $R^*_{D,i} = 0.02$, 8.6k conversational service with $R^*_{D,i} = 0.02$, and 4.75k conversational service with $R^*_{D,i} = 0.05$. In the simulations, we set SIR_0 as $-14dB$, I_{th}/P_i^0 as 52, the permission probability $\tilde{r}_p$ as 0.9, and θ^* as 1.1.

Table I shows the RRI of each connection, the simulated and theoretical packet dropping ratio, and the theoretical maximum number of users in one cell in (a) scenario 1 with single traffic type and (b) scenario 2 with three traffic types. It can be found from Table I (a) that the simulated packet dropping ratio $\hat{R}_D$ is 2.5 times smaller than the theoretical result $\hat{R}^*_D$. As the number of users approaches 54, the $\hat{R}_D$ can be around 0.02. This indicates the proposed RRI is too conservative in radio resource allocation. The RRI attains about 88 percentage of best achievable resource utilization efficiency. It can be found from Table I (b) that the simulated packet dropping ratios of type-1 and type-2 connections are about 3 times less than the theoretical values, and the simulated value of type-3 connections is almost 5 times smaller than the theoretical values. As we increase the number of connections of three

TABLE I

THE THEORETICAL AND SIMULATION RESULTS

Scenario 1				
Parameters	RRI	$\hat{R}_D$	$\hat{R}_D^*$	number of users
Single type	1.08	0.00792	0.02	48

(a) Scenario 1 with single traffic types

Scenario 2				
Parameters	RRI	$\hat{R}_D$	$\hat{R}_D^*$	number of users
Type 1	1.08	0.00626	0.02	25
Type 2	0.67	0.00621	0.02	25
Type 3	0.31	0.0109	0.05	27

(b) Scenario 2 with three traffic types

types, the RRI can get about 80 percentage of the best achievable resource utilization efficiency in average. It can be also seen that the packet dropping ratio of type 3 connections is higher than those of type-1 and type-2 connections because the packet dropping ratio requirement for type 3 connections is set looser and its allocated power can be decreased. The proposed RRI can allocate proper radio resource to the connection according to the specific requirement for each connection without extra computation complexity and the resulting packet dropping ratio can be differentiated to fit its own requirement. From these two simulations, we can concluded that the RRI is a flexible and simple mapping from traffic parameters and QoS requirements in any number types of services, however, the achievable resource utilization efficiency is about 80%. The reasons are that several assumptions and approximations considering the worst case system load are adopted in the derivation of the RRI to simplify the derivation and keep QoS requirements of all connections guaranteed. Also, several bounds in our lemmas have tighter ones in some specific conditions. The precision and the efficiency of the proposed RRI will be improved in the future work.

REFERENCES

[1] 3rd Generation Partnership Project, Technical Specification Group RAN, Working Group 2 (WG2), "Radio Resource Management Strategies," *3GPP TR 25.922*, V3.0.0, Dec. 1999.

[2] S. J. Oh and K. M. Wasserman, "Adaptive resource allocation in power constrained CDMA mobile networks," *IEEE WCNC'99*, pp.510-514, 1999.

[3] A. Sampath, P. S. Kumar and J. M. Holtzman, "Power control and Resource management for a multimedia CDMA wireless system," *PIMRC'95*, pp. 21-25, 1995.

[4] 3rd Generation Partnership Project, Technical Specification Group Services and System Aspects, Working Group 2 (WG2), "QoS Concept and Architecture," *3GPP TS 23.107*, V4.1.0, Jun. 2001.

[5] H. Holma and A. Toskala, WCDMA for UMTS, Chap. 8, John Wiley & Sons, New York, 2000.

[6] C. S. Chang and J. A. Thomas, "Effective Bandwidth in High-Speed Digital Networks," *IEEE J. Select. Area Commun.,* vol. 13, no. 6, AUG. 1995.

[7] A. Dembo and O. Zeitouni., *Large Deviations Techniques and Applications.* Boston: Jones and Barlett Publishers, 1992.

[8] C. S. Chang, "Stability, queue length, and delay of deterministic and stochastic queueing networks," *IEEE Trans. Automa. Control,* vol. 39, no. 5, pp. 913-931, May 1994.
[9] C. S. Chang and T. Zajic, "Effective bandwidth of departure processes from queues with time varying capacities," *Proc. IEEE INFOCOM'1995,* Boston, pp. 1001-1009.
[10] F. P. Kelly, " Effective bandwidths at multi-class queues," *Queue. Sys.,* vol. 9, pp5-16, 1991.
[11] R. J. Gibbens and P. J. Hunt, "Effective bandwidths for the multi-type UAS channel," *Queue. Sys,* vol. 9, pp. 17-27, 1991.
[12] G. de Veciana, G. Kesidis, and J. Walrand, " Resource management in wide-area ATM networks using effective bandwidths," *IEEE J. Select. Area Commun.,* vol. 13, no. 6, pp. 1081-1090, Aug. 1995.
[13] P. Dupuis and R. S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviation*, New York: Wiley, 1997.
[14] R. L. Cruz, "A calculus for network delay, part I: network elements in isolation," *IEEE Trans. Info. Theory*, vol. 37, pp. 114-131, 1991.
[15] R. L. Cruz, "A calculus for network delay, part II: network analysis," *IEEE Trans. Info. Theory*, vol. 37, pp. 132-141, 1991.

A Q-LEARNING-BASED MULTI-RATE TRANSMISSION CONTROL SCHEME FOR RRC IN WCDMA SYSTEMS

Fang-Ching Ren, Chung-Ju Chang, AND Yih-Shen Chen
Department of Communication Engineering
National Chiao Tung University
Hsinchu 300, Taiwan ROC
E-mail: cjchang@cc.nctu.edu.tw
Tel. No.: 886-3-5731923 Fax No.: 886-3-5710116

Abstract

In this paper, a Q-learning-based multi-rate transmission control scheme (Q-MRTC) for radio resource control (RRC) in WCDMA systems is proposed. The RRC problem is modelled as a semi-Markov decision process (SMDP). And we successfully apply a real-time reinforcement learning algorithm, named Q-learning, to accurately estimate the transmission cost for the multi-rate transmission control. For the cost function approximation, we apply the feature extraction method to map the original state space into a more compact set which represents the *resultant interference profile*. Simulation results show that the Q-MRTC can achieve higher system throughput and better users' satisfaction index, by an amount of 87% and 50%, respectively, than the interference-based multi-rate transmission control scheme, while keeping the QoS requirement.

Keywords

multi-rate transmission, radio resource management, Q-learning, and CDMA communication system.

This work is published in Proc. of IEEE PIMRC 2002, Lisbon, Portugal, September 2002, Vol. 3, pp. 1422-1426. The journal version is submitted to *IEEE Transaction on Vehicular Technology.*

I. INTRODUCTION

The objective of a WCDMA system is to provide users a radio access link to services comparable to those currently offered by fixed networks, resulting in a seamless convergence of both fixed and mobile services. The WCDMA system is designed to integrate different types of services with heterogeneous QoS requirements. Therefore, an adequate radio resource control (RRC) is required to enhance the spectrum utilization while meeting those QoS requirements. In this paper, the multi-rate transmission control scheme for RRC is studied.

The multi-rate transmission control in the WCDMA system is to assign *power* and *processing gain* for different service requests so as to maximize the system capacity and to fulfill the users' satisfaction and QoS requirements. There is no absolute number of maximum available channels in the WCDMA system because WCDMA system is interference-limited. Its capacity is affected by multiple access interference (MAI), which is a function of the number of active users, the users' location, and heterogeneous QoS requirements. Many researches for CDMA capacity estimation are based on MAI and other considerations [1]-[3]. In [4], an interference-based CAC scheme in multimedia CDMA cellular systems was proposed. Instead of a fixed system capacity, this interference-based scheme can adaptively assign a channel according to the actual system capacity dependent of interference such that the system utilization can be improved.

Maximizing system capacity while meeting QoS constraints suggests a constrained semi-Markov decision process (SMDP) [5]. The SMDP has successfully applied to many network control problems; however, it requires extremely large state space to model these problems exactly. Also, *a priori* knowledge of state transition probabilities is required. Alternatively, many researchers turned to use the reinforcement learning (RL) algorithms to solve the large state space problems [6] -[8]. The most obvious advantage of RL algorithm is that it could approach an optimal solution from the on-line operation if the RL algorithm is converged. Also, it does not require *a priori* knowledge of state transition probabilities.

In this paper, we propose a Q-learning-based multi-rate transmission control scheme

(Q-MRTC) for RRC in the WCDMA system to maximize the system utilization and fulfill the users' satisfaction, subject to a QoS requirement of packet error probability. For the multi-rate transmission control, a cost function is defined to appraise the cumulative cost of the consecutive decisions for the Q-MRTC. Without knowing the state transition behavior, the cost function is calculated by a real-time RL technique known as Q-learning [9]. To aggregate state space and improve the convergence property, a *feature extraction* method is applied. The state space of the Q-function is mapped into a more compact set which represents *resultant interference profile*. Simulation results show that the Q-MRTC scheme can have higher system throughput and better users' satisfaction than the interference-based scheme [4] in an amount of 87% and 50%, respectively, while keeping the QoS constraints.

II. SYSTEM MODEL

Two types of services are considered in this paper: real-time service as type-1 and non-real-time service as type-2. The system provides connection-oriented transmission for real-time traffic and best-effort transmission rate allocation for non-real-time traffic, as the service discipline adopted in [10]. To guarantee the timely constraint of real-time service, a UE always holds a dedicated physical channel (DPCH) while it transmits real-time packets regardless the variation of the required transmission rate. The real-time UE may generate variable rate information whose characteristics are indicated in its request profile. On the other hand, a UE should contend for the reservation of a DPCH to transmit a burst of non-real-time packets and will release the DPCH immediately while the burst of data is completely transmitted. The non-real-time data are transmitted burst by burst. Due to different service requirements, the RRC performs two kinds of decision. For a real-time request, the request will be accepted or rejected. On the other hand, for a non-real-time request, an appropriate transmission rate will be allocated. A non-real-time request specifies the range of the required transmission rates for itself, and would be blocked if the WCDMA system cannot provide a suitable transmission rate to satisfy its required transmission rate.

The transmission power of a physical channel should be adjusted dependent of its spreading factor, coding scheme, rate matching attributes, and BER requirement. Here, we assume that all physical channels adopt the same coding scheme and have the same rate matching attributes and BER requirement. Therefore, power allocation for a physical channel is simply dependent of its spreading factor and it is in inverse proportion.

III. The DESIGN OF Q-MRTC

A. State, Action, and Cost Function

Since there is no re-transmission for real-time packets, an error real-time packet will be dropped. The error non-real-time packets will be recovered via ARQ (automatic repeat request) scheme. The packet error probability, denoted by P_E, is considered as the system performance measure. And the maximum tolerable packet error probability, denoted by P_E^*, is defined as the system QoS requirement. In this paper, we assume that all packets have the same length. Also, a data packet is assumed to be transmitted in a $10ms$ frame by a basic rate DPCH, and therefore a multi-rate DPCH can transmit multiple data packets in a $10ms$ frame.

As to the service profile, a real-time request provides the mean rate and rate variance to indicate its transmission rate requirement, while a non-real-time request provides the maximum and minimum rate requirements. The mean and variance of the interference from the existing connections, denoted by I_m and I_v, respectively, can be used as an interference profile for indicating the system loading condition [2]. As noted, radio resource control of the WCDMA system can be regarded as a discrete-time SMDP problem, where major events are request arrivals.

The request arrivals are treated as events that trigger the state transition in which the radio resource control is executed. For the arrival of the k-th request, the *system state* at x_k is defined as

$$x_k = (I_m, I_v, i, \mathbf{R}_i), \tag{1}$$

where $\mathbf{R}_i$ is transmission rate requirement of the type-i request, $i = 1, 2$. The $\mathbf{R}_1 = (r_m, r_v)$, where r_m and r_v denote the mean rate and the rate variance of a real-time request, respectively; the $\mathbf{R}_2 = (r_{\max}, r_{\min})$, where $r_{\max}$ and $r_{\min}$ denote the maximum rate and the minimum rate requirements of a non-real-time request, respectively.

Based on the system state x_k, the radio resource controller will determine an *action*, denoted by A_k, for the k-th request arrival. If the request is non-real-time, $A_k \in \{0, r\}$ and $A_k = r$ is to accept the request given with rate r, $r_{\min} \le r \le r_{\max}$, while $A_k = 0$ is to reject the request. For the state-action pair (x_k, A_k), an immediate cost is given by

$$c(x_k, A_k) = [P_E(x_k, A_k) - P_E^*]^2, \tag{2}$$

where $P_E(x_k, A_k)$ is the packet error probability if the state-action pair (x_k, A_k) has been selected. We further define a cost function, denoted by $Q(x, A)$, which is the total expected discounted cost counted from the initial state-action pair (x, A) over an infinite time. It is given by

$$Q(x, A) = E\left\{\sum_{k=0}^{\infty} \gamma^k c(x_k, A_k) | x_0 = x, A_0 = A\right\}, \tag{3}$$

where $0 \le \gamma < 1$ is a discounted factor. The optimal multi-rate transmission control is to determine an optimal action which minimizes the Q value with respective to the current state. The minimization of Q value represents the maximization of the system capacity while the QoS requirements are satisfied.

Let $P_{xy}(A)$ be the transition probability from state x with action A to the next state y. Then $Q(x, A)$ can be expressed as

$$Q(x, A) = C(x, A) + \gamma \sum_y P_{xy}(A) Q(y, B), \tag{4}$$

where $C(x, A) = E\{c(x, A)\}$. Eq. (4) indicates that the Q function of the current state-action pair can be represented in terms of the immediate cost of the current state-action pair and the Q function of the next state-action pairs.

Based on the principle of Bellman's optimality [12], the optimal action, denoted by A^*, can be obtained by a two-step optimality operation. The first step is to find an intermediate minimal $Q(x, A)$, denoted by $Q^*(x, A)$, where the intermediate cost function for every possible next state-action pair (y, B) is minimized and the optimal action is performed with respective to each next state y. For all (x, A), $Q^*(x, A)$ is given by

$$Q^*(x,A) = C(x,A) + \gamma \sum_y P_{xy}(A) \left\{ \underset{B}{\text{Min}} \ [Q^*(y,B)] \right\} \tag{5}$$

Then we can determine the optimal action A^* with respective to the current state x such that $Q^*(x,A)$ is minimal, which can be expressed as

$$Q^*(x,A^*) = \underset{A}{\text{Min}} \ [Q^*(x,A)]\,. \tag{6}$$

However, it is difficult to find the $C(x,A)$ and $P_{xy}(A)$ to solve Eq. (5). In this paper, we adopt the Q-learning algorithm [9], [11] to find the optimal resource allocation without *a priori* knowledge of $C(x,A)$ and $P_{xy}(A)$.

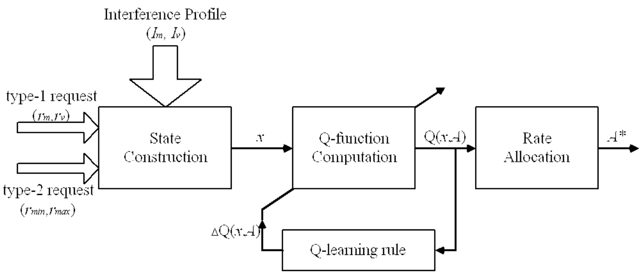


Fig. 1. Structure of the Q-learning-based multi-rate transmission control scheme (Q-MRTC)

B. Q-MRTC

Fig. 1 shows the implementation structure of the Q-learning-based multi-rate transmission control scheme (Q-MRTC). The Q-learning algorithm tries to find optimal value $Q^*(x,A)$ in a recursive method using available information (x, a, y, $c(x,a)$), where x and y are the current and next states, respectively; and $a \in A$ and $c(x,a)$ are the action for current state and its immediate cost of the state action pair, respectively. The Q-learning rule is formulated by

$$Q(x,a) = \begin{cases} Q(x,a) + \eta \Delta Q(x,a), & \text{if } (x,a) \text{ is chosen} \\ Q(x,a) & \text{otherwise.} \end{cases} \tag{7}$$

where η is the learning rate, $0 \le \eta \le 1$, and

$$\Delta Q(x,a) = \left\{ c(x,a) + \gamma \underset{B}{\text{Min}} \ [Q(y,B)] \right\} - Q(x,a). \tag{8}$$

It has been shown in [9] that if the value of each admissible pair is visited infinitely often and the learning rate is decreased to zero in a suitable way, then as the learning time goes to infinity, $Q(x, a)$ in Eq. (7) converges to $Q^*(x, A)$ with probability 1.

The rate allocation block of the Q-MRTC in Fig. 1 is the implementation of Eq. (6), which determines the adequate rate allocation or call rejection for the new call request.

As noted, if the state space is too large, some states would be less visited and the convergence of Q-learning algorithm would take a long time consequently. To tackle this problem, the *feature extraction* method is applied in the proposed Q-MRTC. *Feature extraction* is a method that maps the original state space into some feature vectors associated with it. Feature vectors are used to represent the important characteristics of the state space[12]. In the WCDMA system, after the RRC decision is made, the change of interference is the most obvious corresponding response. That is, the value of $Q(x, A)$ mainly depends on the resultant interference profile if the state-action pair (x, A) is performed. Therefore, $Q(x, A)$ can be represented as a function of resultant interference profile.

This resultant interference profile of (x, A) is denoted by $(I_m+\Delta I_m, I_v+\Delta I_v)$, where (I_m, I_v) indicates the system loading of existing connections at state x and $(\Delta I_m, \Delta I_v)$ indicates the change of interference profile due to action A. Then $(\Delta I_m, \Delta I_v)$ is obtained by

$$(\Delta I_m, \Delta I_v) = \begin{cases} (r_m, r_v), \text{ accept a real-time request} \\ (r, 0), \text{ accept a nonreal-time request} \\ (0, 0), \text{ reject a request} \end{cases} \tag{9}$$

Now, the state-action pair (x, A) is converted to interference profile $(I_m + \Delta I_m, I_v + \Delta I_v)$ which requires less state space.

IV. SIMULATION EXAMPLE AND DISCUSSIONS

In this simulation, we consider the WCDMA communication system supporting two types of services: real-time and non-real-time services with QoS requirement P_E^*=0.01. Two kinds of traffic are transmitted via the real-time service: one is 2-level transmission rate traffic and the other is M-level transmission rate traffic.

TABLE I
THE SIMULATED TRAFFIC PARAMETERS

Traffic Type	Traffic Parameters
2-level real-time	Call holding time: 30s Mean talkspurt duration: 1.0s Mean silence duration: 1.35s
M-level real-time	Call holding time: 30s Peak rate: 4-fold of basic rate Mean rate: 2-fold of basic rate
Non-real-time	Mean burst size: 200 packets $r_{\min}$: 1-fold of basic rate $r_{\max}$: 8-fold of basic rate

They are modelled by 2-level and M-level MMDP (Markov modulated deterministic process),respectively. On the other hand, the non-real-time service is considered to transmit variable-length data bursts. The arrival process of the data burst is Poisson process and the data length is assumed to be with a geometric distribution.

The detail traffic parameters are listed in Table. I. The traffic intensities of the services are 20%(2-level real-time), 40%(M-level real-time), and 40%(non-real-time). A basic rate in the WCDMA system is assumed to be a physical channel with SF=256. For each connection, DPCCH is always active to maintain the connection reliability.

For evaluation the performance of the Q-MRTC, the conventional interference-based scheme [4] is used as a benchmark. Fig. 2 illustrates the throughput versus the request arrival rate. The Q-MRTC has higher throughput than the interference-based scheme. Generally speaking, the Q-MRTC can improve the maximum throughput by an amount of 87% over the interference-based scheme. The reason is that the Q-MRTC performs an on-line reinforcement learning algorithm to take the behavior of interference variation into consideration for multi-rate transmission control. That is, the Q-MRTC takes advantage of the multiplexing gain from the variable-rate services. Actually, some existing connections may terminate or handoff between two consecutive arrivals. The interference decreases consequently. Therefore, the multi-rate transmission cost would be over-estimated in the interference-based scheme.

We further define an overall users' satisfaction index (USI) which is a linear combination of $\frac{A_{a1}}{A_{d1}}$ (type-1) and $\frac{A_{a2}}{A_{d2}}$ (type-2), where the A_{a1} (A_{a2}) is the admitted transmission rate for type-1 (type-2) and the A_{d1} (A_{d2}) is the desired transmission rate for type-1 (type-2); $A_{d1} = 1$ and $A_{d1} = r_{\max}$. That is, USI is expressed as

$$USI = \alpha \frac{A_{a1}}{A_{d1}} + (1 - \alpha) \frac{A_{a2}}{A_{d2}}, \tag{10}$$

where α is the weighting factor. Fig. 3 depicts the USI versus the request arrival rate for $\alpha = 0.5$. It can be found that the Q-MRTC has higher USI than the interference-based scheme. This is because the Q-MRTC can accurately estimate the multi-rate transmission cost.

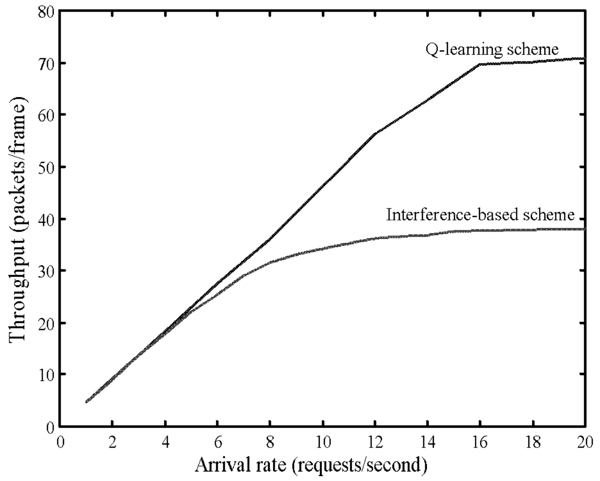


Fig. 2. Throughput versus the request arrival rate

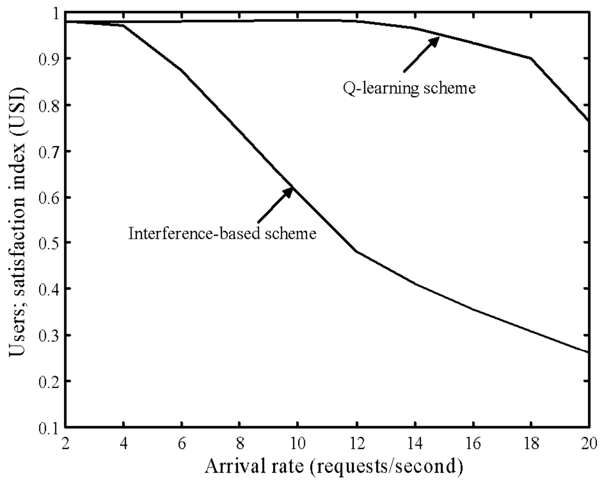


Fig. 3. The users' satisfaction index versus the request arrival rate

Fig. 4 depicts the packet error probability versus the request arrival rate. It can be seen that the average packet error probability of the Q-MRTC is larger than that of the interference-based scheme which is almost zero. However, the Q-MRTC can still hold the packet error probability within the QoS requirement.

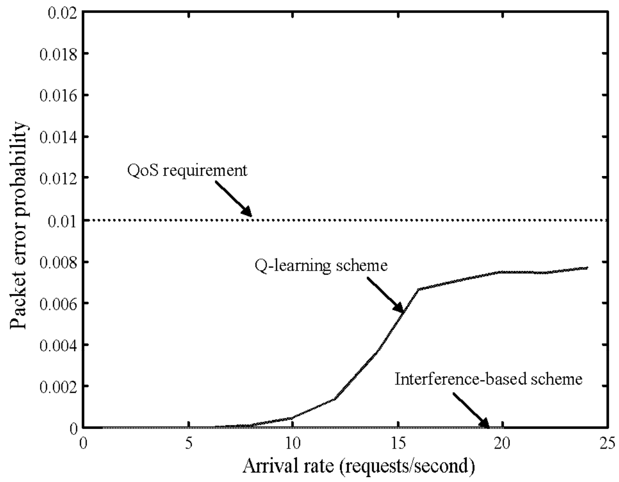


Fig. 4. packet error probability versus the request arrival rate

V. CONCLUDING REMARKS

In this paper, we propose a Q-learning-based multi-rate transmission control scheme for radio resource control in WCDMA communication systems. We successfully apply the Q-learning algorithm to accurately estimate the transmission cost for the multi-rate transmission control. We also apply the feature extraction method to efficiently map the original state space into the *resultant interference profile*. Compared with the interference-based scheme, Q-MRTC can improve the throughput of the WCDMA system by an amount of 87% under the constraint of the QoS requirement. In addition, the Q-MRTC provides better users' satisfaction by an amount of 50%. Since the Q-learning algorithm performs a closed-loop control by applying the system performance measurement as a feedback to adjust the multi-rate transmission cost, the Q-MRTC can have self-tuning capability to adaptively estimate the transmission cost.

ACKNOWLEDGEMENTS

This work was supported by the National Science Council, Taiwan under contract number NSC 91-2219-E-009-006, Chung Hwa Telecommunication Laboratory under contract number TL-90-1101, and Lee and MTI Center, National Chiao Tung University.

References

[1] K. S. Gilhousen, I. M. Jacobs, R. Padovani, A. J. Viterbi, L. A. Weaver, and C. E. Wheatley, "On the capacity of a cellular CDMA system," *IEEE Trans. Veh. Technol.*, vol. 40, no. 2, pp. 303-312, May. 1991.

[2] W. Huang and V. K. Bhargava, "Performance evaluation of a DS-CDMA cellular system with voice and data services," *IEEE PIMRC'96*, pp. 588-592.

[3] J. B. Kim and M. L. Honig, "Resource allocation for multiple classes of DS-CDMA traffic," *IEEE Trans. Veh. Technol.*, vol. 49, no. 2, pp. 506-519, March 2000.

[4] N. Dimitriou and R. Tafazolli, "Quality of service for multimedia CDMA," *IEEE Comm. Mag.*, vol. 38, no. 7, pp. 88-94, July 2000.

[5] K. W. Ross, Multiservice Loss Models for Broadband Communication Networks. Berlin, Germany: Springer-Verlag, 1995.

[6] J. Nie and S. Haykin, "A Q-learning-based dynamci channel assignment technique for mobile Communication systems," *IEEE Trans. Veh. Technol.*, vol. 48, no. 5, pp. 1676-1687, Sep. 1999.

[7] H. Tong and T. X. Brown, "Admission call admission control under quality if service constraints: a reinforcement learning solution," *IEEE Trans. Veh. Technol.*, vol. 18, no. 2, pp. 209-221, Feb. 2000.

[8] P. Marbach, O. Mihatsch, and J. N. Tsisiklis, "Call admission control and routing in integrated services networks using neuro-dynamic programming," *IEEE Trans. Veh. Technol.*, vol. 18, no. 2, pp. 197-208, Feb. 2000.

[9] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, pp. 279-292, 1992.

[10] S. Choi and K. G. Shin, "An uplink CDMA system architecture with diverse QoS guarantees for heterogeneous traffic," *IEEE/ACM Trans. Netorking*, vol. 7, no. 5, pp. 616-628, Oct. 1999.

[11] S. Haykin, Neural Networks 2nd. Prentice Hall, 1999.

[12] D. P. Bertsekas and J. N. Tsitsiklis, Neuro-Dynamic Programming. Athena Scientific, 1996.

Achieving Weighted Fairness for Wireless Multimedia Services

Chih-Yung Shih[+], Ray-Guang Cheng[++], Chung-Ju Chang[+] , and Yih-Shen Chen[+]
[+]Dept. of Communication Engineering, National Chiao-Tung University, Hsinchu, TAIWAN
[++]Dept. of Electronic Engineering, National Taiwan University of Science and Technology, Taipei, TAIWAN
crg@ieee.org, cjchang@cc.nctu.edu.tw

Abstract

It is important to provide differentiated quality of services (QoSs) for multimedia services over Wireless LAN (WLAN). Fairness is one of the key issues for QoS supporting. In this paper, we proposed a method to achieve weighted fairness for two classes of service operating under the enhanced distributed coordinator function (EDCF) mode of 802.11e. A queueing model is adopted to analyze the behavior of the two classes. The analytical results were verified by computer simulation. Numerical results showed that the weighted goals are easily achieved by adjusting both the arbitration inter frame space (AIFS) and the contention window (CW).

Keywords

weighted fairness,WLAN, EDCF

This work was published in IEEE VTC-Fall 2003, Orlando, Florida.

I. INTRODUCTION

In order to provide differentiated QoSs in Wireless LAN (WLAN), it is essential to have a method that can allocate bandwidth for classes of stations (STAs) according to their priorities. A weighted fairness is achieved if the bandwidth could be allocated according to a predetermined goal.

In IEEE 802.11, two mechanisms are defined to access the channel: a contention-based distributed coordinator function (DCF) and a polling-based point coordinator function (PCF). In order to support different priorities, enhanced DCF (EDCF) mode is further defined in 802.11e. DCF is based on CSMA/CA protocol with slotted binary exponential backoff scheme. In DCF and EDCF modes, the bandwidth is shared by all of the STAs and the probability for a STA to access the WLAN is depended on the number of active stations, the contention window (CW) size, and the inter-frame space (IFS) time. The major difference between EDCF and DCF is that the CW and IFS are the same for all STAs in DCF but could be different in EDCF.

Several algorithms were proposed for investigating the behavior of DCF mode. Bianchi, Fratta, and Oliveri [1] proposed a mechanism to adaptively adjust the CW according to the estimated contending stations. It showed that a better throughput is achieved as the increased of network loading. Cali, Conti, and Gregori [2] proposed a dynamic tuning of the backoff algorithm to achieve a theoretical throughput limit. In [3], the authors designed a throughput enhancement mechanism for DCF by adjusting the contention window-resetting scheme. In these papers, the authors were focused only on the behavior of a single class. Vaidya, Bahl, and Gupta [4] presented a distributed packet scheduling algorithm. The bandwidth of different flows was allocated in proportion to their weights. Banchs and Perez [5] proposed an extension of the DCF function to provide weighted fairness by tuning the CW. However, these two schemes only focused on fairness and did not consider the enhancement of channel utilization. Qiao and Shih [6] attempted to deal with both weighted fairness and maximized utilization simultaneously by analytical method, but their model did not consider backoff mechanism and cannot be backward-compatible to DCF mode.

In this paper, we propose a method to achieve weighted fairness for two classes of services operating under EDCF mode. We derive the relationship between throughput, conditional collision probability, and channel busy probability, for high- and low-class stations, respectively. The rest of the paper is organized as follows. In Section II, basic concepts of DCF and EDCF modes are described. Section III presents the analytical results of the weighted fairness problems for two classes of services in EDCF mode. In Section IV, numerical examples and simulation results are presented to verify the effectiveness of the analysis. Finally, the concluding remarks are given in Section V.

II. BACKGROUND

In this section, we briefly review some background information on 802.11. Basic concepts of DCF and EDCF modes will be introduced.

The time interval between frames, named InterFrame Spaces (IFSs), are used to control the priority for accessing the channel in 802.11. Three types of IFSs are defined in 802.11: Short IFS (SIFS), the Point coordination function IFS (PIFS), and the Distributed coordination function IFS (DIFS). SIFS is the shortest interval and is used for transmission of acknowledgments (ACKs), polling responses in point coordination function (PCF) mode, and fragments belonging to the same MAC service data unit (MSDU). The PIFS, which is greater than SIFS but smaller than DIFS, is used to initiate the Contention Free Period in PCF mode.

In DCF mode, a STA with a new packet is allowed to transmit only if the channel is sensed to be idle for DIFS. Otherwise, the transmission is deferred and the exponential backoff procedure is invoked. The exponential backoff procedure is implemented via using a backoff counter C calculated by

$$C=Rand(0,w-1), \qquad (1)$$

where w is set to be equal to CWmin, at the first transmission attempt and is doubled after each unsuccessful transmission until it reaches a maximum value CWmax = 2m CWmin. C is decreased whenever the channel is sensed idle for , is frozen when any packet transmission is detected, and is reactivated when the medium is sensed idle for DIFS again. The STA transmits immediately when C reaches to zero. At the end of the receiving packet, the destination STA immediately acknowledges the successful reception by transmitting an ACK after SIFS. Since the SIFS is shorter than DIFS, the other STAs will not detect the channel as idle until the end of ACK. The originating STA assumes that the transmission is failed if it does not receive ACK within a pre-defined period or it detects packets transmitted by other STAs. For the failed transmission, the originating STA will reschedule the packet according to the backoff procedure described above.

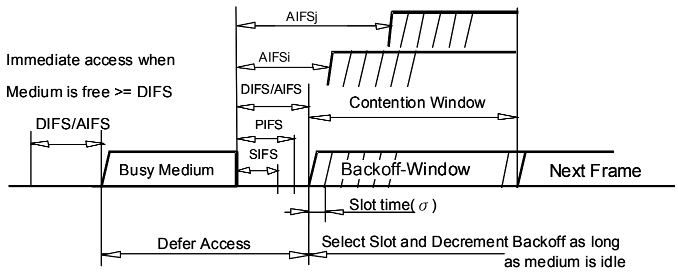
The access mechanism of 802.11e EDCF, as shown in Fig. 1, is similar to that of DCF. Four different priorities, called access categories (ACs), are supported in EDCF. Each AC has its associated values of CWmin, CWmax, and arbitration IFS (AIFS). The AIFS for the i-th AC, denoted by AIFSi, is defined by

$$\mathrm{AIFS}_i=\mathrm{SIFS}+L_i \times \sigma, \ \text{for } 1 \leqq i \leqq 4, \qquad (2)$$

where Li is an integer ranging from 1 to 255. Specifically, AIFSi is PIFS and DIFS for *Li* equal to 1 and 2, respectively. One should note that an AC with a smaller CWmin or AIFS implies a higher priority to access the channel. In EDCF, the backoff counter for priority i, denoted by ci, is modified as

$$c_i = Rand\ (0,w_i\text{-}1) + X, \text{ for } 1 \leqq i \leqq 4, \qquad (3)$$

where w_i is the CW for priority *i; X* is equal to 1 if *Li*=1, otherwise, it's set to be 0. *X* is introduced to ensure



that operation of EDCF mode will not disturb the PCF mode.

Figure 1. Basic access mechanism under 802.11e EDCF MAC protocol.

III. THEORETICAL ANALYSIS

In this section, we investigate the weighted fairness of 802.11 under EDCF mode. We consider a system with NL low-class STAs and NH high-class STAs, each STA adopts a full queue traffic model [7]. For backward compatible to DCF, AIFSH=PIFS and AIFSL=DIFS are assigned for the two classes. An ideal channel condition without hidden terminals and with error-free transmission is assumed. We adopted the weighted fairness function given by [6]

$$\frac{ST_H}{ST_L} = \frac{\phi_H}{\phi_L}, \qquad (4)$$

where ϕ_H, ϕ_L, ST_H, and ST_L are the assigned weights and the successful transmission probabilities for high- and low-class STAs, respectively. We assumed that the average frame length for both classes is the same. Therefore, the traffic flows for each class may share the channel according to the pre-defined weights and the weighted fairness is then achieved if Eq. (4) can be guaranteed.

A. System Parameters and Observation Points

According to the backoff procedure, the decrement of backoff counter is stopped if the channel is sensed busy. Therefore, the time interval between two consecutive backoff counter decrements is not fixed. Due to the fact that AIFSH<AIFSL, we define a slot time as the (variable) time interval between two consecutive

backoff counter decrements for the high-class STAs. The observation points are then selected at the end of each time slot such that the backoff counter for either low- or high-class STAs can only be decreased at the observation points.

Let $c_L(n)$ and $c_H(n)$ be the stochastic processes representing the backoff counter of a given low- and high-class STA saw at the observation point n, respectively. The first property we found is that the $c_H(n)$ is always decreased but $c_L(n)$ could be frozen for any observation point n. Since AIFSH=PIFS, the new backoff counter for high-class STAs is initially chosen in range of $(1, WH_i)$ after a successful or a collided transmission. However, the selection of initial backoff counter of a high-class STA must be done in a slot and should be decreased by 1 at the observation point. Thus, we have the second property that $c_H(n)$ will fall into the range of $(0, WH_i\text{-}1)$.

$c_L(n)$ and $c_H(n)$ are non-Markovian because the backoff counter depends also on its retransmission history. Therefore, we adopt the definition of "backoff stage," which is defined as the number of retransmission attempts for a frame, to account for the retransmission history [8]. Let M_L (M_H) and WL_0 (WH_0) be the maximum backoff stage and the CWmin of the low-class (high-class) STAs, respectively. We can calculate the CW of the low-class (high-class) STAs at the i-th backoff by WL_i =2i× WL_0 ($WH_i = 2^i \times WH_0$), where i is called the backoff stage and i≦ M_L (M_H).

B. *Behavior of a Single Station with Different AIFS*

Let $s_L(n)$ and $s_H(n)$ be the stochastic processes representing the backoff stage for a given low-class (high-class) STA at time n, respectively. We first consider the behavior of a single low-class STA with $c_L(n)$ and $s_L(n)$ at observation point n. Similar to the approximation adopted in [8], we assume that at each transmission attempt, each frame (of a low-class STA) collide with a constant and independent probability P_L regardless of the number of retransmissions. P_L is referred to as conditional collision probability of a low-class STA, meaning that a collision seen by a frame (of a low-class STA) being transmitted on the channel. In other words, it is the probability that at least one of the other STAs (i.e. N_H high-class STAs and N_L-1 low-class STAs) counts down to zero while the low-class STA transmits. We further assume that a low-class STA with nonzero backoff counter may sense the channel as busy with a constant and independent probability q. q is referred to as channel busy probability sensed by a low-class STA. In other words, it's the probability that at least one of other STAs (i.e. N_H high-class STAs and N_L-1 low-class STAs) transmits while the low-class STA has nonzero backoff counter. Therefore, q is independent of the number of retransmissions.

Once independence is assumed, and P_L and q are supposed to be constant values, it is possible to model the bi-dimensional process $\{s_L(n), c_L(n)\}$ with the discrete-time Markov chain as depicted in Fig. 2. Here, we assumed that the CW of a frame originates from a low-class (high-class) STA will be reset to WL_0 (WH_0) if has been retransmitted for M_L (M_H) times. In this Markov chain, the non-null one-step transition probabilities of a single low class station are

$$\begin{cases} P\{i,k \mid i,k\} = q, & k \in (1, WL_i - 1) \quad i \in (0, M_L), & (5) \\ P\{i,k \mid i,k+1\} = 1-q, & k \in (0, WL_i - 2) \quad i \in (0, M_L), & (6) \\ P\{0,k \mid i,0\} = (1-P_L)/WL_0, & k \in (0, WL_0 - 1), \; i \in (0, M_L - 1), & (7) \\ P\{i,k \mid i-1,0\} = P_L / WL_i, & k \in (0, WL_i - 1) \quad i \in (1, M_L), & (8) \\ P\{0,k \mid M_L, 0\} = 1/WL_0, & k \in (0, WL_0 - 1). & (9) \end{cases}$$

where $P\{i_1, k_1 \mid i_0, k_0\} \equiv P\{s(n+1) = i_1, c(n+1) = k_1 \mid s(n) = i_0, c(n) = k_0\}$.

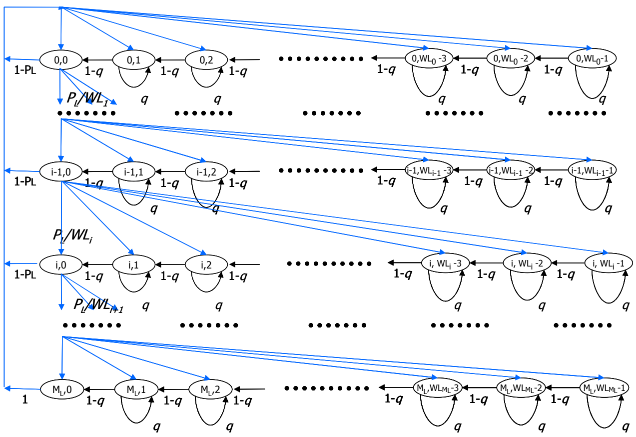


Figure 2. Markov chain model for the backoff counter with $AIFS_L$=DIFS.

Let $b_{i,k} = \lim_{n\to\infty} \Pr\{s_L(n) = i, c_L(n) = k\}, i \in (0, M_L), k \in (0, WL_i - 1)$ be the steady-state probability of the low-class STA. We can derive $b_{i,k}$ and $b_{0,k}$ by

$$b_{i,0} = P_L \cdot b_{i-1,0} = P_L^{\,i} \cdot b_{0,0}, \qquad i \in (1, M_L), \qquad (10)$$

$$b_{i,k} = \frac{(WL_i - k) \cdot P_L^{\,i} \cdot b_{0,0}}{WL_i \cdot (1-q)}, \qquad i \in (1, M_L), k \in (1, WL_i - 1), \qquad (11) \qquad b_{0,k} = \frac{(WL_0 - k) \cdot b_{0,0}}{WL_0 \cdot (1-q)}, \qquad k \in (1, WL_0 - 1). \qquad (12)$$

Then, by using the normalization condition for stationary probabilities, we have

$$b_{0,0} = \frac{2(1-q)(1-2P_L)(1-P_L)}{WL_0(1-P_L)(1-(2P_L)^{M_L+1}) + (1-2q)(1-2P_L)(1-P_L{}^{M_L+1})}. \quad (13)$$

Similarly, consider the stochastic process $c_H(n)$ and $s_H(n)$ for a high-class STA observed at n. We also assumed that at each transmission attempt, each frame (of a high-class STA) collide with a constant and independent probability P_H regardless of the number of retransmissions. P_H is referred to as conditional collision probability of a high-class STA, meaning that a collision seen by a frame (of a high-class STA) being transmitted on the channel. In other words, it's the probability that at least one of the other STAs (i.e. N_H–1 high-class STAs and N_L low-class STAs) counts down to zero while the high-class STA transmits.

P_H is supposed to be a constant value because of the independence assumption. It is also possible to model the bi-dimensional process $\{ s_H(n), c_H(n) \}$ with the discrete-time Markov chain as depicted in Fig. 3. Similarly, the non-null one-step transition probabilities are

$$\begin{cases} P\{i,k \mid i,k+1\} = 1, & k \in (0, WH_i - 2) \;\; i \in (0, M_H), \quad (14) \\ P\{0,k \mid i,0\} = (1-P_H)/WH_0, & k \in (0, WH_0 - 1) \;\; i \in (0, M_H - 1), \quad (15) \\ P\{i,k \mid i-1,0\} = P_H / WH_i, & k \in (0, WH_i - 1) \;\; i \in (1, M_H), \quad (16) \\ P\{0,k \mid M_H, 0\} = 1/WH_0, & k \in (0, WH_0 - 1). \quad (17) \end{cases}$$

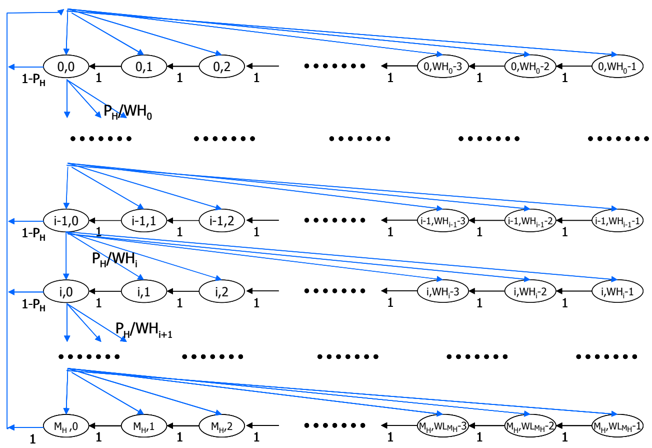


Figure 3. Markov chain model for the backoff counter with $AIFS_H$=PIFS.

Here, we do not have P$\{i,k|i,k\}$ as in Eq.(5) due to the first observed property that $c_H(n)$ is always decreased but $c_L(n)$ could be frozen for any observation point n. Also, the range of $c_H(n)$ (i.e. k) has to be modified to (0,WH_i-1) according to the second observed property.

Let $d_{i,k} = \lim_{n\to\infty} \Pr\{s_H(n) = i, c_H(n) = k\}, i \in (0, M_H), k \in (0, WH_i - 1)$ be the steady-state probability of the high-class STA.

$$d_{i,0} = P_H \cdot d_{i-1,0} = P_H{}^i \cdot d_{0,0},\ i \in (1, M_H), \quad (18) \qquad d_{i,k} = \frac{(WH_i - k) \cdot P_H{}^i d_{0,0}}{WH_i},\ \ i \in (1, M_H), k \in (0, WH_i - 1), \quad (19)$$

$$d_{0,k} = \frac{(WH_0 - k) \cdot d_{0,0}}{WH_0}, \qquad k \in (0, WH_0 - 1). \quad (20)$$

And, $d_{0,0}$ can be found as

$$d_{0,0} = \frac{2(1-2P_H)(1-P_H)}{WH_0(1-P_H)(1-(2P_H)^{M_H+1}) + (1-2P_H)(1-P_H{}^{M_H+1})}. \quad (21)$$

C. Frame Transmission Probabilities with Different AIFS

Denote τ_H as the transmission probability that a high-class STA transmits in a randomly chosen time slot, τ_H can be obtained by summarizing of the state probability $d_{i,0}$ found in (18) as

$$\begin{aligned}\tau_H &= \sum_{i=0}^{M_H} d_{i,0} \\ &= \frac{2(1-2P_H)}{WH_0(1-P_H)(1-(2P_H)^{M_H+1})/(1-P_H{}^{M_H+1}) + (1-2P_H)}. \end{aligned} \quad (22)$$

Since AIFS$_H$<AIFS$_L$, the low-class STA may not affect the high-class STA. Therefore, τ_H is equal to the probability that more than one high-class STA that choose the same backoff counter. Similarly, we can derive the probability that more than one low-class STA that choose the same backoff counter value, denoted by τ'_L, as

$$\begin{aligned}\tau_L{}' &= \frac{2 \cdot \sum_{i=0}^{M_L} P_L{}^i}{\sum_{i=0}^{M_L} P_L{}^i (2^i \cdot WL_0 + 1)} \\ &= \frac{2(1-2P_L)}{WL_0(1-P_L)(1-(2P_L)^{M_L+1})/(1-P_L{}^{M_L+1}) + (1-2P_L)}. \end{aligned} \quad (23)$$

The transmission probability that a low-class STA transmits in a randomly chosen time slot, denoted by τ_L, is the sum of $b_{i,0}$ found in (10) and is given by

$$\begin{aligned}\tau_L &= \sum_{i=0}^{M_L} b_{i,0} \\ &= \frac{2(1-q)(1-2P_L)}{WL_0(1-P_L)(1-(2P_L)^{M_L+1})/(1-P_L^{M_L+1}) + (1-2q)(1-2P_L)}. \end{aligned} \quad (24)$$

Then we can derive the P_H, P_L, and q based on their definition. P_H is the probability that a ready-to-transmit high-class STA collides with any of the N_H-1 high-class STAs or N_L low-class STAs. q is the probability that the channel is sensed busy by a low-class STA with nonzero backoff counter. The channel will be busy if any of the N_H high-class STAs or N_L-1 low-class STAs transmits at the same time. P_L is the probability that a ready-to-transmit low-class STA collides with any of the N_H high-class STAs or N_L–1 low-class STAs. In this case, the transmission probability for low-class STAs is τ'_L because the counter of the ready-to-transmit low-class STA is zero. Then we can have,

$$\begin{cases} P_H = 1-(1-\tau_H)^{N_H-1}(1-\tau_L)^{N_L}, \\ P_L = 1-(1-\tau_H)^{N_H}(1-\tau'_L)^{N_L-1}, \\ q = 1-(1-\tau_H)^{N_H}(1-\tau_L)^{N_L-1}. \end{cases} \tag{25}$$

Finally, we can easily derive successful transmission probabilities of high-class and low-class STAs, respectively, as

$$\begin{cases} ST_H = \tau_H \cdot (1-P_H), \\ ST_L = \tau_L \cdot (1-P_L). \end{cases} \tag{26}$$

Comparing with Eq. (4), we can use the numerical method to find the relationship between WH_0 and WL_0 from Eqs.(22), (23), (24), and (25) that can satisfied with Eqs. (4) and (26) by fixing the values of ϕ_H, ϕ_L, N_H, N_L, M_H, and M_L.

IV. Numerical Results

To validate the analysis, simulations were performed based on MATLAB. The values of PHY-related parameters were referred to IEEE 802.11b [9]. The symbol transmission rate was set to 11 Mbps. The frame format was the one defined by the 802.11e MAC specifications, and the PHY header and IFS intervals were those defined for 802.11b PHY. The PHY overhead time including preamble and header length is 196 μs, σ is 20 μs, SIFS is 10 μs, and the propagation delay is 1 μs. The length of the MAC header and ACK packet is 36 and 14 bytes, respectively. Unless otherwise specified, a constant frame payload size of 1028 bytes, which includes 1000 bytes application data payload, 20 bytes IP header, and 8 bytes UDP header, were used in the simulations. The full queue traffic model was assumed to apply to all stations. The maximum backoff stage M_H and M_L were both set to be equal to 5 throughout this section. Unless otherwise specified, the numerical results were depicted in solid and dash lines and the simulation results were depicted with hollow and full symbols.

The accuracy of the analysis is verified by simulation results. In the following examples, we fix the sum of N_H and N_L to be 10 and set WL_0 to be 32. The effect of high- and low-class STAs (i.e. N_H and N_L, respectively) for different values of WH_0 was investigated. In Fig.4, the transmission probabilities of the two classes (i.e. τ_H and τ_L) were depicted for different N_H and WH_0. It was found that the larger N_H would lead to smaller transmission probabilities for both classes because the high-class STAs have more chances to access the channel. It was observed that, if WL_0 was fixed, a small WH_0 resulted in a high τ_H but a low τ_L. Fig. 5 showed the weighted fairness, ST_H/ST_L, for different cases. It was showed that ST_H/ST_L was highly depended on the selection of (WH_0, WL_0). The smaller WH_0 would lead to the larger ST_H/ST_L when WL_0 was fixed. Moreover, the growth of N_H might lead to the decrease of ST_H/ST_L. Moreover, it was found that, the ST_H/ST_L could be greater than one even if WH_0=WL_0. The differentiation of throughput was achieved by setting different AIFSs.

Based on the results observed above, we might choose a suitable WL_0 based the given N_H, N_L, M_H, and M_L. Then we may adopt any numerical method to find the WH_0 based on the desired weighted goal ϕ_H/ϕ_L. In the simulations, WH_0 was initially set to be equal to WL_0. For a given WH_0, ST_H/ST_L was calculated. The new value of WH_0 was decreased if $ST_H/ST_L < \phi_H/\phi_L$ and was increased otherwise. The iteration continued until $|ST_H/ST_L-\phi_H/\phi_L|$ was minimized.

In the following, we show that the fairness is achieved using the proposed method. A total of 10 STAs are considered. The weighted goal ϕ_H/ϕ_L is 2 and the WL_0 is fixed to be 32. WH_0 is then chosen based on the weighted goal for different combination of (N_H, N_L). Fig. 6 showed that the weight goal was achieved for different combination of high- and low-class STAs. Due to the constraint that WH_0 was an integer, therefore, it resulted in a little fluctuation of ST_H/ST_L.

V. Concluding Remarks

In this paper, we proposed an analytical method to obtain parameters required to achieve weighted fairness for services operating under the enhanced distributed coordinator function (EDCF) mode. A system with full queue traffic model and supported two classes of services was considered. Specifically, the length of AIFS was set to be DIFS and PIFS for low- and high-class stations, respectively, to backward compatible with 802.11. In the queueing analysis, a discrete-time Markov-chain was adopted to model the behavior of backoff counters for the two classes and the steady-state probabilities were derived. We further explored the relationship between throughput, conditional collision probability, and conditional busy medium probability for the two classes. With the information, the size of the contention window was adjusted to achieve the weighted goal. The accuracy of the analytical solution is verified by simulation for different number of active stations. It can conclude that, for different combination of high- and low-class STAs, the weighted fairness is easily achieved by employing the proposed method.

REFERENCES

[1] G. Bianchi, L. Fratta, and M. Oliveri, “Performance Evaluation and Enhancement of the CSMA/CA MAC Protocol for 802.11 Wireless LANs,” IEEE PIMRC, Taipei, Taiwan, Oct. 1996, pp. 407-411.

[2] F.Cali, M.Conti, and E.Gregori, “Dynamic Tunig of the IEEE 802.11 Protocol to Achieve a Theoretical Throughput Limit,” IEEE/ACM Trans. on Networking, vol. 8, no. 6, Dec. 2000.

[3] H. Wu, S. Cheng, Y. Peng, K. Long, and J. Ma, “IEEE 802.11 Distributed Coordination Function (DCF) Analysis and Enhancement,” IEEE ICC, vol. 1,pp. 605-609, 2002.

[4] N. H. Vaidya, P. Bahl, and S. Gupta, “Distributed Fair Scheduling in Wireless LAN,” IEEE MOBICOM, Boston, MA, Aug. 2000, pp. 167-178.

[5] A. Banchs and X. Perez, ”Distributed Weighted Fair Queuing in 802.11 Wireless LAN,” IEEE ICC, vol. 5 , pp. 3121 -3127, 2002.

[6] D. Qiao and K.G. Shin, ”Achieving Efficient Channel Utilization and Weighted Fairness for Data Communications in IEEE 802.11 WLAN under the DCF ” Tenth IEEE International Workshop on Quality of Service, pp. 227 -236, 2002.

[7] 3GPP TSG-RAN-1, Nortel Networks, “Nortel Networks’ reference simulation methodology for the performance evaluation of OFDM/WCDMA in UTRAN”, Document R1-03-0518, Meeting #32, Paris, France, May 19-23, 2003.

[8] G. Bianchi, “Performance Analysis of the IEEE 802.11 Distributed Coordination Function,” IEEE Journal on Selected Area in Communications, vol.18, no.3, pp. 535-547, Mar. 2000.

[9] “IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Higher-Speed Physical Layer Extension in the 2.4 GHz Band,” IEEE Std 802.11b-1999, Sep. 1999.

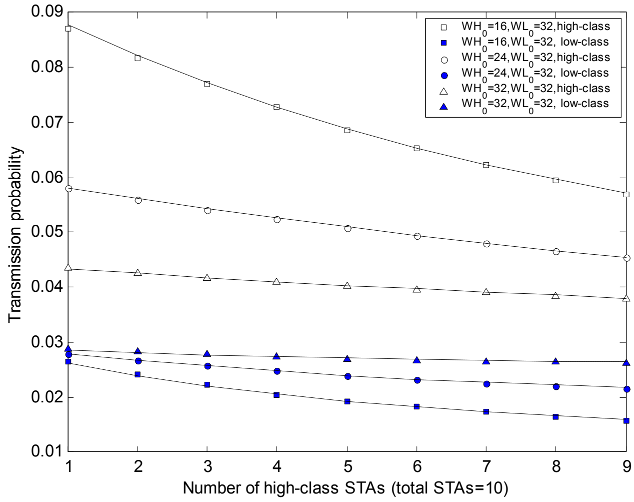


Figure 4. Transmission probability: analytical versus simulation results.

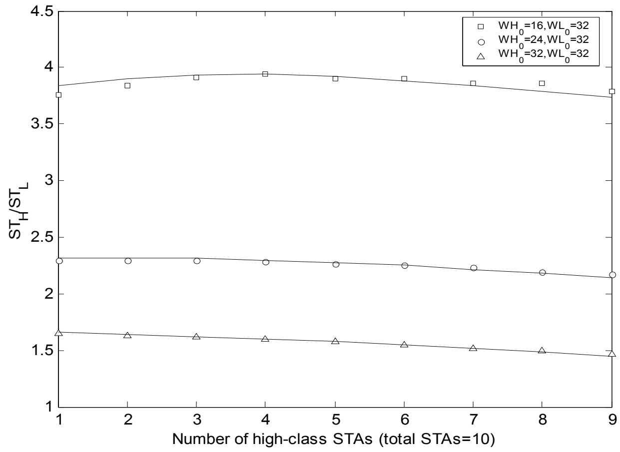


Figure 5. The ST_H/ST_L: analytical versus simulation results.

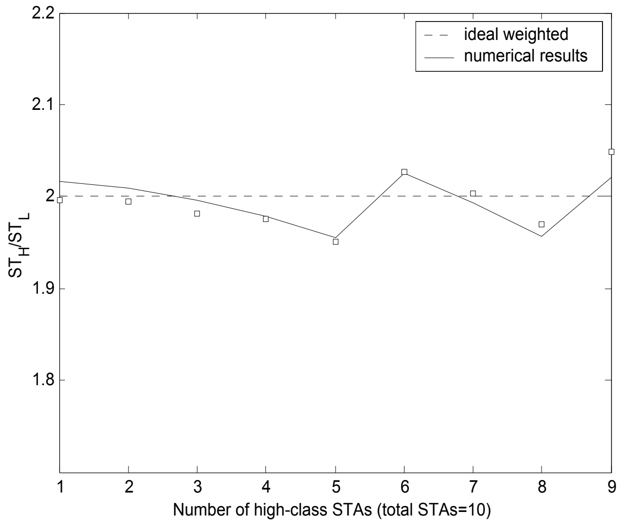


Figure 6. Performance of ST_H/ST_L under a total 10 STAs: ϕ_H/ϕ_L=2

A Joint Power and Rate Assignment Algorithm for Multirate Soft Handoff in WCDMA Heterogeneous Cellular Systems

Ching-Yu Liao, Chung-Ju Chang, and Yi-Shen Chen
Department of Communication Engineering
National Chiao Tung University
Hsinchu 300, Taiwan ROC
E-mail: cjchang@cc.nctu.edu.tw
Tel. No.: 886-3-5731923 Fax No.: 886-3-5710116

Department of Communication Engineering,

National Chiao Tung University, Taiwan, ROC.

Email: cjchang@cc.nctu.edu.tw

Tel: +886-3-5731923 Fax: +886-3-5710116

Abstract

The paper proposes a joint power and rate assignment (JPRA) algorithm to deal with multirate soft handoff in WCDMA heterogeneous cellular systems. This JPRA algorithm, containing a constrained unequal power allocation (CUPA) scheme and an evolutionary computing rate assignment (ECRA) method, can determine an appropriate allocation of power and service rate for multirate soft handoffs, respectively. It can achieve power balancing between cells for soft handoffs better than the conventional site-selection diversity transmission (SSDT) scheme. Simulation results show that the JPRA algorithm can improve the forced termination probability of soft handoffs by 61.0 %, and the total throughput by 2.4 %. Besides, considering measurement error happened during the active set selection, the JPRA algorithm can further improve the forced termination probability of soft handoffs by 76.8 %, and the total throughput by 6.7 %.

Keywords

multirate soft handoff, heterogeneous cellular system, power allocation, rate assignment.

Part of this work is submitted to *IEEE ICC'04*. The journal version is submitted to *IEEE Trans. on Mobile Computing*.

I. Introduction

Soft handoff is one of the most important features in WCDMA cellular mobile communication systems. When mobile users move from one cell to another cell, the soft handoff technique can provide seamless connections and better signal qualities for users in the cell boundary. However, in the forward link, systems often have to consume more power to serve soft handoff users than that to serve non-handoff users. Since the power resource are shared between non-handoff users and soft handoff users, the radio resource management would be a critical problem especially when soft handoff users are with multirate services. Generally, a multirate soft handoff costs much more power resource to satisfy the required quality of service.

Furthermore, in heterogeneous WCDMA cellular systems, microcells, which are with stringent power budget, may easily exhaust their total transmission power because of serving soft handoff users in the forward link [1], [2], [3], and then there is no extra power resource for serving other users in the system. This situation would become worse for multirate WCDMA heterogeneous cellular systems. If power balancing can be achieved between macrocells and microcells through effectively managing radio resources for multirate soft handoffs, there are more power resource can be allocated for users in the congested microcells. Consequently, the system performance can be improved.

In this paper, we propose a joint power and rate assignment (JPRA) algorithm for multirate soft handoff users in WCDMA heterogeneous cellular systems. Many literature discussed the issue of joint power and rate assignment for all users in the cellular system in the sense of global optimization problem [4], [5]. However, they focused on the reverse link and did not concern about multirate soft handoffs. Reference [6] discussed radio resource management in multiple-chip-rate direct sequence CDMA systems supporting multiclass services, and it developed call admission control to arrange handoff in the same subsystem or execute inter-frequency handoff. Kim [7] dealt with rate-regulated power control in the reverse link without concerning handoff. Kim and Sung [8] proposed a handoff management scheme for multirate services using guard channels and reservation on demand queue control. But a hard handoff scheme was considered. Reference [9] and [10] proposed joint power and rate allocation algorithms in the downlink CDMA systems, however no handoff mechanism was concerned.

The proposed JPRA algorithm is composed of a two-phase process. In the first phase, a constrained unequal power allocation (CUPA) scheme is designed for soft handoffs. A conventional site selection diversity transmission (SSDT) scheme was proposed for soft handoff in [11]. It dynamically selects one base station with best link quality in the active set to serve the soft handoff. The SSDT sometimes cannot offer enough power required for multirate soft handoff users because of the maximum allocation power constraint. Furthermore, since SSDT is a single site transmission scheme at one time, it easily consumes more power when suffering measurement error during active set selection. The advantage of power saving feature would disappear. The proposed CUPA is a multi-site diversity transmission scheme. It distributes the required allocation power for one soft handoff user unequally among all the active base stations, based on link quality between the active base station and the soft handoff user. The base station with better link quality will allocate more power than others with worse link qualities. In addition, the allocation power is constrained by the maximum allocation power of the base station. In the second phase of JPRA algorithm, an evolutionary computing rate assignment (ECRA) method is proposed to formulate an integer and discrete optimization problem under a predefined total power constraint for soft handoffs in each cell. It is well known that conventional optimization methods can hardly cope with problems with integer and discrete variables, whereas evolutionary computing methods are very efficient for these problems for reducing the searching complexity [12].

In the meantime, a multirate removal (MRV) algorithm is proposed to pick out a user who consumes system resource most and to reduce its service rate or even block it when the system resource is insufficient. Several removal algorithms had been proposed. Among these, the link-based and received signal-strength based removal algorithms were only suitable for single service [13], [14]. The prioritized removal algorithm, based on predefined service priority, did not consider service rate tuning for users in the reverse link of a multiservice cellular system [15]. On the other hand, a new multi-quality balancing power allocation (MQBPA) algorithm for non-handoff users with multiple service rates is also developed. Previous work for quality balancing power allocation technique were studied only for a single service rate and the same required signal quality [1], [3], [16]. Consequently, system operation for forward-link power and rate assignment containing JPRA for soft handoff is shown in Figure 1.

Simulation results show that proposed JPRA algorithm effectively achieves power balancing between macrocells and microcells in WCDMA heterogeneous cellular systems. Because of the power balancing characteristic, JPRA can improve not only the forced termination probability of soft handoffs by 61.0 % but also the total throughput by 2.4 %, as compared to the conventional SSDT scheme. Moreover, with the concern of 1.5 dB measurement error happened during the active set selection, it further improves the forced termination probability of soft handoffs by 76.8 %, and the total throughput by 6.7 %. The proposed JPRA algorithm can achieve better power balancing between cells and owns less sensitivity to the measurement error during the active set selection in multirate WCDMA heterogeneous cellular systems.

The remaining parts of this paper are organized as follows. In section II, the JPRA algorithm for multirate soft handoff is proposed. Simulation results are presented and discussed in section III.

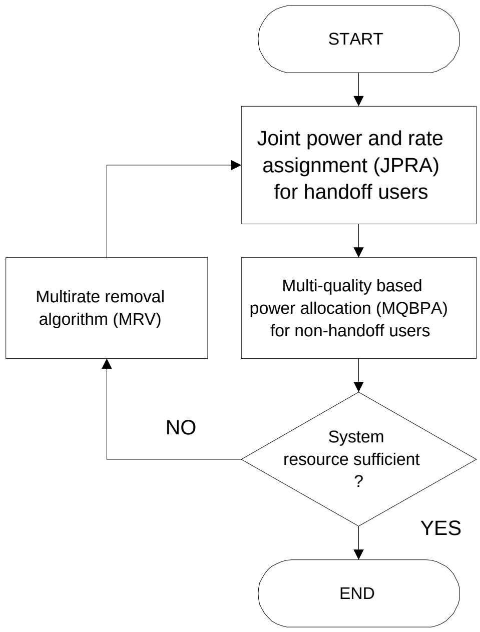


Fig. 1. The system operation of forward-link power and rate assignment

Finally, section VI concludes this paper. Moreover, the proof of CUPA convergence is provided in the Appendix A; the MQBPA and MRV algorithms are provided in the Appendix B and C, respectively.

II. The JPRA Algorithm

The JPRA algorithm is mainly composed of the constrained unequal power allocation (CUPA) scheme in the first phase and the evolutionary computing rate assignment (ECRA) method in the second phase. Before describing them, the received bit-energy-to-noise ratios of multirate soft handoff is defined.

In WCDMA cellular systems, the received bit-energy-to-noise ratio (E_b/N_o) of user j in base station i and with service rate r, denoted by $\gamma_{i,j}(r)$, must be larger than or equal to the required signal quality, denoted by $\gamma^*(r)$. For bandwidth W, the $\gamma_{i,j}(r)$ can be expressed as

$$\gamma_{i,\,j}(r) = \frac{q_{i,j}(r) \cdot L_{i,j} \cdot G(r)}{\sum\limits_{k} P_k \cdot L_{k,j} + \eta_o} \geq \gamma^*(r)\,, \tag{1}$$

where $q_{i,j}(r)$ is the allocation power from base station i to user j; $P_k = \sum\limits_{j=1}^{N} q_{k,j}$ is the total downlink transmission power for N users in cell k; $L_{i,j}$ is the link quality from cell i to user j; $G(r) = W/r$ is the processing gain; and η_o is background noise. Assume that the link quality includes only the effect of both path loss and shadowing. For a soft handoff user h with service rate r, using the maximum ratio combining (MRC) method to combine signals from all serving base stations in the active set D_h, its received E_b/N_o, $\gamma_h(r)$, can be obtained by

$$\gamma_h(r) = \sum_{i \in D_h} \gamma_{i,h}(r). \tag{2}$$

A. The CUPA Scheme

The constrained unequal power allocation (CUPA) scheme estimates the required allocation power for soft handoff user h, $q_h(r)$; then it distributes $q_h(r)$ to all serving base stations in D_h under the constraint of maximum allocation power to each user by base station $i \in D_h$, $\widehat{q}_i$. The $q_{i,h}(r)$ is proportional to the link quality between the serving base station i and the soft handoff user h [3]. If the allocation power of one link reaches to the constraint of maximum allocation power, the CUPA will compensate the required power through other links. The CUPA scheme is an iterative method to distribute $q_h(r)$ to all serving base stations so that the required signal quality can be satisfied. The design is to try to accomplish power balancing between cells in the heterogenous cellular system with different cell sizes. Besides, it is noteworthy that because of the maximum allocation power constraint of each forward link, there exists a forced termination situation for the soft handoff because the soft handoff user cannot obtain required signal quality even though all active links are allocated with maximum power. If the soft handoff is forced to terminate, the $q_{i,h}(r)$ of each link i in the active set D_h are reset to zero. The CUPA scheme is stated in more details in the following.

[**The CUPA Scheme**]

Step 0: [Exam soft handoff feasibility]

- Allocate maximum power $\widehat{q}_i$ for each active links i .
- Calculate received signal quality $\gamma_h(r)$ based on (1) and (2).
- IF $\gamma_h(r) > \gamma^*(r)$, THEN Goto **Step 1**.

ELSEIF $\gamma_h(r) = \gamma^*(r)$, THEN Set $q_{i,h}(r) = \widehat{q}_i$, $i \in D_h$, DONE.

ELSE the soft handoff user h is forced to terminate ($q_{i,h}(r) = 0$, $i \in D_h$), DONE.

Step 1: [Initialize]

- Initialize the required transmission power $q_h^n(r)$, $n = 0$, for soft handoff user h to be the summation of the maximum allocation power, $\widehat{q}_i$, of each serving base station i by

$$q_h^0(r) = \sum_{i \in D_h} \widehat{q}_i. \tag{3}$$

Step 2: [Calculate weighting factor]

• Set the weighting factor of the allocation power from base station i in D_h, based on the link quality between the cell i and the soft handoff user h, by

$$w_{i,h} = \frac{L_{i,h}}{\sum\limits_{i \in D_h} L_{i,h}}. \tag{4}$$

Step 3: [Calculate allocation power]

• Determine the power that base station i in D_h allocates to the soft handoff user h at iteration time n, $q_{i,h}(r)$, by

$$q_{i,h}^n(r) = \mathrm{Min}\{\ q_h^n(r) \times w_{i,h},\ \widehat{q}_i\ \},\ \forall\, i \in D_h. \tag{5}$$

Step 4: [Compute received E_b/N_o and tuning factor]

• Compute the corresponding $\gamma_h^n(r)$ in (2) at iteration time n, and set tuning factor ρ_h^n by

$$\rho_h^n = \frac{\gamma_h^*(r)}{\gamma_h^n(r)}. \tag{6}$$

Note that $\gamma_h^*(r)$ is the required signal quality received at the soft handoff user h.

Step 5: [Check Stop Criterion]

• IF $\rho_h^n \neq 1.0$, THEN

  – Let

$$q_h^{n+1}(r) = \rho_h^n \times q_h^n(r) \tag{7}$$

  – Set $n = n + 1$ and Goto **Step 3**.

ELSE DONE. ■

The CUPA scheme is convergent, that is proven in the Appendix A. The required allocation power of each active link, $q_{i,h}(r)$, $i \in D_h$, for all soft handoff users with all kinds of service rates can be obtained through the CUPA scheme.

B. The ECRA Method

The ECRA method performs the rate assignment for multirate soft handoff users. It formulates the rate assignment issue as a constraint optimization problem with an objective to maximize the total throughput of multirate soft handoffs. Note that the total power budget for a cell is limited, and the total power allocated to soft handoffs in cell i would be constrained by a maximum, denoted by $\widehat{Q}_i$, and $\widehat{Q}_i < P_i$. Since the computation time may be far behind system's requirements especially for a larger number of multirate soft handoff users being managed, In this paper an evolutionary computing [12], which is a promising intelligent technique to find global optimal solution effectively, is adopted.

Assume there are N_d and N_v soft handoff users with data and voice services, respectively, and N_b cells in the system. Assume there are m kinds of data service rates, and the searching complexity is $(m+1)^{N_d}$ by using exhaustive method, in which 1 means zero service rate for suspending transmission. For example, if N_d is 10 and m is 4, there are nearly 10^7 searching complexity. This is far behind the requirement of computation time. In order to reduce complexity of exhaustive search, the evolutionary computing technique [12] is applied and depicted in the following. Represent the service rate of each user as a chromosome in a population. For m kinds of data service rates, each rate is encoded into $\lfloor \log_2(m+1) \rfloor$ binary digits, denoted by x, and the decoder function of the service rate is $s(x)$. For soft handoff data user h, its data service rate is thus $s(x_h)$, and its corresponding allocation power is $q_h(s(x_h))$, in which the allocation power from active link i is $q_{i,h}(s(x_h))$ by the CUPA scheme. Denote r_v to be the service rate for soft handoff voice users and $q_{i,h}(r_v)$ is the corresponding allocation power from active link i to soft handoff user h. The ECRA method is to find an optimal rate assignment vector (decision vector) of N_d soft handoffs, $\mathbf{x}^* = [x_1^*, x_2^*, \cdots, x_{N_d}^*]$, for maximizing the objective function $O(\mathbf{x})$, which is defined to be the total throughput of soft handoff data users, given by

$$O(\mathbf{x}) = \max \left\{ \sum_{h=1}^{N_d} s(x_h) \right\}, \tag{8}$$

subject to constraints:

$$\sum_{h=1}^{N_v} q_{i,h}(r_v) + \sum_{h=1}^{N_d} q_{i,h}(x_h) \le \widehat{Q}_i,\ 1 \ \le \ i \le \ N_b, \tag{9}$$

and

$$\gamma_h(s(x_h)) \ge \ \gamma^*(s(x_h)) \ ,\ \forall\, h. \tag{10}$$

Because of these constraints, some decision vectors may be out of the feasible domain. A violation function, which is proportional to the square of violation, is used to rank violated constraints of solution [12]. The values of the constraint violation function indicate how far the solutions deviate from the feasible region. This constraint violation function is defined as

$$\Psi(\mathbf{x}) = \frac{\sum_{h=1}^{N_d} H_h[\gamma_h(s(x_h))]^2}{2N_d} + \frac{\sum_{i=1}^{N_b} M_i[Q_i^d(\mathbf{x})]^2}{2N_b}, \tag{11}$$

where H_i, and M_i are the Heaveside operators [12], in which $H_i(\cdot) = 1$ $[M_i(\cdot) = 1]$ whenever the constraint in (10) and (9) is violated, and $H_i(\cdot) = 0$ $[M_i(\cdot) = 0]$ otherwise. The evolutionary computing is a more advanced genetic algorithm, which uses stochastic searches through simulating natural genetic processes of living organisms, including selection, mutation, and crossover, to solve difficult optimization problem in real-world. Based on the formulation of constraint optimization problem, the optimal decision vector, $\mathbf{x}^*$, can be found for maximizing the objective function, O($\mathbf{x}$). The ECRA method is described in the following. Noticeably, the allocation powers for the soft handoffs are corresponding to the ones obtained by the CUPA scheme.

[The ECRA method]

Step 1: [Initialize]

• Set the crossover rate p_c, the mutation rate p_u, and the maximum number of generations T.

• Initialize the generation $t = 1$, the optimal objective value $O^* = 0$, and the optimal decision vector $\mathbf{x}^*$ to be a zero pattern.

• Generate K populations that are randomly selected decision vector $\mathbf{x}_k = [x_1^k, \cdots, x_{N_d}^k]$, $1 \leq k \leq K$.

Step 2: [Constraint tournament selection]

• Choose K tournament pairs randomly among all populations.

• Calculate violation function (11) for each competitive pair, and determine one winner, which owns a smaller value of violation function.

• Replace each population $\mathbf{x}_k$ with the winner population of each competitive pair, thus form K new populations.

Step 3: [Variable point crossover]

• Choose $K/2$ crossover pairs from adjacent population $\mathbf{x}_k$ and $\mathbf{x}_{k+1}$, where k is odd.

• Generate a random number c in $[\,0, 1\,]$ for each chromosome in each crossover pair.

• For the chromosome with $c < p_c$, generate crossover point randomly in $[1, \lfloor \log_2(m+1) \rfloor]$, and make crossover operation within this crossover chromosome.

Step 4: [Uniform mutation]

• Generate a random number u in $[0, 1]$ for every bit in each population, and mutate the bits whenever $u < p_u$.

Step 5: [Calculate objective function of resulting new population]

• Calculate violation function value for each population.

• Find feasible population $\{\, \mathbf{x}_f \,\}$ with zero violation among K populations.

• IF $\{\, \mathbf{x}_f \,\}$ is not empty set, THEN Calculate the objective function value $\{\, O(\mathbf{x}_f) \,\}$.

 – IF $\max \{\, O(\mathbf{x}_f) \,\} > O^*$, THEN

  Set $O^* = \max\{O(\mathbf{x}_f)\}$ and the optimal decision vector $\mathbf{x}^* = \arg\max\limits_{\mathbf{x}_f}\{\, O(\mathbf{x}_f) \,\}$.

 ELSE Goto **Step 6**.

ELSE Goto **Step 6**.

Step 6: [Check stop criterion]

• IF $t < T$, THEN Set $t = t + 1$, and Goto **Step 2**.
ELSE DONE. ■

III. Simulation Results and Discussion

A. Simulation Model

Consider a heterogeneous cellular system with 9 wrap-around squared cells, including 4 microcells in the central and 8 macrocells in the neighboring cells as shown in Figure 2. The radii of macrocell and microcell are 1 km (R_M) and 0.5 km (R_μ), respectively, thus the cell radius ratio (φ) between microcell and macrocell, R_μ/R_M, is 0.5. The antenna heights of macrocell and microcell are 20 meters and 10 meters, respectively, and the antenna height of mobile stations is 1.5 meters. For the propagation channel model, only path loss and long-term shadowing are taken into account, in which two slope path loss exponents are 2 dB and 4 dB, and standard deviations of two slope shadowing are 4 dB and 8 dB [17]. For the power budget design, the maximum transmission power, $\widehat{P}_i$, for macrocell (microcell) i is 20 (10) watt; and the maximum allocation power for each user in macrocell (microcell) is 1 (0.5) watt. Each user determines its active set members based on received signal strength by soft handoff algorithm which is based on the difference of received signal strength between users and cells, $\widetilde{P}_i L_{i,j} - \widetilde{P}_k L_{k,j} < \eta, i \neq k$, where η is soft handoff threshold and $\widetilde{P}_i$ is the transmission power of pilot signal of base station i, which is related to the cell size. Here, η is 2 dB and the maximum active set size is 3. In the simulations, two cases without and with measurement error during the active set selection are concerned. For the measurement error case, the received signal strength of user is added one Gaussian distributed random variable with zero mean and 1.5 dB standard deviation.

Assume there are 40 users uniformly distributed in each cell, and each user moves in a constant speed 36 km/hr. The probability of moving direction change for users is 0.2 and the direction update is among ±45 degree [18]. During the mobility, the correlated shadowing effect is based on Gudmundson model [19], in which the normalized autocorrelation function $A(d)$ between two correlated points with distance d can be described accurately by an exponential function:

$$A(d) = \exp\left\{ -\frac{|\,d\,|}{d_{corr}} \ln 2 \right\}, \tag{12}$$

where d_{corr} is the decorrelation length equal to 20 meters in a vehicular environment. Figure 2 shows an example of mobility trajectory. Assume the shadowing factor will not be changed when the moving distance is less than 4 meters and there are five averaging windows in each snapshot, for 36km/hr mobility speed, the correlated shadowing duration is 400 msec. Assume the power allocation duration

TABLE I

Multirate Service

Service	r (kbps)	$\gamma^*(r)$(dB)	encoded x
Voice	12.2	5	
Data	0	N/A	(100)
Data	16	4	(001), (101)
Data	32	3	(011), (111)
Data	64	2	(010), (110)
Data	144	1.5	(000)

is equal to one frame time (10 msec), the allowable iteration is about 40 times. These performance measurements are averaging from 2000 independent instances of user location and shadowing, and each snapshot has 5 times correlated instances. In the simulations, two essential performance measures are investigated. One is the forced termination probability of soft handoffs, which indicates the service continuity for soft handoffs. It is evaluated by counting the proportion of soft handoff users that are terminated by the system due to insufficient power resource for soft handoffs, temporarily. The other is the total system throughput, which is obtained by summing all service rates of users in the system.

For parameters in ECRA method, the total power constraint for soft handoffs, $\widehat{Q}_i$, is assumed to be 0.3 times the total transmission power of each cell i [20]. The population size (K) is 100, the crossover rate (p_c) is 0.5, the mutation rate (p_m) is 0.05, and the stop generation (T) is 20. The computing time of ECRA method is far behind that of the exhaustive searching method. The ECRA method will search whole 2000 searching patterns to find optimal rate sets for multirate soft handoffs in the very first iteration. Since the system gradually approaches to convergent point, the optimal rate sets only need less searching patterns afterward.

B. Discussion

We compare the proposed JPRA algorithm with the site-selection diversity transmission (SSDT) algorithm and the multiple site-selection diversity transmission (MSSDT) algorithm. The major idea of site selective transmit diversity (SSDT) [11] is to dynamically choose one base station with the best link quality in the active set for transmission in order to mitigate interference caused by multiple site transmission. The SSDT sometimes cannot offer enough required power for soft handoff users because of the maximum allocation power constraint of each link. The MSSDT is designed to compensate the required power for users with allocated rates so as to satisfy their required signal quality, and then it sorts the active links by link quality and uses better links to transmit signals in order. Since MSSDT uses diversity transmission for soft handoff users when necessary, the ECRA method is applied in MSSDT for rate assignment in the simulations.

Figures 3(a) and 3(b) show the average forced termination probability of soft handoffs without and with 1.5 dB measurement error, respectively. It can be seen from Figure 3(a) that JPRA and MSSDT outperform SSDT by 61.0 % and 45.8 %, respectively, when there are 2 data users in each cell. The main reason is that SSDT easily uses up the maximum allocation power for *the best link* in microcells, as compared to JPRA and MSSDT, so that SSDT cannot support enough required allocation power often for these multirate soft handoff users in microcells. Notice that the power budget (maximum allocation power) for microcells to serve multirate soft handoffs is stringent and critical in the heterogeneous cellular system. Besides, because of the signal quality constraint in ECRA method, only if soft handoff users can satisfy required signal quality of the allocated service rates, the ECRA method helps to serve soft handoff users as many as possible. The signal quality constraint results in a reduction of forced termination probability of soft handoffs. However, the performance improvement is decreased as soon as the number of data users is increased because high interference is induced accordingly. In the case of 1.5 dB measurement error, it can be found from Fig. 3(b) that JPRA further improves the forced termination probability of soft handoffs over SSDT and MSSDT by 76.8 % and 26.8 %, respectively, at the case of 2 data users in each cell. The results imply that if the measurement error causes a wrong selection of best link in the active set, the damage of power waste is more serious in SSDT and MSSDT than in JPRA. It is because the CUPA scheme in JPRA *distributes* the required allocation power to *all* the cells in the active set in proportional to their link qualities. Thus, because of the power balancing feature, JPRA can not only provide better service continuity performance but also own a capability of resistance to measurement error better than SSDT and MSSDT.

Figures 4(a) and 4(b) present the total throughput for cases without and with 1.5 dB measurement error, respectively. It can be found that JPRA and MSSDT can enhance the average total throughput than SSDT by 2.4 % in the case of error free, and they can support higher average total throughput than SSDT by 6.7 % and 4.2 %, respectively, in the case of 1.5 dB measurement error. The superiority of JPRA over the SSDT and MSSDT is mainly because the CUPA scheme in JPRA owns the aforementioned power balancing feature; in addition, the total power constraint of the ECRA method in JPRA prevents wasting too much power resources for multirate soft handoffs so that the power resource can be preserved to serve non-handoff users with higher service rates.

Figure 5 (a) and (b) show the average total transmission power ratio for macrocells and microcells, respectively, where the ratio of cell i is defined as its total transmission power relative to the maximum transmission power. It is found that SSDT and MSSDT allocate more power for the best link of microcells than that of macrocells, whereas JPRA allocates nearly equal ratio of total transmission power to macrocells and microcells. Thus, JPRA achieves power balancing better than SSDT and MSSDT. This is because JPRA distributes power between cells by applying the CUPA algorithm which is to let all base stations in the active set cooperatively allocate power, proportional to their link qualities, to serve soft handoff users. As regards the case of measurement error, it may result in a wrong selection of the best link in the active set, and base stations have to allocate more power to serve soft handoffs. As a result, it incurs higher ratio of total transmission power than the case without measurement error. In order to demonstrate power balancing performance, we further define a power balancing index PBI as

$$PBI = E\left[\frac{\sum_{i=1}^{N_b}(\mathbf{b}_i - \overline{\mathbf{b}})^{\mathbf{2}}}{N_b - 1}\right], \tag{13}$$

where $\mathbf{b}_i$ is total transmission power ratio for cell i, $P_i/\widehat{P}_i$; $\overline{\mathbf{b}}$ is the average total transmission power ratio for all N_b cells, $(\sum_{i=1}^{N_b}\mathbf{b}_i)/N_b$. The PBI represents the variance of $\mathbf{b}_i$ which means the lower value of PBI, the better power balancing between cells. Figure 6 shows the PBI results with and without measurement error cases. It can be found that JPRA accomplishes excellent power balancing performance for both with and without measurement error cases, which justify the superiority of JPRA algorithm for previous results in terms of forced termination probability of soft handoff and total throughput.

Noticeably, if the cell radio ratio between macrocell and microcell turns out to be smaller or the measurement error becomes worse in heterogeneous cellular systems, the JPRA algorithm could enhance the system performance more. On the other hand, we also investigate all the algorithms in the homogeneous cellular systems ($\varphi = 1$). Because the power budgets of maximum allocation power constraint and maximum transmission power are the same for all cells in the homogeneous cellular systems, the power balancing advantages of JPRA would not be so significant between mixed-size cells in the heterogeneous cellular systems. On the contrary, the power saving characteristic is more important than the power balancing for the homogeneous cellular systems, but SSDT and MSSDT have weak resistance capability of measurement error which may degrade the advantage of power saving characteristic.

IV. Conclusion

In this paper, a joint power and rate assignment (JPRA) algorithm is proposed to deal with multirate soft handoffs in WCDMA heterogeneous cellular systems. JPRA allocates transmission power based on the constrained unequal power allocation (CUPA) scheme and achieves optimal service rate according to the evoluationary computing rate assignment (ECRA) method, for the multirate soft handoffs. In order to support good service quality for multirate soft handoffs, a proposed multirate removal algorithm (MRV) is activated to reduce service rate or even block users whenever system radio resource is insufficient. In the meantime, a new multi-quality balancing power allocation (MQBPA) scheme is also developed to allocate power for non-handoffs users with multirate services.

Simulation results show that, JPRA accomplishes excellent power balancing feature between cells, and it thus can improve the forced termination probability of soft handoffs by 61.0 %, and the total throughput by 2.4 % , as compared to conventional SSDT. Moreover, JPRA is less sensitive to the measurement error than SSDT in the active set selection; so the former can further enhance the forced termination probability of soft handoffs and the total throughput by 76.8 % and 6.7 % over the latter, respectively. The above advantages of JPRA are more conspicuous in WCDMA heterogeneous cellular systems with smaller cell radius ratio between microcell and macrocells.

Also, the computational complexity is worthwhile to take into account. Assume there are 10 multirate soft handoff users in the system. The ECRA method and the exhaustive searching method need to take 2.0×10^3 and 4.8×10^7 searching patterns, respectively, for finding optimal rate set of multirate soft handoffs. According to the parameters set in ECRA method, it takes 20 float-point additions, 50 compare operations, and 2450 bit operations per generation. By using the INTEL Pentium 4 processor with speed 2.0 GHz and MATLAB codes, the average computing time of ECRA method is about 9.5 msec. The rate assignment procedure can be completed among one frame time. If special purpose CPU or DSP processors with pipeline architecture or optimized computation capabilities are adopted, less computing time should be taken to execute rate assignment. Therefore, the computational complexity is reduced significantly, and the practical implementation would be feasible in WCDMA heterogeneous cellular systems with multirate services.

Appendix A

Convergence Proof of the CUPA Scheme

[**Definition**]: A function F is "**standard**" if it satisfies the following conditions for all non-negative power vectors [22]:

- Positivity : $F(\,y\,) \;>\; 0$,

- Monotonicity : $y_1 \geq y_2 \Rightarrow F(y_1) \geq F(y_2)$,
- Scalability : $\forall \alpha > 1, \alpha F(y) \geq F(\alpha y)$. ∎

[**Proposition**]: A "**standard**" power control algorithm will converge to a unique "**effective**" power vector that achieves γ_h^* for any initial power q_h. The standard power control algorithm means that the power allocation function is standard. ∎

The CUPA scheme has an iterative process for the required power allocation for the multirate soft handoff user h, q_h, which is described by

$$q_h^{n+1} = I(q_h^n), \tag{14}$$

where the superscript of q_h^n denotes the number of iteration and I denotes the power allocation function.The notation of service rate r is ignored here for convenience. From (2), (6), and (7), I is given by

$$I(q_h^n) = \frac{\gamma_h^*}{\sum\limits_{i \in D_h} \gamma_{i,h}} \times q_h^n. \tag{15}$$

Based on (1), $\gamma_{i,h}$ is the function of $q_{i,h}$, $i \in D_h$, which should obey the constraint of maximum allocation power given in (5). In order to represent these relationship, we denote $\gamma_{i,h}$ as $\Gamma(q_{i,h})$, instead of (1).

Since all the link gains and background noise between soft handover mobile station h and serving base stations i, $i \in D_h$, are positive, the power allocation function given in (14) has the positivity and monotonicity properties. As for the scalability property, there are two kinds of cases in the resulting allocation power vector $q_{i,h}$, $i \in D_h$, considering the effect of maximum allocation power constraint:

Case 1: $q_{i,h} = \min(q_h w_{i,h}, \widehat{q}_{i,h}) < \widehat{q}_i, \forall\, i$.

Since $\alpha q_h w_{i,h} > q_h w_{i,h}$, for $\alpha > 1$,

it can be found that $\Gamma(\alpha q_h w_{i,h}) > \Gamma(q_h w_{i,h})$, and $\sum\limits_{i \in D_h} \Gamma(\alpha q_h w_{i,h}) > \sum\limits_{i \in D_h} \Gamma(q_h w_{i,h})$.

Thus,

$$I(\alpha q_h) = \frac{\gamma_h^*}{\sum\limits_{i \in D_h} \Gamma(\alpha q_h w_{i,h})} \times (\alpha q_h) < \alpha\, I(q_h).$$

Case 2: $\exists\ k,\ k \in D_h$ s.t. $q_{k,h} = \min(q_h w_{k,h}, \widehat{q}_{k,h}) = \widehat{q}_k$.

It can be found that $\sum\limits_{i \in D_h} \Gamma(q_h w_{i,h}) = \sum\limits_{\substack{i \neq k, \\ i \in D_h}} \Gamma(q_h w_{i,h}) + \sum\limits_{k} \Gamma(\hat{q}_k)$,

then $\sum\limits_{i \in D_h} \Gamma(\alpha q_h w_{i,h}) = \sum\limits_{\substack{i \neq k, \\ i \in D_h}} \Gamma(\alpha q_h w_{i,h}) + \sum\limits_{k} \Gamma(\widehat{q}_k) \geq \sum\limits_{i \in D_h} \Gamma(q_h w_{i,h})$.

Thus,

$$I(\alpha q_h) = \frac{\gamma_h^*}{\sum\limits_{i\in D_h} \Gamma(\alpha q_h w_{i,h})} \times (\alpha q_h) \ \leq \ \alpha\, I(q_h).$$

■

The interference function also possesses the scalability property. Therefore the proposed CUPA scheme is a standard power control algorithm, and it always exist an effective solution q_h for soft handoff user h.

Appendix B

The MQBPA Algorithm

The multi-quality balancing power allocation (MQBPA) algorithm is to provide each non-handoff user the required signal quality of itself. Assume each service rate r has the required signal quality $\gamma^*(r)$; denote C_i (Q_i) as the total transmission power for non-handoff (soft handoff) in cell i; and $C_i + Q_i = P_i$. The MQBPA algorithm assigns the non-handoff user j in cell i with service rate r an amount of power, $q_{i,j}(r)$, by

$$q_{i,j}(r) = \frac{w_{i,j}}{\sum\limits_{j\in \mathbf{U}_i} w_{i,j}} \cdot C_i, \tag{16}$$

where $\sum\limits_{j\in \mathbf{U}_i} q_{i,j}(r) = C_i$, $\mathbf{U}_i$ is the set of non-handoff user in cell i, and $w_{i,j}$ is defined as

$$w_{i,j} = \frac{\sum\limits_k P_k L_{i,k} + \eta_o}{L_{i,j} \cdot G(r)} \cdot \gamma^*(r). \tag{17}$$

Substituting (16) and (17) into (1), the received signal quality of the non-handoff user j in cell i can be yielded as

$$\gamma_{i,j}(r) = \frac{C_i}{\sum\limits_{j\in \mathbf{U}_i} w_{i,j}}\, \gamma^*(r). \tag{18}$$

From (18), it can be found that any user j with service rate r in cell i will receive the same signal quality, $\gamma_{i,j}(r)$, which depends on required signal quality of service rate r, $\gamma^*(r)$. Thus, all users in cell i will have the same referenced signal quality $\gamma_{i,j}(r)/\gamma^*(r)$ value, no matter what kind of service rate r is. Assume that γ_o is the target of balanced signal quality for the system, and the relative balanced signal quality, $\widetilde{\gamma}_i$, for users with any service rate in cell i will be

$$\widetilde{\gamma}_i = \frac{\gamma_{i,j}(r)}{\gamma^*(r)} \cdot \gamma_o. \tag{19}$$

The balancing target is to let $\widetilde{\gamma}_i \geq \gamma_o$ so that each user with service rate r will have $\gamma_{i,j}(r) \geq \gamma^*(r)$. If $\widetilde{\gamma}_i$ is not equal to γ_o, the total allocation power C_i should be adjusted by tuning factor ψ_i, which is given by

$$\psi_i = \frac{\gamma_0}{\widetilde{\gamma}_i}. \tag{20}$$

The MQBPA algorithm is described in the following.

[The MQBPA Algorithm]

Step 1: [Initialize]

- Initialize iteration time $n = 0$, and the total transmission power P_i^0 to the maximum total transmission power $\widehat{P}_i$ for each cell i.
- Calculate the total allocation power Q_i for handoff users in each cell i after executing JPRA algorithm.

Step 2: [Calculate $w_{i,j}$]

- Calculate $w_{i,j}$, based on (17), for user j in cell i.

Step 3: [Calculate allocation power]

- Calculate the total allocation power C_i^n for non-handoff users in each cell i, which is equal to $(P_i^n - Q_i)$.
- Calculate allocation power $q_{i,j}^n(r)$= min($q_{i,j}^n(r)$, $\widehat{q}_i$) for each non-handoff user j with service rate r in cell i based on (16).

Step 4: [Calculate balanced signal quality]

- Calculate the relative balanced signal quality $\widetilde{\gamma}_i^n$ for users in cell i based on (19).

Step 5: [Calculate tuning factor]

- Calculate tuning factor ψ_i^n for each cell i based on (20).

Step 6: [Check Stop Criterion for each cell i]

- IF any $\psi_i^n \neq 1.0$ and convergence is not met, THEN
  - Adjust total transmission power as $P_i^{n+1} = \min(\ \psi_i^n \times C_i^n + Q_i,\ \widehat{P}_i\ )$.
  - Set $n = n + 1$ and Goto **Step 2**.

ELSE DONE. ■

Since the balanced signal quality for each cell in (19) is independent of serving users' service rates, the proposed MQBPA algorithm is standard [1]. It will converge to a desired solution if there exists an effective individual power allocation for all users such that they can obtain their required signal qualities. If an effective solution does not exist, the issue becomes how to find a subset of users that can obtain their required signal qualities. Then, the MRV algorithm will be activated, which is described in the Appendix C.

Appendix C

The MRV Algorithm

The multirate removal (MRV) algorithm defines a novel removal index for user j with service rate r, denoted by $J_j(r)$, as

$$J_j(r) = \frac{\gamma^*(r)}{\widetilde{P}_i \cdot L_{i,j} \cdot G(r)}\,, \tag{21}$$

where $\widetilde{P}_i$ is the transmission power of pilot signal of base station i, which is related to the cell size. The removal index shows how much the system resource is required to serve user j. The worse the received signal strength, the higher the service rate and the required signal quality are, the larger the removal index value will be.

In order to provide higher priority for voice users, the proposed MRV algorithm removes system resource from data users first unless all the data users are reduced to basic service rate. The flowchart of MRV algorithm is shown in Figure 7. At first, the MRV scheme will check if all data users are with basic rate. If there exists one data user not with the basic rate, the MRV scheme will choose the data user with the maximum removal index. If the service rate of the selected user is with basic rate, then the system will remove it directly, otherwise reduce its rate to the next lower service rate. If all data users are with basic rate, the system will remove the user which is with the maximum removal index.

References

[1] D. Kim, “A simple algorithm for adjusting cell-site transmitter power in CDMA cellular systems,” *IEEE Trans. Veh. Technol.*, vol. 48, no. 4, pp.1092-1098, July 1999.

[2] L. C. Wang, C. Y. Liao, and C. J. Chang, “Performance study of soft handover with WCDMA heterogeneous cellular architectures,” *Multiaccess, Mobility and Teletraffic for Wireless Comm., MMT’02 Conf.*, pp. 87-102, June 2002.

[3] L. C. Wang, C. Y. Liao, and C. J. Chang, “Downlink Soft Handoff and Power Allocation for CDMA Heterogeneous Cellular Networks,” *IEEE GLOBECOM’02.*, vol. 2, pp. 1830-1834, Nov. 2002.

[4] S. L. Kim, Z. Rosberg, and J. Zander, “Combined power control and transmission rate selection in cellular networks,” *IEEE VTC’99*, pp. 1653-1657, 1999.

[5] C. W. Sung, W. S. Wong, ”Power control and rate management for wireless multimedia CDMA systems,” *IEEE Trans. Commun.* vol. 49, no. 7, pp. 1215-1226, July 2001.

[6] Y. W. Kim, D. K. Kim, J. H. Kim, S. M. Shin, and D. K. Sung ”Radio resource management in multiple-chip-rate DS/CDMA systems supporting multiclass services,” *IEEE Trans. Veh. Techno.*, vol. 50, no. 3, pp. 723-736, May 2001.

[7] D. Kim, ”Rate-regulated power control for supporting flexible transmission in future CDMA mobile networks,” *IEEE JSAC.*, vol. 17, no. 5, pp. 968-977, May 1999.

[8] D. K. Kim, and D. K. Sung, ”Handoff management in CDMA systems with a mixture of low rate and high rate traffics,” *IEEE VTC’98*, pp. 1346-1350, 1998.

[9] D. I. Kim, E. Hossain, and V. K. Bhargava, “Downlink joint rate and power allocation in cellular multirate WCDMA systems,” *IEEE Trans. Wireless Commun.*, vol. 2, no. 1, pp. 69-80, Jan. 2003.

[10] S. Kahn, M. K. Gurcan, and O. O. Oyefuga, “Downlink throughput optimization for wideband CDMA systems,” *IEEE Communi. Lett.*, vol. 7, no. 5, pp. 251-253, May 2003.

[11] H. Furukawa, K. Hamabe, and A. Ushirokawa, “SSDT − Site Selection Diversity Transmission Power Control for CDMA Forward Link,” *IEEE JSAC.*, vol. 18, no. 8, pp. 1546-1554, Aug. 2000.

[12] A. Osyczka, Evolutionary algorithms for single and multicriteria design optimization, *Physica-Verlag*, 2002.

[13] M. Andersin, Z. Rosberg, and J. Zander, “Gradual removals in cellular PCS with constrained power control and noise,” *ACM/Baltzer Wireless Networks J.*, vol. 2, no. 1, pp. 27-43, 1996.

[14] F. Berggren, R. Jantti, and S. L. Kim, “A generalized algorithm for constrained power control with capability of temporary removal,” *IEEE Trans. Veh. Techno.*, vol. 50, no. 6, pp. 1604-1612, Nov. 2001.

[15] S. L. Kim, “Optimization approach to prioritized transmiter removal in a multiservice cellular PCS,” *IEEE PIMRC’98*, vol. 3, pp. 1565-1569, 1998.

[16] S. A. Grandhi, J. Zander, and R. D. Yates, “Constrained power control,” *Wireless Personal Commun.*, vol. 1, no. 4, pp. 257-270, 1995.

[17] V. Erceg, S. Ghassemzadeh, M. Taylor, D. Li, and D. L. Schilling, “Urban/suburban out-of-sight propagation modeling,” *IEEE Commun. Mag.*, vol 30. , no. 6, pp. 56-61, June 1992.

[18] Unviersal Mobile Telecommunication System (UMTS); “Selection procedures for the choice of radio transmission technologies of the UMTS,” *UMTS 30.03*, version 3.2.0, TR 101 112, pp. 54-55, 1998.

[19] M. Gudmundson, “Correlation model for shadow fading in mobile radio systems,” *Electronics Letters*, vol. 27, no. 23, pp. 2145-2146, Nov. 1991.

[20] A. J. Viterbi, “CDMA:priciples of spread spectrum communication,” *Addison-Wesley*, pp. 218-224, June 1995.

[21] Shapira J., “Microcell engineering in CDMA cellular networks,” *IEEE Trans. Veh. Technol.*, vol. 43, no. 4, pp. 817-825, Nov. 1994.

[22] R. Yates, “A Framework for uplink power control in cellular radio systems,” *IEEE JSAC.*, vol. 13, no. 7, pp. 1341-1347, Sept. 1995.

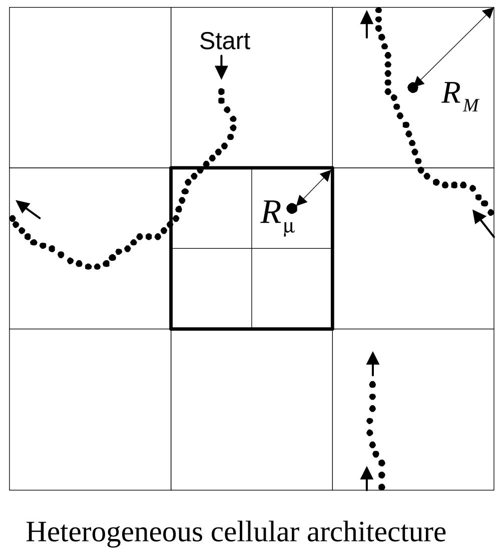


Fig. 2. Heterogeneous cellular model with an example of mobility trajectory.

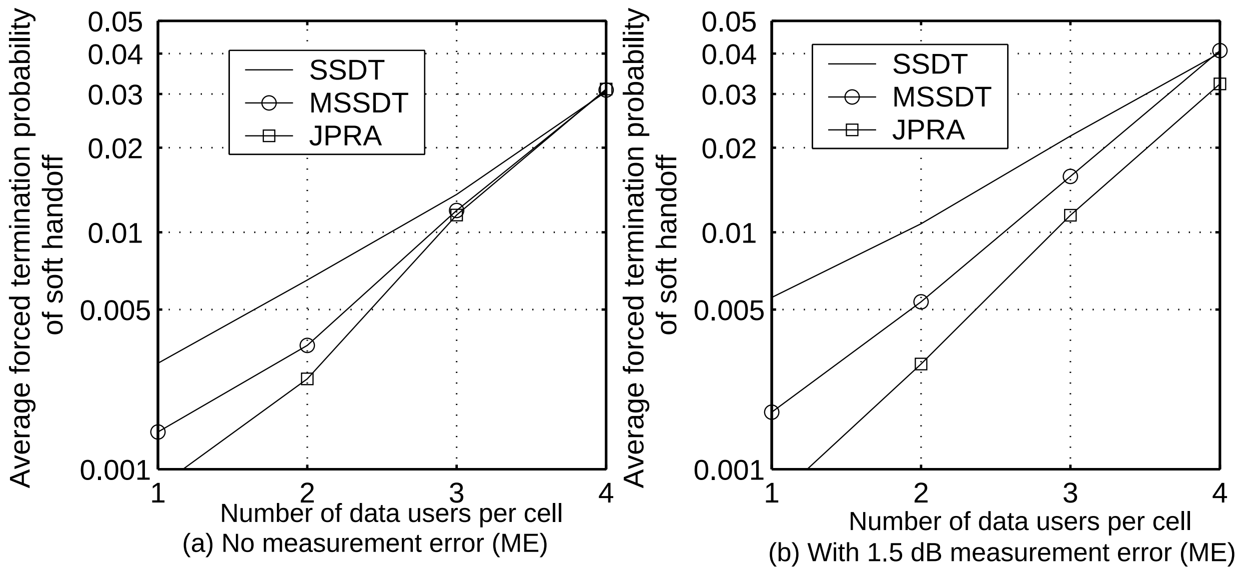


Fig. 3. Averaged forced termination probability of soft handoff (a) without measurement error (ME) and (b) with 1.5 dB measurement error (ME) during active set selection

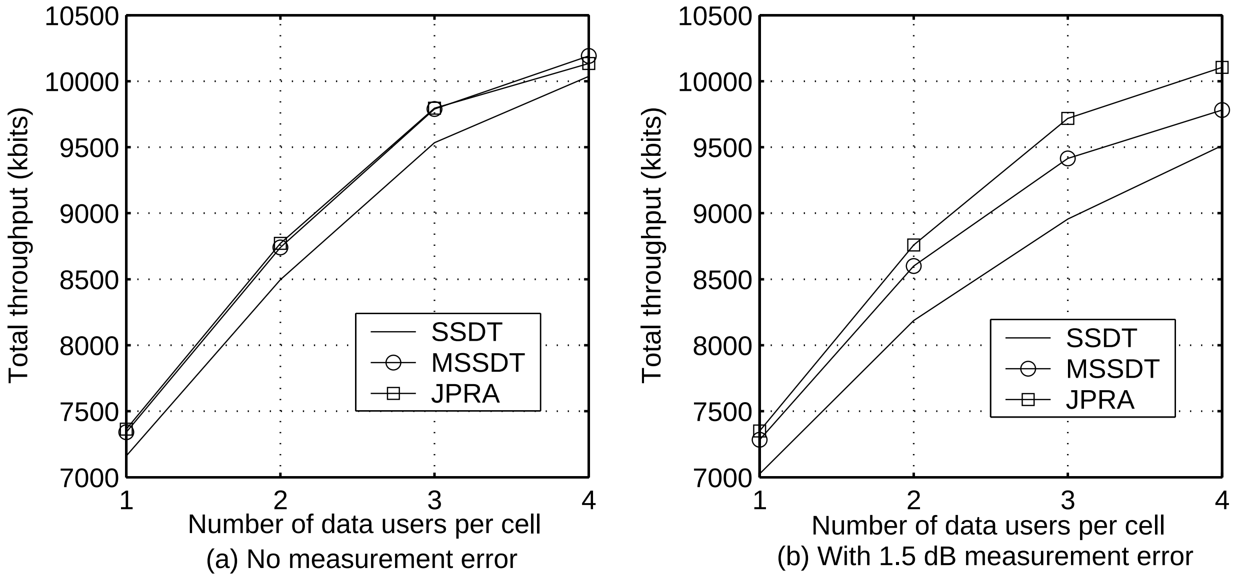


Fig. 4. The total throughput versus the number of data users per cell (a) without measurement error (ME) and (b) with 1.5 dB measurement error (ME) during active set selection

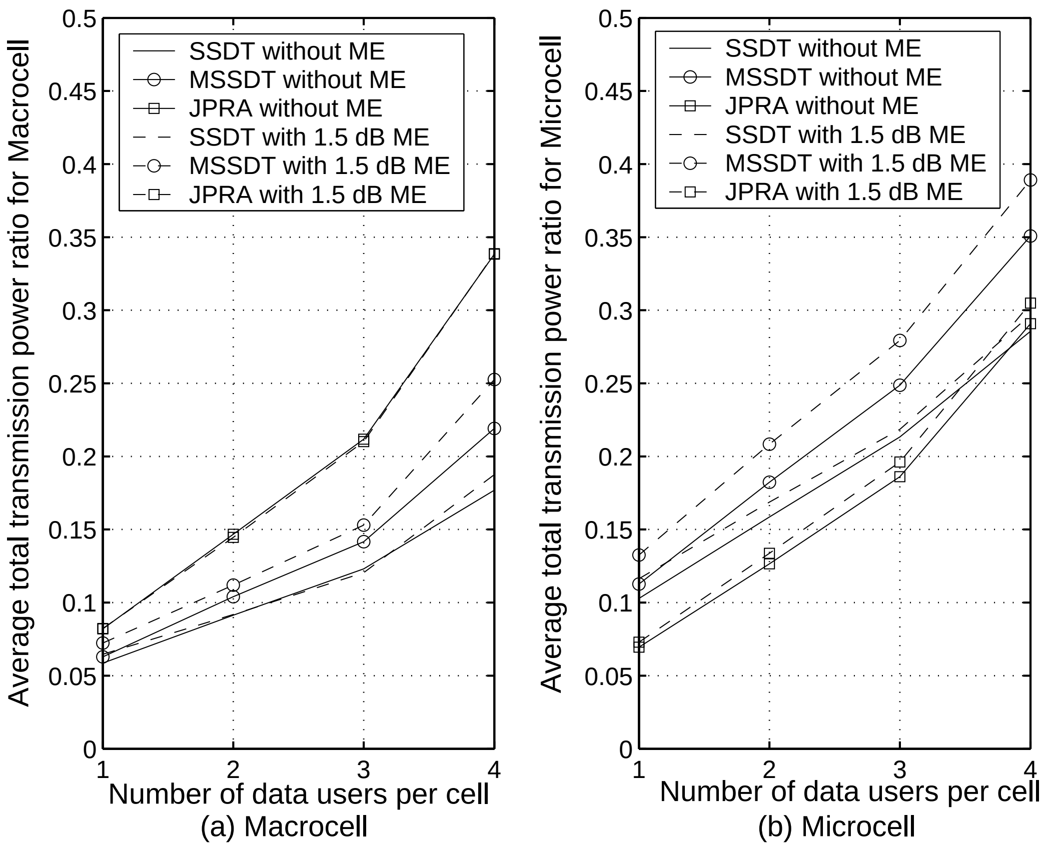


Fig. 5. Average total transmission power ratio with/without 1.5 dB measurement error (ME) for (a) macrocell and (b) microcell

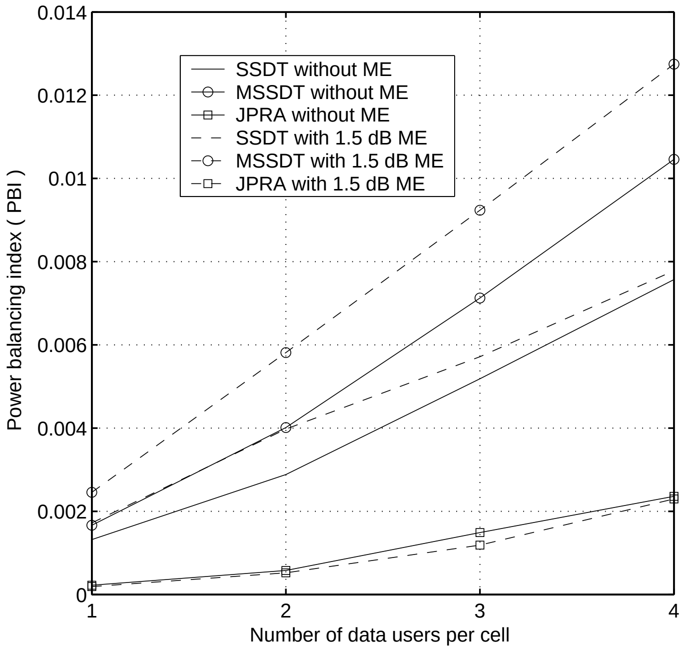


Fig. 6. Power balancing index (PBI) with/without 1.5 dB measurement error (ME)

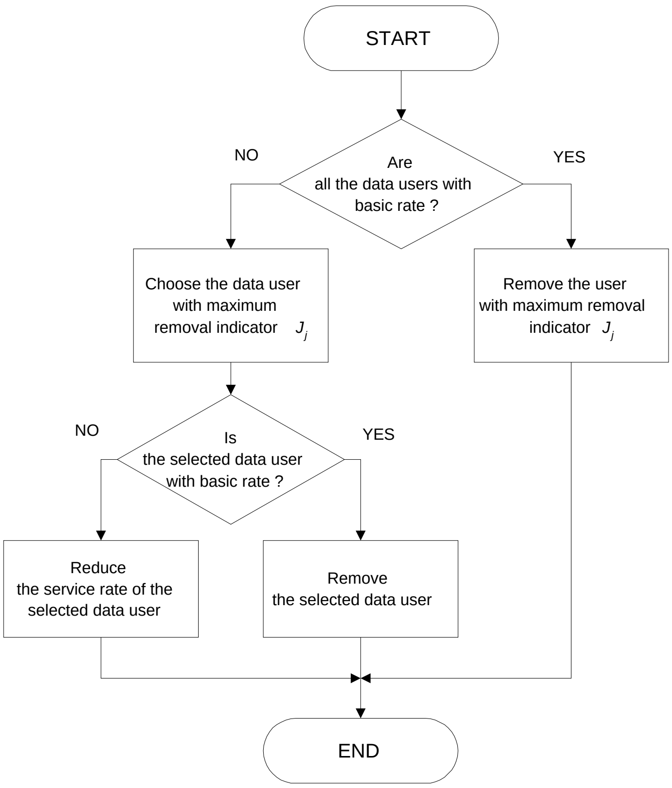


Fig. 7. Flowchart of the MRV algorithm

Conclusion

The radio resource management technologies for heterogeneous access network are investigated in this project. Two kinds of network architecture are considered: WCDMA/WLAN and macro-cell/micro-cell. For the WCDMA/WLAN, we develop novel admission/access control schemes for proper system resource allocation. Since the WLAN lacks the capability of QoS, in the first year of the project, we treat the WCDMA and WLAN separately. Firstly, a radio resource management index (Part 1) and a Q-learning-based multirate controller are proposed for the WCDMA system (Part 2), and the weighted fairness for differentiated services is studied for the WLAN system (Part 3). For the macro-cell/micro-cell heterogeneous access network (hierarchical network), a joint power and rate assignment (JPRA) algorithm for multirate soft handoff is proposed (Part 4).

In Part 1, we propose RRI scheme as a unified resource allocation metric. From computer simulation results, it can be found that RRI is a flexible and simple mapping from traffic parameters and QoS requirements for any number types of services. Moreover, the achievable resource utilization efficiency is about 80%. The reasons are that several assumptions and approximations considering the worst case system load are adopted in the derivation of the RRI to simplify the derivation and keep QoS requirements of all connections guaranteed. Also, several bounds in our lemmas have tighter ones in some specific conditions. The precision and the efficiency of the proposed RRI will be improved in the future work.

In Part 2, we propose a Q-learning-based multirate controller for multirate transmission control in WCDMA system. In the Q-MRCT, the Q-learning algorithm is being successfully applied to accurately estimate the transmission cost for the multi-rate transmission control. Also, the feature extraction method is applied to efficiently map the original state space into the *resultant interference profile*. The computer simulation results show that, compared with the interference-based scheme, Q-MRTC can improve the throughput of the WCDMA system by an amount of 87% under the constraint of the QoS requirement. In addition, the Q-MRTC provides better users' satisfaction by an amount of 50%.

In Part3, the weighted fairness for differentiated services is studied. An analytical method is proposed to obtain parameters required to achieve weighted fairness for services operating under the enhanced distributed coordinator function (EDCF) mode. In the queuing analysis, a discrete-time Markov-chain was adopted to model the behavior of backoff counters for the two classes and the steady-state probabilities were derived. The analytic model can interpret the relationship between access probability and contention window. The accuracy of the analytical solution is verified

by simulation for different number of active stations. It can conclude that, for different combination of high- and low-class STAs, the weighted fairness is easily achieved by employing the proposed method. With our proposed Markov analytic model, we will further investigate the throughput enhancement in WLAN.
In Part 4, a joint power and rate assignment (JPRA) algorithm is proposed to deal with multirate soft handoff in WCDMA heterogeneous cellular systems. JPRA allocates transmission power based on the constrained unequal power allocation (CUPA) scheme and achieves optimal service rate according to the evolutionary computing rate assignment (ECRA) method, for the multirate soft handoff. In order to support good service quality for multirate soft handoff, a proposed multirate removal algorithm (MRV) is activated to reduce service rate or even block users whenever system radio resource is insufficient. In the meantime, a new multi-quality balancing power allocation (MQBPA) scheme is also developed to allocate power for non-handoff users with multirate services. Simulation results show that, JPRA accomplishes excellent power balancing feature between cells, and it thus can improve the forced termination probability of soft handoff by 61.0%, and the total throughput by 2.4 %, as compared to conventional SSDT. Moreover, JPRA is less sensitive to the measurement error than SSDT in the active set selection.

We can conclude that (1) the RRI scheme provides a mathematical background for resource allocation in wireless network; (2) the Q-MRTC provides an intelligent technique approach for burst mode transmission control in WCDMA system; (3) adaptive adjusting the contention window is the key to provide QoS capability in WLAN.

In the second year of the project, we will work toward the heterogeneous admission control/access control scheme in WCDMA/WLAN network architecture. Also, we will investigate the dynamic cell planning for macro-cell/micro-cell heterogeneous access network using smart antenna techniques. The concept of situation-aware network planning emerges because it is hard to precisely partition the cell size in advance. The actual service area in the macro-cell/micro-cell heterogeneous access network should be dynamically adjusted according to the fluctuation of traffic so as the other RRM technologies in the heterogeneous network. We believe that the concept of situation-aware will play an important role in the development of B3G RRM technologies.

Self-assessment of the Project

In this project, our researches focus on the development of radio resource management technologies for heterogeneous access network. The time schedule of the project is keeping pace with our project proposal. In the first year's study, have completed five working items, including:

- Call admission control in heterogeneous network
- Radio resource index for heterogeneous network
- Bandwidth management for WLAN
- Novel power control algorithm
- Soft handoff algorithm for heterogeneous network

Generally speaking, we have done a great deal of works and some of them have been published in international conferences and journals [1]-[5]. As to the detail of our works, please refer to these published papers. In accordance with our proposal, our research in the coming year will keep on development of the radio resource management technologies for WLAN/WCDMA and hierarchical WCDMA systems. Based one the current research results, we envision that the highlights of our researches are three-folded:

- Call admission control for WCDMA/WLAN heterogeneous access network
- Soft handoff algorithms for hierarchical WCDMA systems usinf smart antenna
- A novel situation-aware/location-aware packet scheduling algorithm for bust-mode data transmission in heterogeneous access network.

Reference

[1] F. C. Ren, C. J. Chang and Y. S. Chen "A Q-learning-based multi-rate transmission control scheme for RRC in WCDMA systems", IEEE *PIMRC* 2002.

[2] S. Shen and C. J. Chang, "The Radio Resource Index for WCDMA Cellular Systems", submitted to IEEE *Communication Letter*

[3] C. Y. Shih, R. G. Cheng, and C. J. Chang, "Achieving Weighted Fairness for Wireless Multimedia Services", to be appeared in IEEE *VTC* 2003-*Fall*.

[4] C. H. Lee and C. J. Chang, "An accurate method for approximating the interference statistics for DS/CDMA cellular systems with power control over frequency selective fading channels", submitted to IEEE *Transaction on Wireless Communication*

[5] C. C. Chen, C. Y. Liao, and C. J. Chang, " The enhanced soft handoff algorithm for multirate DS-CDMA cellular systems", *CDMA International Conference 2002 (CIC 2002)*, Korea