A Resource Allocation Framework With Credit System And User Autonomy Over Heterogeneous Wireless Networks

Zhu Han, Z.Jane Wang, and K.J. Ray Liu,

Electrical and Computer Engineering Department, University of Maryland, College Park.

Abstract—Future wireless networks will support the growing demands of heterogeneous services. Dynamic resource allocation is essential to guarantee quality of service (QoS) and enhance the network performance. We propose a novel resource allocation framework to cope with the time-varying channel conditions, co-channel interferences, and different QoS requirements in various kinds of services. We define a QoS measurement for delay sensitive applications. We introduce a credit system, where users have their autonomy to decide when and how to use their resources, and users can borrow or lend resources from the system. We also develop a simple feedback mechanism to report the system with the users' QoS satisfaction levels and channel conditions. Then the system will adapt its resource allocation strategy according to the users' feedbacks to favor the users with the bad QoS satisfaction levels or the good channels. We develop adaptive algorithms at both the user and system levels. From simulations, the proposed algorithms efficiently allocate the resources to different types of users. The users' delay constraints are satisfied and the links can survive under a long period of bad channels.

I. INTRODUCTION

The future wireless systems are expected to provide other information services beyond voice, such as telecommuting, video conferencing, interactive media, real-time Internet games, etc., at anytime, anywhere. To satisfy the growing demand of heterogeneous applications, one of the key objectives is to deliver flexible, variable rate services with high spectral efficiency. The time-varying channel conditions, co-channel interferences (CCI), and different quality of service (QoS) requirements are the potential challenges for the system design. Dynamic resource allocation is a general strategy to enhance the system performance.

Current wireless systems choose single-to-interferencenoise ratio (SINR) as the QoS measure for voice communications. The resource allocation problem in the context of voice communications becomes power control problem [1], [2], [3], [4], [5], where the transmitted powers are constantly adjusted to achieves the users' target SINR. It has been shown in [6], [7], [8], [9], [10] that jointly considering power control and adaptive modulation can provide a variable rate and variable power ability to combat with the time varying channel and CCI. In this paper, we concentrate on the resources such as the transmitted powers and throughput of MQAM modulation.

The goal of this paper is to develop a framework of dynamic resource allocation with credit system and user autonomy for heterogeneous types of users, based on a QoS measure for delay sensitive applications. We view the problem at two levels: the macro system level and the micro user level. We also develop a feedback mechanism between the two levels. The motivations and how the framework operates are explained as follows:

- 1. Micro User Level: The goal is to let each user have the autonomy to decide when and how to use his resources according to the channel conditions and his application type. A credit system is constructed, where each user can borrow and lend resources from the system to transmit his information during different periods of times. By doing so, his resources can be "water filled" in time during the transmission, which not only guarantees the QoS, but also ensures the survival of link during the long period of bad channel conditions.
- 2. Macro System Level: The goal is to create an environment to improve the overall network performance under the users' QoS constraints. It receives the feedbacks from the users to adapt the strategy for the environment, so that the user with the bad QoS satisfaction level or good channel condition can be allocated with more resources. Moreover, the system should encourage some users to sacrifice their performance temporarily, so that the overall network performance can be improved. These users may have the incentive to sacrifice in hope for the long-term payback.
- 3. Feedback Mechanism: The goal for feedback mechanism is to provide a simple but efficient way for each user to report his level of QoS satisfaction and channel condition, on which the system will be based to modify the optimization strategies.

The rest of the paper is organized as follows: In Section II, we give the system model and MQAM modulation throughput approximation. In Section III, we explain our resource allocation framework. In Section IV, we have numerical studies. In Section V, we offer the conclusions.

II. System Model and Approximation

For the purpose to illustrate the idea and performance of our proposed framework, we consider a K-user uplink Direct-Sequence CDMA system in a single cell where each user is assigned with a signature sequence and an antenna array of L elements is employed at the base station (BS). For simplicity, we assume a synchronous system with processing gain H. For uplink, over one bit period, the received signal vector of the antenna array at the BS is:

$$\mathbf{y}(t) = \sum_{k=1}^{K} \sqrt{P_k G_k} b_k s_k(t) \mathbf{a}_k + \mathbf{n}_0(t)$$
(1)

where P_k , b_k , and s_k are the transmit power, bit, and signature of user k, respectively, G_k is the uplink gain from user

0-7803-7974-8/03/\$17.00 © 2003 IEEE

k to the BS, the spatial signature \mathbf{a}_k is the array response vector of user k, and $\mathbf{n}_0(t)$ represents the white Gaussian noise vector. We apply the chip rate filtering and sample at the chip rate. The sampled output is represented as:

$$\mathbf{Y} = \sum_{k=1}^{K} \sqrt{P_k G_k} b_k \mathbf{s}_k \mathbf{a}_k^T + \mathbf{N}_0 \tag{2}$$

where **Y** has the size *H*-by-*L*, whose l^{th} column represents the outputs of the l^{th} antenna element, \mathbf{s}_k is the signature sequence of user *k*, and **N**₀ represents the space and time white noise with zero mean and variance σ^2 .

Suppose we apply a two-dimensional temporal-spatial linear filter \mathbf{X}_i to decode the bit b_i in the MMSE sense [8]. The filter \mathbf{X}_i with size *H*-by-*L* is:

$$\mathbf{X}_{i} = \arg \min_{\mathbf{X}_{i}} E[|tr(\mathbf{X}_{i}^{H}\mathbf{Y}) - b_{i}|^{2}]$$
(3)

where $tr(\cdot)$ is the trace operation. The i^{th} user's SINR at the output of the joint temporal-spatial filter is given by:

$$\Gamma_i = \frac{P_i G_i |tr(\mathbf{X}_i^H \mathbf{s}_i \mathbf{a}_i^T)|^2}{\sum_{k \neq i} P_k G_k |tr(\mathbf{X}_i^H \mathbf{s}_k \mathbf{a}_k^T)|^2 + \sigma^2 tr(\mathbf{X}_i^H \mathbf{X}_i)}.$$
 (4)

Adaptive modulation provides the system with the ability to adjust the effective bit rate (throughput), according to the interference and channel conditions. MQAM is a modulation method that has high spectrum efficiency. Without loss of generality, we assume that each user has the unit bandwidth and the throughput is continuous. Let T_i denote the i^{th} user's throughput, which is the number of bits sent within each transmitted symbol. The BER can be approximated as a function of the received SINR and throughput [6], [7], [8], [11] given by:

$$BER_i \approx c_1 e^{-c_2 \frac{\Gamma_i}{2^{T_i} - 1}} \tag{5}$$

where $c_1 \approx 0.2$ and $c_2 \approx 1.5$ for MQAM when BER is small. From (5), for a specific BER, the i^{th} user's throughput is

$$T_i = \log_2\left(1 + c_3^i \Gamma_i\right) \tag{6}$$

where $c_{3}^{i} = -\frac{c_{2}^{i}}{ln(BER_{i}/c_{1}^{i})}$.

III. RESOURCE ALLOCATION FRAMEWORK

A. Problem Formulations

In order to implement the proposed ideas and the framework for resource allocation, we propose to formulate and solve the problems heuristically at the micro user level and the macro system level as:

1. Micro User Level: According to the transmission history, the users calculate user satisfaction factor (USF) for their QoS. The users' tolerance for delay will affect the value of USF. At time n, according to the USF and his current channel condition, the i^{th} user feedbacks the system with an acceptable throughput range $[T_i^{min}(n), T_i^{max}(n)]$. The problems include how to define USF and how to update the acceptable throughput range. 2. Macro System Level: The system employs adaptive algorithms to optimally assign different users their shares of resources according to their throughput ranges and other constraints such as the system feasibility [9], [10] and the maximum power. We assume perfect estimations of channel conditions. The problem is given by:

$$\max_{\gamma_i, P_i} \sum_{i=1}^{K} T_i(n) \tag{7}$$

s.t.
St.
St.
St.
Feasibility:
$$(\mathbf{I} - \mathbf{DF})\mathbf{P} \ge \mathbf{u}$$
,
Throughput: $T_i^{min}(n) \le T_i(n) \le T_i^{max}(n)$,
Maximum Power: $P_i \le P_{max}$,

where $\mathbf{P} = [P_1 \dots P_K]^T$, $\mathbf{u} = [u_1 \dots u_K]^T$ with $u_i = \gamma_i \sigma^2 tr(\mathbf{X}_i^H \mathbf{X}_i)/(G_i | tr(\mathbf{X}_i^H \mathbf{s}_i \mathbf{a}_i^T)|^2)$, γ_i is the targeted SINR such that $\Gamma_i \geq \gamma_i$, $\mathbf{D} = diag\{\gamma_1 \dots \gamma_K\}$, and

$$\left[\mathbf{F}_{ij}\right] = \begin{cases} 0 & \text{if } j = i, \\ \frac{G_j |tr(\mathbf{X}_i^H \mathbf{s}_j \mathbf{a}_j^T)|^2}{G_i |tr(\mathbf{X}_i^H \mathbf{s}_i \mathbf{a}_i^T)|^2} & \text{if } j \neq i. \end{cases}$$

B. User Satisfaction Factor

In this subsection, we will address how to quantify the USF which shall help adjust the resource allocation strategies, i.e., the system adapts its algorithms so that the resources are more likely to be allocated to the unsatisfied users in the future. Due to the concerns on bandwidth and real-time feature, only limited feedback is allowed, therefore, the USF should be represented efficiently, for example, by a simple real value.

Suppose that data stream is transmitted in frames. Each frame has the length of M. In this paper, the USF represents whether or not a user can transmit its frame within the desired time. We define N as the the transmit time with the strictest delay constraint. The time when the frame is completely transmitted is n' and $n' \ge N$. The current time is n. For each user, a parameter α is selected when he is admitted to the network, where α depicts the tolerance of delay for this user. We assume at each time $n \ge N$, the user has probability of $1 - \alpha$ to finish the current frame. Then we can depict the probability for the total frame transmit time n' as a geometric distribution:

$$P_r(n' = N + i) = (1 - \alpha)\alpha^i, \quad i = 0, 1, \dots$$
(8)

Different types of payloads have different delay tolerances, which are categorized as:

- 1. Strict Delay Constraint: In this case, $\alpha = 0$, P(n' = N) = 1, which means the frame must be transmitted before or at time N. It fits the voice payload.
- 2. Soft Delay Constraint: Here $0 < \alpha < 1$, the estimated time to transmit the frame is $\overline{N} = N 1 + 1/(1 \alpha)$. It fits the video/image or data payload.
- 3. No Delay Constraint: $\alpha = 1$, so $P(n' = N + i) = 0, \forall i \ge 0$, which means the user can suffer arbitrary transmission delay. It fits some generic data payload that is not time sensitive.

In the traditional wireless network, when a user is admitted to the system, his parameters are predefined to the system. Then the system assigns the resources to the user, according to his parameters. There is no feedback from the user to the system during the transmission to reflect whether or not the user really gets the desired QoS, even if the wireless channels may fluctuate. So we need to define USF for user's real QoS satisfaction such that the system can adapt its resource allocation scheme under different conditions. Define $T_i^{his}(n-1) = \sum_{j=1}^{n-1} T_i(j)$. We define the *i*th user's proposed USF at time *n* as:

$$\mathrm{USF}_{i}(n) = \alpha^{(\frac{M(n-1)}{T_{i}^{his}} - N)}.$$
(9)

If the i^{th} user maintains the average rate $T_i^{his}/(n-1)$, the estimated time to finish the frame is $n_i^{est} = M(n-1)/T_i^{his}$. So the physical meaning of USF is the probability that the user can transmit after n_i^{est} if $n_i^{est} \ge N$. If $n_i^{est} < N$, the user is over satisfied and USF > 1. The value of USF represents the user's QoS satisfaction level and has the following implications:

- 1. USF > 1: user can finish transmission even before time N and is over satisfied. He can use a lower rate to transmit during the rest of times.
- 2. **USF** = 1: in this case, user's QoS is exactly satisfied. If he uses the average rate M/\overline{N} , he can finish the frame at time \overline{N} .
- 0 ≤ USF < 1: when USF becomes smaller, the user becomes more unsatisfied and has to transmit more in the rest of times.

C. Credit System, User Autonomy, Resource Awareness

Similar to the economy system, we introduce concepts of credit system, user autonomy, and resource awareness to resource allocation. At a specific time, since the channel varies, the user may transmit more or less than the desired throughput. A credit system is constructed to allow lending or borrowing resources and record user's transmission history. If the user experiences a bad channel, he will be more aggressive to transmit in the future when the channel becomes better. In the proposed approach, the user will provide a higher acceptable throughput range to demand more resources. On the other hand, if the channel is still bad, he will delay requesting resources until the channel becomes better. So the user has his own autonomy to decide when and how to use the resources.

In order to optimize the users' autonomy for resource usages, the users need to know their current channel conditions, i.e., they have resource awareness. If the channels are good, users prefer to spend more resources for transmission, else they will wait until the channels become better. Suppose \hat{T}_i^{max} and 0 be the maximum and minimum allowable throughput provided by the system for the i^{th} user. To quantify the resource awareness, we define:

$$\kappa_i(n) = \frac{T_i(n-1)}{M/\overline{N}}.$$
(10)



Fig. 1. Throughput Range vs. USF for different κ_i

The physical meaning of $\kappa_i(n)$ is the ratio of the most current throughput at time n-1 over average desired throughput, which can represent the relative channel condition.

In the micro user level, the i^{th} user's goal is to report the system with the current acceptable throughput range $[T_i^{min}(n), T_i^{max}(n)]$, according to his USF and channel condition. If USF ≥ 1 , there is no need for the user to transmit at the rate larger than M/\overline{N} . So we have $T_i^{min}(n) = 0$. We assume the USF is uniformly distributed from $[1, \infty]$ and we select exponential function for $T_i^{max}(n)$ as:

$$T_i^{max}(n) = (M/\overline{N})e^{-(\text{USF}_i - 1)/\kappa_i(n)}.$$
 (11)

So the average $T_i^{max}(n)$ for this USF_i is equal to the throughput $T_i(n-1)$. If $0 \leq \text{USF} < 1$, we use power function to determine the throughput as:

$$T_i^{max}(n) = \hat{T}_i^{max} - (\hat{T}_i^{max} - M/\overline{N})(\text{USF}_i)^{\kappa_i(n)}$$
(12)

$$T_i^{min}(n) = (M/\overline{N})(1 - (\text{USF}_i)^{\kappa_i(n)}).$$
(13)

In Fig. 1, we give an example on how the throughput ranges change with different USF and channel conditions. Here $M/\overline{N} = 2$, $\hat{T}_i^{max} = 4$, and $\kappa_i = 0.5, 1, 2$ respectively.

By jointly considering the USF and channel condition $\kappa_i(n)$, the adaptive algorithm for each user is given by:

D. Adaptive Algorithm for Macro System Level

At the macro system level, the goal is to select the best throughput allocation method to different users to generate the maximum overall system throughput under the constraints. In [10], we developed a projected gradient method. In this paper, we will develop a much faster barrier method by using the idea from semi-definition programming [12].

GLOBECOM 2003

0-7803-7974-8/03/\$17.00 © 2003 IEEE

The basic idea for the barrier method is to add barrier functions to the optimization goal such that the sum approaches negative infinity if the constraints are not satisfied. On the other hand, if the constraint is satisfied, the barrier function doesn't affect the optimization goal. The barrier function is commonly approximated by logarithmic barrier functions given by:

$$I_{constraint} \approx \Phi_1 + \Phi_2 + \Phi_3 + \Phi_4 \tag{14}$$

where Φ_1 is for $T_i(n) > T_i^{min}(n)$, Φ_2 is for $T_i(n) < T_i^{max}(n)$, Φ_3 is for feasibility, and Φ_4 is for P_{max} :

$$\Phi_1 = \begin{cases} \sum \ln(T_i(n) - T_{min}(n)), \ T_i(n) > T_{min}(n), \\ -\infty, \text{ otherwise.} \end{cases}$$
(15)

$$\Phi_2 = \begin{cases} \sum \ln(T_{max}(n) - T_i(n)), \ T_i(n) < T_{max}(n), \\ -\infty, \text{otherwise.} \end{cases}$$
(16)

$$\Phi_3 = \begin{cases} \ln det (\mathbf{I} - \mathbf{DF}), \text{ if } (\mathbf{I} - \mathbf{DF}) > 0, \\ -\infty, \text{ otherwise.} \end{cases}$$
(17)

$$\Phi_4 = \begin{cases} \sum_{i=1}^{K} \ln (P_{max} - P_i), \text{ if } P_i < P_{max}, \\ -\infty, \text{ otherwise.} \end{cases}$$
(18)

The approach for barrier method is to solve the constrained optimization problem by a sequence of unconstrained problems. We rewrite (7) as:

$$\max_{\gamma_i, P_i} f = \tilde{t} \sum_{i=1}^{K} T_i(n) + I_{constraint}$$
(19)

where \tilde{t} is a value that increases from iteration to iteration. The barrier functions become more and more like the ideal barrier function, when \tilde{t} is increasing. So the solution is more and more close to the optimal solution. Within each iteration, we use Newton method [12] to solve the unconstrained optimization problem. The algorithm is given by:

Barrier Method for Macro Throughput Maximization

where m is the iteration number for barrier method, δ determines the accuracy of the proposed algorithm, t' is the optimal step for the Newton method, t_0 is the initial value for barrier function, whose value determines the convergence rate of the first iteration, and β is the constant that \tilde{t} is multiplied in each iteration.

IV. SIMULATION RESULTS

We assume a linear array of omni directional antennas with L = 4 elements equispaced at half a wavelength. All K = 80 users are uniformly distributed within the range of $[r_0, r]$ with $r_0 = 50$ m being the closest distance and r = 1000m being the cell radius. H = 64. The mobile users move in arbitrary directions with speeds uniformly distributed in the range [0, 40] kph. We consider three phenomena in the propagation model: the path loss factor is 3.5 and a constant factor is chosen to yield a 30dB loss at 1m; the slow shadowing fading is modelled as a lognormal distribution with 3dB standard deviation; three paths with equal power Rayleigh fading with negligible delay spreads are considered. The fading is generated by the Jakes model with a $\pi/10$ angle spread. The update is taken every 10ms.

In Fig. 2 and Fig. 3, we show the throughput and USF for different types of users at different transmission times. Here the packet size M = 30, $\hat{T}_i^{max} = 4$, $\hat{T}_i^{min} = 0$, $\forall i$, and N = 15. We assume user 1 to 20 have $\alpha = 0$, user 21 to 40 have $\alpha = 0.9$, user 41 to 60 have $\alpha = 0.95$, and user 61 to 80 have $\alpha = 1$. The figure is brighter when the throughput is large and USF is large. The behaviors of different types of users are summarized as:

- 1. $\alpha = 0$: USF is equal to 0 or 1 and the transmission rate is always high because each user has to transmit his frame before the strict deadline.
- 2. $0 < \alpha < 1$: The transmission rate is determined by the channel condition. When the users with good channel conditions finish their frames early, their USF will be high so that they demand less throughput in the future.
- 3. $\alpha = 1$: USF is always equal to 1. The transmission is concentrated when the system is less busy. For example, the throughput is high around time 50 when most of users from No. 21 to 60 finish their transmissions.

From the simulation results, we can see that the proposed algorithms allocate system resources according to the service types, USF, and channel conditions.

Fig. 4 shows typical delay spreads for three schemes: our proposed scheme, Round Robin [13], and greedy scheduling (Traditional scheduling to maximize system throughput with $T_i(n) \in [0, \hat{T}_i^{max}], \forall i$). We order the users from the best channel to the worst. M = 100, N = 55, and $\alpha = 0.9$. Round Robin is a strict fair scheduling, but it has the poorest performance. Scheduling has the highest system throughput, but the users suffer arbitrary delays. While in the proposed scheme, the delays are more strict around the desired value.

In Fig. 5, we show the throughput loss for different types of services and different α . If all the users have arbitrary delay constraint $\alpha = 1$, system will have the largest average throughput and we use this value to compare with the other situations. When all the users have the same delay constraint (0%), if α is too small, the system will be infeasible. This is because the links can not survive in the long bad channel conditions. When $\alpha > 0.83$, the users can



Fig. 2. Throughput for Different Payloads vs. Transmit Times



Fig. 3. USF for Different Payloads vs. Transmit Times

survive by borrowing from our credit system. On the other two curves, we have 25% and 50% of users with $\alpha = 1$. We can see that the strict delay will degrade the system performance most. The proposed algorithm can perform better if users' service types are more diversified.

V. Conclusions

We propose a resource allocation framework for heterogeneous types of services. We define the QoS measure for delay sensitive applications. We introduce the concepts of credit system, user autonomy, and resource awareness. The users can borrow or lend resources from the credit system and decide when and how to use their resources within their transmission time. An adaptive algorithm is developed for the user level to feedback users' demands for throughput according to their USF and current channel conditions. An adaptive algorithm is developed for the system level to adapt resource allocation strategy according to the users' feedbacks. From the simulation results, the proposed algorithms allocate the resources to different types of users to maximize the system performance and guarantee QoS. The links can survive in the long bad channel conditions.

References

 R. Yates, "A framework for uplink power control in cellular radio systems", *IEEE J. Sel. Areas Commun.*, vol.13, no.7, pp.1341-



Fig. 4. An Example of Delay Spread



Fig. 5. Throughput Loss For Different Types of Payloads

1348, Sep. 1995.

- [2] J. Zander and S.L. Kim, "Radio resource managment for wireless networks", Artech House Publishers, 2001.
- [3] G. J. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence", *IEEE Trans. Vehicular Technology*, vol.42, no.4, pp.641646, Nov. 1993.
- [4] W. Yu, W. Rhee, and J. M. Cioffi, "Optimal power control in multiple access fading channels with multiple antennas", *IEEE ICC*, pp. 575579, Jun. 2001.
- [5] C. U. Saraydar, N. B. Mandayam, and D. J. Goodman, "Efficient power control via pricing in wireless data networks", *IEEE Trans. Commun.*, vol.50, no.2, pp.291-303, Feb. 2002.
- [6] X. Qiu and K. Chawla, "On the performance of adaptive modulation in cellular systems", *IEEE Trans. Commun.*, vol.47, no.6, pp.884–895, Jun. 1999.
- [7] M. H. Ahmed, et.al, "Performance enhancement of joint adaptive modulation, coding and power control using cochannel-interferer assistance and channel reallocation", WCNC'03, Mar. 2003.
- [8] S. Ulukus, E. Biglieri and M. Z. Win, "Optimum modulation and multicode formats in CDMA systems with multiuser receivers", *IEEE Infocom*, Anchorage, AK, Apr. 2001.
 [9] Z. Han and K.J.R. Liu, "Power minimization under constant
- [9] Z. Han and K.J.R. Liu, "Power minimization under constant throughput constraint in wireless networks with beamforming", *Proc. of IEEE VTC.*, Vancouver, Sep. 2002.
- [10] Z. Han and K.J.R. Liu, "Throughput maximization using adaptive modulation in wireless networks with fairness constraint", *Proc. of WCNC*, New Orleans, Mar. 2003
- [11] S.T. Chung and A.J. Goldsmith, "Degrees of freedom in adaptive modulation: a unified view", *IEEE Trans. Commun.*, vol.49, pp.1561-1571, Sep. 2001.
- [12] L. Vandenberghe and S. Boyd, "Semidefinite programming", SIAM Review 38, pp. 49-95, Mar. 1996.
- [13] S. Ramakrishna and J.M. Holtzman, "A scheme for throughput maximization in a dual- class CDMA system", *IEEE Journal on Selected Areas in Commun.*, vol.16, no.6, pp.830-844, Aug. 1998.