Utility-Based Resource Allocation in Wireless Networks

Wen-Hsing Kuo and Wanjiun Liao

***Abstract*— In this paper, we study utility-based maximization for resource allocation in the downlink direction of centralized wireless networks. We consider two types of traffic, i.e., best effort and hard QoS, and develop some essential theorems for optimal wireless resource allocation. We then propose three allocation schemes. The performance of the proposed schemes is evaluated via simulations. The results show that optimal wireless resource allocation is dependent on traffic types, total available resource, and channel quality, rather than solely dependent on the channel quality or traffic types as assumed in most existing work.**

***Index Terms*— Utility-based maximization, resource allocation, wireless networks.**

I. Introduction

RESOURCE allocation is an important research topic in wireless networks [1-10, 12-16]. In such networks, radio resource is limited, and the channel quality of each user may vary with time. Given channel conditions and total amount of available resource, the system may allocate resource to users according to some performance metrics such as throughput and fairness [1], [2] or according to the types of traffic [3]. "Throughput" and "fairness," however, are conflicting performance metrics. To maximize system throughput, the system will allocate more resource to the users in better channel conditions. This may cause radio resource monopolized by a small number of users, leading to unfairness. On the other hand, to provide fairness to all users, the system tends to allocate more resource to the users in worse channel conditions so as to compensate for their shares. As a result, the system throughput may be degraded dramatically. The work in [4-5] show that the system can behave either "throughput-oriented" or "fairness-oriented" by adjusting certain parameters. However, they do not describe how to determine and justify the value of these parameters, leaving this trade-off unsolved.

In this paper, we focus on "user satisfaction" for resource allocation to avoid such a "throughput-fairness" dilemma. Since it is unlikely to fully satisfy the different demands of all users, we turn to maximize the total degree of user satisfaction. The degree of user satisfaction with a given amount of resource can be described by the utility function $U(r)$, a non-decreasing function with respect to the given amount of resource r. The more the resource is allocated, the more the user is satisfied. The marginal utility function defined by $u(r) = \frac{dU(r)}{dr}$ is the derivative of the utility function $U(r)$ with respect to the given amount of resource r. The exact expression of a utility function may depend on traffic types, and can be obtained by studying the behavior and feeling of users. We leave the work of finding utility functions to psychologists and economists, and focus on maximizing the total utility for a given set of utility functions.

We are not the pioneer to study utility-based resource management in wireless networks. Proposals [6-10] are the examples which are based on the utility functions of different objectives for wireless networks. In [6], a utility-based power control scheme with respect to channel quality is proposed. In that scheme, users with higher SIR values have higher utilities, and thus are more likely to transmit packets. Therefore, the wireless medium can be better utilized and the transmission power can be conserved. The work in [7] propose a utility-based bandwidth allocation scheme, which can adapt to channel conditions and guarantee the minimum utility requested by each user. In [8-9], the authors design a utility-based fair allocation scheme to ensure the same utility value for each user. However, letting users with different traffic demands to achieve an identical level of satisfaction may not be an efficient way of using wireless resource. Worse, traffic which is difficult to be satisfied tends to consume most of the system resource, leading to another kind of unfairness. In [10], a utility-based scheduler together with a Forward Error Correction (FEC) and an ARQ scheme is proposed. That work gives lagging users more resource and thus results in a similar performance level (i.e., fixed utility value) for each user. The work in [19-20] targets at multi-hop wireless networks.

Utility functions have also been widely used in Internet pricing and congestion control [11]. Many bandwidth pricing schemes have been proposed for wireless networks [12-16]. The typical approach is to set a price to radio resource and to allocate tokens to users. The objective is then to maximize the "social welfare" through a bidding process. These kinds of bidding schemes, while useful for Internet pricing and congestion control, may not be practical for wireless networks. In wireless environments, the types of traffic, the number of users, and channel conditions are all time-varying. It would be very expensive to implement a wireless bidding process because the users would have to keep exchanging control messages for real-time bidding, and the control protocols of the wireless system would also have to be modified to accommodate this process. Finally, the complexity and efficiency of wireless bidding have not been analyzed. It is hard to estimate the time elapsed to achieve the Nash equilibrium.

In this paper, we study the wireless resource allocation

Manuscript received December 1, 2005; revised September 15, 2006; accepted December 22, 2006. The associate editor coordinating the review of this paper and approving it for publication was J. Hou. This work was supported by National Science Council (NSC), Taiwan, under a Center Excellence Grant NSC95-2752-E-002-006-PAE, and under Grant Number NSC95-2221-E-002-066.

W.-H. Kuo and W. Liao (corresponding author) are with the Department of Electrical Engineering and the Graduate Institute of Communication Engineering, National Taiwan University, Taipei, Taiwan (email: wjliao@ntu.edu.tw).

Digital Object Identifier 10.1109/TWC.2007.05942.

1536-1276/07$25.00 © 2007 IEEE

Authorized licensed use limited to: National Taiwan University. Downloaded on January 20, 2009 at 23:15 from IEEE Xplore. Restrictions apply.

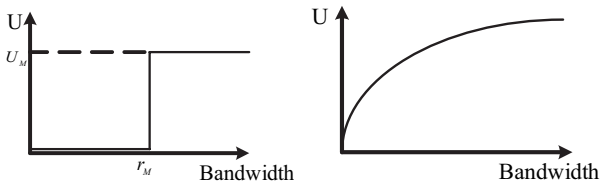
problem in the downlink of a wireless network with a central control system, such as a cellular base station, and attempt to maximize the total utility of all users without a bidding process. We consider two common types of traffic: hard QoS and best effort traffic, and propose three allocation algorithms[1] for these two types of traffic, namely, 1) the $HQ\ allocation$ for hard-QoS traffic, 2) the $elastic\ allocation$ for best effort traffic, and 3) the $mixed\ allocation$ for the co-existence of both types of traffic. These three allocation schemes are all polynomial time solutions and proved to be optimal under certain conditions, and in any case, the difference between the total utilities obtained by our solutions and the optimal utility are bounded. We also develop some theorems as the general design guidelines for utility-based resource allocation in wireless networks. The performance of the proposed schemes is validated via simulations. The results show that optimal wireless resource allocation depends on the traffic demand, total available resource, and wireless channel quality, rather than solely dependent on channel quality or traffic type as assumed in most existing work.

The rest of the paper is organized as follows. In Sec. II, three wireless allocation schemes are proposed and proved to be optimal under certain conditions. Some theorems are developed as design guidelines for utility-based resource allocation schemes. In Sec. III, the performance of the proposed schemes is validated via simulations. Finally, the paper is concluded in Sec. IV.

II. Utility-based Allocation Schemes for Two Types of Traffic

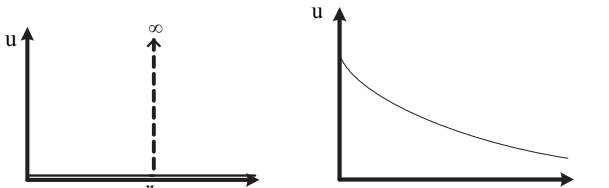
A. Problem Statement and Definitions

Suppose that there are n users served by a base station. Let r_{total} denote the total amount of radio resource available at the base station, and r_i, the amount of resource to be allocated to user i. Users with the same kind of traffic may not feel the same way by given the same amount of resource because the wireless channel quality for each user may not be identical. Let q_i denote the channel quality[2] of user i, $0 \le q_i \le 1$, and $i = 1, 2, \cdots n$. The smaller the value of q_i, the worse the channel quality. Given an amount of resource r_i and channel quality q_i, the amount of resource actually beneficial to user i is given by $\theta_i = r_i \cdot q_i$. Let $T(i)$ denote the type of traffic of user i. The utility function of user i is expressed by $U_i(r_i) = U_{T(i)}(r_i q_i)$, where $U_{T(i)}(.)$ is the utility function of traffic $T(i)$ and $U_i(.)$ is the utility function for the type of traffic described by $U_{T(i)}(.)$ but taking into account the channel quality of user i. The marginal utility function of $U_i(.)$ is $\frac{dU_{T(i)}(r_i q_i)}{dr_i} = q_i \cdot u_{T(i)}(r_i q_i)$, and that of $U_{T(i)}(.)$ is $u_{T(i)}(.)$.



(a) Hard QoS (b) Best effort

Fig. 1. The utility functions of two types of traffic.



(a) Hard QoS (b) Best effort

Fig. 2. The marginal utility functions of two types of traffic.

Our objective is to maximize $\sum_{i=1}^{n} U_i(r_i)$, subject to $\sum_{i=1}^{n} r_i \le r_{total}$ and $\forall r_i \ge 0$. An optimal allocation for n users with total available resource r_{total} is defined as follows. Note that the optimal allocation may not be unique in the system.

Definition 2.1: A resource allocation $\Re^* = \{r_1, r_2, \cdots, r_n\}$ for n users is an optimal allocation if for all feasible allocations $\Re_a = \{r'_1, r'_2, \cdots, r_n'\}$, $U(\Re^*) \ge U(\Re_a)$, where $U(\Re^*) = \sum_{i=1}^{n} U_i(r_i)$ and $U(\Re_a) = \sum_{j=1}^{n} U_j(r'_j)$.

Definition 2.2: A unit-step utility function $U_{step}(r)$ refers to a utility function whose $u_{step}(r_M) = \infty$ if $r = r_M$, and $u_{step}(r) = 0$, otherwise, where $u_{step}(r) = \frac{dU_{step}(r)}{dr}$.

Definition 2.3: A concave utility function $U_{concave}(r)$ refers to a utility function whose $u_{concave}(r) > 0$ and $u'_{concave}(r) < 0$ for all r, where $u_{concave}(r) = \frac{dU_{concave}(r)}{dr}$ and $u'_{concave}(r) = \frac{du_{concave}(r)}{dr}$.

By definition, a unit-step function is a discrete function, and a concave utility function is a non-decreasing and continuous function with respect to resource r. Fig. 2 plots the marginal utility functions for the two types of traffic shown in Fig. 1 More terminology used in this paper is defined as follows.

Definition 2.4: $\Re = \{r_1, r_2, \cdots r_n\}$ is a *full* allocation if $\sum_{\forall r_i \in \Re} r_i = r_{total}$.

Definition 2.5: For $\Re = \{r_1, r_2, \cdots r_n\}$, all users i with $r_i > 0$ (i.e., those who are allocated resources) are referred to as allocated users and all users j with $r_j = 0$ (i.e., those who are not allocated resources) are unallocated users.

Definition 2.6: $\Re = \{r_1, r_2, \cdots r_n\}$ is *marginally fair* if it satisfies the following two conditions:

1) Each allocated user i (i.e., with $r_i > 0$) in $\Re$ must have the same marginal utility value.
2) For each unallocated user j (i.e., with $r_j = 0$) in $\Re$, its marginal utility value $u_j(0)$ cannot exceed $u_i(r_i)$, which is the marginal utility of each allocated user i.

[1]These algorithms can help determine and validate the parameter settings of existing schedulers. Most of the existing wireless schedulers focus on the design of scheduling systems which provide such performance guarantees as delay bounds of packets or fairness among users, but leave the weights of the queues (or users) undecided. This missing component can be provided by optimal resource allocation schemes as proposed in this paper.

[2]The channel quality parameter q_i is provided by the lower layers to indicate the proportion of effective transmission with a given amount of radio resource. Since q_i may be time-variant, the proposed algorithm should be operated periodically to adapt to the current channel condition. For simplicity, but without loss of generality, we will treat q_i as a constant for each user i in the rest of the paper.

Authorized licensed use limited to: National Taiwan University. Downloaded on January 20, 2009 at 23:15 from IEEE Xplore. Restrictions apply.

Definition 2.7: For a marginally fair allocation $\Re = \{r_1, r_2, \cdots r_n\}$, the marginal utility value of each allocated user in $\Re$ is equal, and is referred to as the allocated marginal utility $u_m(\Re)$. Thus, for each allocated user i in $\Re$, $u_i(r_i) = u_m(\Re)$, and for each unallocated user j, $u_j(0) < u_m(\Re)$.

B. HQ Allocation for Hard QoS Traffic

Suppose that there are n users in the queue, all with hard QoS traffic. Let $r_{residue}$ denote the residual resource in the system. The resource allocation algorithm designed for users whose utility functions are all unit-step functions is referred to as the $HQ\ allocation$ and the output is denoted by $\Re_{HQ} = \{r_1, r_2, \cdots r_n\}$. Given the total available resource in the system r_{total}, the channel quality q_i and utility function $U_{T(i)}(.)$ for all user i, $\Re_{HQ}$ can be obtained as follows.

1) Initialize $r_i \leftarrow 0$, $i = 1, 2, \cdots, n$; $r_{residue} \leftarrow r_{total}$.
2) Sort all users i in the queue in descending order of $\frac{U_{M_i} q_i}{r_{M_i}}$.
3) Repeat Steps (4) and (5) until the queue becomes empty.
4) Pop out user i who is now at the head of the queue.
5) If $r_{residue} > \frac{r_{M_i}}{q_i}$, then
   $r_i = \frac{r_{M_i}}{q_i}$; $r_{residue} = r_{residue} - r_i$.

The utility function for user i with hard QoS traffic is described by $U_{T(i)}(r) = U_{M_i} \times f_u(q_i r - r_{M_i})$, where $f_u(.)$ is a unit-step function, q_i is the channel quality of this user, M_i is the kind of QoS traffic, r_{M_i} is the preferred amount of resource to be allocated.

The allocation rule of the $HQ\ allocation$ is to assign resources to users in descending order of $\frac{U_{M_i} q_i}{r_{M_i}}$, subject to $\sum_{i=1}^{k} r_i \leq r_{total}$, where k is the largest value satisfying this constraint, as illustrated in Fig. 3. The allocation problem for users with arbitrary unit-step utility functions can be proved to be NP-complete. The performance of the $HQ\ allocation$ can be proved to be close to the optimum. When the utility functions of all users are identical, the HQ allocation can be proved to be optimal.

Lemma 2.1: The allocation problem is NP-complete if the utility functions for all users are arbitrary unit-setp functions.

Proof: This problem can be reduced from the 0/1 knapsack problem, an NP-complete problem. Consider a knapsack with capacity $c > 0$ and n items. Each item has a value of $v_i > 0$ and a weight of $w_i > 0$. The problem is then to find a selection of items that maximize $\sum_{i=1}^{n} \delta_i v_i$ subject to $\sum_{i=1}^{n} \delta_i w_i \leq c$, where $\delta_i = 1$ if the item is selected, and 0, otherwise. Therefore, any instance of the knapsack can be reduced to an instance of our problem by substituting $r_{total} = c$, $U_i = v_i \times f_{step}(q_i r - w_i)$ and $q_i = 1$, for $i = 1, 2, \cdots, n$. Since an optimal solution to our problem is also a solution to the given knapsack problem, the knapsack problem is a special case of our problem, it follows that our problem is NP-hard.

Next, it can be observed that our problem is an NP problem because any given solution can be verified as a feasible solution and bounded at a given utility value u in polynomial time. Since our problem is both in NP and NP-hard, it is an NP-complete problem. ■

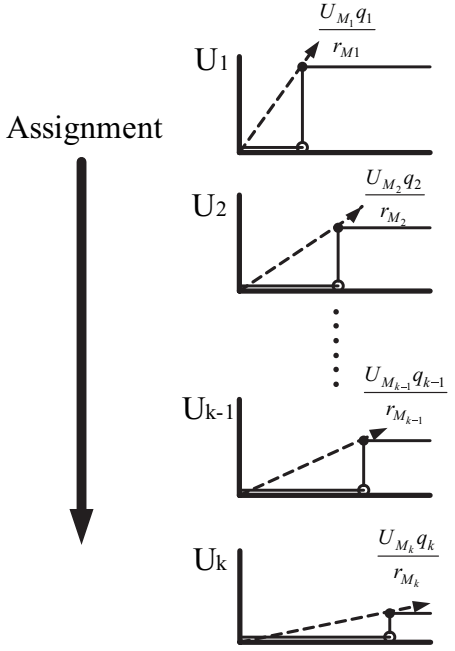


Fig. 3. Allocation ordering of k users in the *HQ allocation*

Theorem 2.2: $U(\Re_{HQ}) \geq U(\Re_{op}) - U_{\max}$, where $\Re_{HQ}$ and $\Re_{op}$ are the proposed solution and the optimal solution, respectively, of this $HQ\ allocation$ problem, $U(x)$ is the total utility of all users for solution x, and U_{max} is the maximum utility value taken over all users, i.e., $U_{\max} = \max_{1 \leq k \leq n} \{U_{M_k}\}$.

Proof: Let p denote the original $HQ\ allocation$ problem, which is NP-complete as proved in Lemma 2.1, and let p' denote the problem by relaxing the integrity constraint of problem p (i.e., each user in p' can be served fractionally). To better indicate the optimal solution to each problem, we let $\Re_{op}^{p}$ and $\Re_{op}^{p'}$ represent the optimal solutions to problems p and p', respectively. Since p is a maximization problem, we have $U(\Re_{op}^{p'}) \geq U(\Re_{op}^{p}) \geq U(\Re_{HQ}^{p})$, where $\Re_{HQ}^{p}$ means $\Re_{HQ}$.

The optimal solution to p', $\Re_{op}^{p'}$, is obtained by sorting QoS users in the queue in decreasing order of $U_{M_i} q_i / r_{M_i}$ as in $\Re_{HQ}^{p}$. The difference between $\Re_{op}^{p'}$ and $\Re_{HQ}^{p}$ is only in those users fractionally served in $\Re_{op}^{p'}$. It follows that $U(\Re_{HQ}^{p}) + U_x > U(\Re_{op}^{p'})$, where U_x is the utility value of the unallocated user whose $\frac{U_{M_i} q_i}{r_{M_i}}$ in $\Re_{HQ}^{p}$ is the largest. Thus, $U_x \leq U_{max}$. Since $U(\Re_{HQ}) + U_{\max} \geq U(\Re_{HQ}^{p}) + U_x > U(\Re_{op}^{p'}) \geq U(\Re_{op}^{p})$, we obtain $U(\Re_{HQ}^{p}) > U(\Re_{op}^{p}) - U_{\max}$. ■

Theorem 2.3: The $HQ\ allocation$ solution $\Re_{HQ}$ is an optimal allocation if the unit-step utility functions $U_{step}(r)$ of all n users in the queue are identical.

Proof: Let U_M denote the unit-step utility function for all users. The inequality in Theorem 2.2 can be rewritten as $nU_M + U_M > U(\Re_{op}^{p'}) \geq U(\Re_{op}^{p}) \geq nU_M$, i.e., $n + 1 > \frac{U(\Re_{op}^{p'})}{U_M} \geq \frac{U(\Re_{op}^{p})}{U_M} \geq n$. Since both $\frac{U(\Re_{op}^{p})}{U_M}$ and n are integral, $U(\Re_{op}^{p}) = nU_M$. Therefore, $\Re_{HQ}$ must be optimal. ■

Theorem 2.4: The time complexity of $\Re_{HQ}$ is $O(n \log n)$.

Authorized licensed use limited to: National Taiwan University. Downloaded on January 20, 2009 at 23:15 from IEEE Xplore. Restrictions apply.

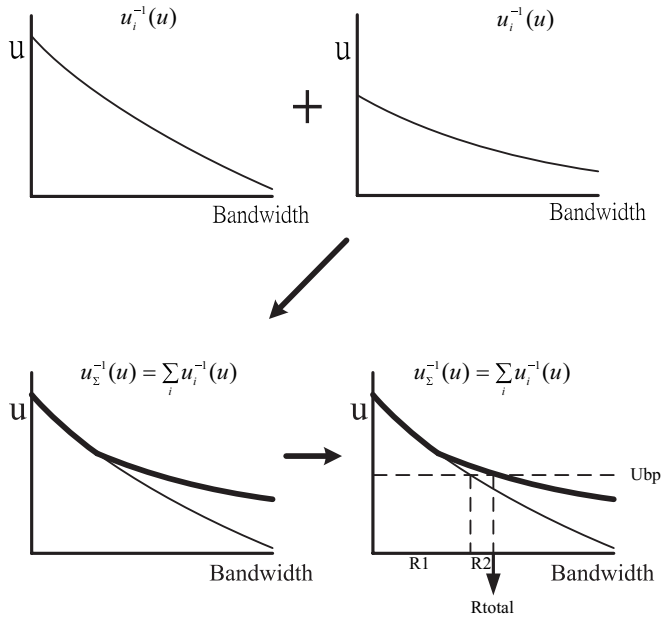


Fig. 4. An example of the *elastic allocation*. First, the marginal utility functions of users i and j are inverted (i.e., $u_i^{-1}(r)$ and $u_j^{-1}(r)$). Then, the two inverted marginal utility functions are summed, i.e., $u_\Sigma^{-1}(u) = \sum_i u_i^{-1}(u)$. Finally, the relationship among u_{BE}, r_{total}, r_1, and r_2 is decomposed.

Proof: The time complexity of $\Re_{HQ}$ can be expressed by a function of the number of users in the network. Since the complexity of Step (2) (sorting) is $O(n \log n)$ and this iteration dominates the operation, the overall complexity of $\Re_{HQ}$ is $O(n \log n)$. ■

C. Elastic Allocation for Best Effort Traffic

We next consider the best effort traffic. The resource allocation algorithm for users with concave utility functions $U_i(r)$ is referred to as the *elastic allocation* and the output is denoted by $\Re_{elastic} = \{r_1, r_2, \cdots r_n\}$. Given the total available resource r_{total}, the channel quality q_i and marginal utility function $u_i(r)$ for each user i, $\Re_{elastic}$ can be obtained as follows.

1) For each user i, derive $u_i^{-1}(u)$, the inverted function of $u_i(r)$.
2) Derive $u_\Sigma^{-1}(u)$ by summing up $u_i^{-1}(u)$ over all users i, i.e., $u_\Sigma^{-1}(u) = \sum_i u_i^{-1}(u)$.
3) Find $u_\Sigma(r)$, the inverted function of $u_\Sigma^{-1}(u)$.
4) Find u_{BE}, which is equal to $u_\Sigma(r_{total})$.
5) For all r_i, $i = 1, 2, \cdots, n$,
   if $u_{BE} < u_i(0)$, then $r_i = u_i^{-1}(u_{BE})$;
   else $r_i = 0$.

The allocation rule of this scheme is to 1) derive the aggregated utility function from the inverse functions of all users, 2) calculate the allocated marginal utility from the aggregated utility function, and 3) determine r_i for each user. As an example, Fig. 4 illustrates the *elastic allocation* algorithm with two best effort traffic .

Theorem 2.5: The *elastic allocation* $\Re_{elastic}$ is optimal if the utility functions for all users are concave utility functions.

Proof: In the *elastic allocation*, 1) u_{BE} is the marginal utility value at which point the total resource has been fully allocated, and 2) for all allocated users i in $\Re_{elastic}$, $u_{BE} < u_i(0)$ and $u_i(r_i) = u_{BE}$; for all unallocated users j, $u_{BE} \geq u_j(0)$. With concave utility functions, i.e., function with $u(r) > 0$ for all r, it can be easily proved by contradiction that an optimal allocation for elastic users, i.e., $\Re_{elastic}$, must be *full*. Similarly, by using contradiction again, we can prove that $\Re_{elastic}$ must also be *marginally fair*. Since all users' utility functions are increasing, only one allocation can be both *full* and *marginally fair*. Therefore, if an allocation is *full* and *marginally fair*, it must be $\Re_{elastic}$. ■

Theorem 2.6: The time complexity of the algorithm *elastic allocation* is $O(n)$.

Proof: Since the operation at each step at most takes time $O(n)$, the time complexity of the *elastic allocation* algorithm is $O(n)$. ■

D. A Mixture of Hard QoS and Best Effort Traffic

Finally, we consider the co-existence of QoS and best effort traffic in the system, which is referred to as $mixed\ allocation$ and the output of which is denoted by $\Re_{mix} = \{r_1, r_2, \cdots r_n\}$. Let r_{BE} denote the amount of residual resource to be given to best effort traffic, and ΔU_i, the utility gain by allocating resource r_i to QoS user i. Other notations remain the same as in the *HQ* and the *elastic allocations*. Given the total available resource r_{total}, the channel quality q_i and marginal utility function $u_i(r)$ for each user i, $\Re_{mix}$ can be obtained as follows.

1) Initialize $r_i \leftarrow 0$, $i = 1, 2, \cdots, n$; and $r_{BE} \leftarrow r_{total}$.
2) Sort all QoS users i in descending order of $\frac{U_{M_i} q_i}{r_{M_i}}$, and store them in the queue.
3) For each best effort user j, derive $u_j^{-1}(u)$ from $u_j(r)$;
   Find $u_\Sigma^{-1}(u)$ by summing up $u_j^{-1}(u)$ over all users j;
   Find $u_\Sigma(r)$, the inverted function of $u_\Sigma^{-1}(u)$.
4) If the queue is not empty, then
   pop out the QoS user i at the head of the queue;
   else go to Step (8).
5) For the popped user i:
   if $r_{BE} \geq \frac{r_{M_i}}{q_i}$, then $r_i = \frac{r_{M_i}}{q_i}$;
   else $r_i = 0$; go to Step (4).
6) $\Delta U_i = U_{M_i} - \int_{r_{BE} - r_i}^{r_{BE}} u_\Sigma(r) dr$.
7) If $(\Delta U_i > 0)$, then
   $r_{BE} = r_{BE} - r_i$; go to Step (4);
   else
   $r_i = 0$; go to Step (8);
8) If $u_\Sigma(r_{BE}) < u_j(0)$, then $r_j = u_j^{-1}(u_\Sigma(r_{BE}))$;
   else $r_j = 0$.

The allocation rule of this *mixed allocation* is to: 1) allocate resource to the first k QoS users at the sorted queue, and 2) then allocate the residual bandwidth (i.e., $r_{total} - r_{QoS}$) to all best effort users based on the *elastic allocation*. The value of k is determined based on the requirement that there is sufficient resource for this QoS user and the utility gain ΔU_k is positive (i.e., $r_{BE} - r_k > 0$) .

We have proved in Lemma 2.1 that the allocation problem for hard QoS traffic is NP-complete. It follows that the

Authorized licensed use limited to: National Taiwan University. Downloaded on January 20, 2009 at 23:15 from IEEE Xplore. Restrictions apply.

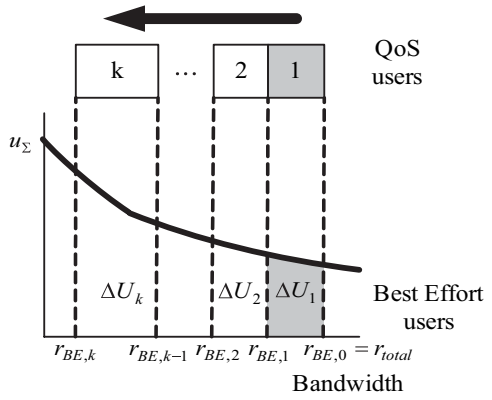


Fig. 5. An example of *mixed allocation*.

allocation problem for the co-existence of best effort and hard QoS traffic is also NP-complete. Again, the proposed *mixed allocation* algorithm can achieve a performance lower bounded by $U(\Re_{op}) - U_{max}$. When all QoS users have an identical utility function, the $mixed\ allocation$ can be proved to be optimal.

Theorem 2.7: For the *mixed allocation* problem, $U(\Re_{mix}) \geq U(\Re_{op}) - U_{\max}$, where $\Re_{mix}$ is the proposed solution, $\Re_{op}$ is the optimal solution, $U(x)$ is the total utility of all users for solution x, and $U_{\max}$ is the maximum utility value taken over all users, i.e., $U_{\max} = \max_{1 \leq k \leq n} \{U_{M_k}\}$.

Proof: This theorem can be proved by relaxing the constraint in the original problem as in Theorem 2.2. Therefore, the different in the total utility between $\Re_{mix}$ and the relaxed problem is bounded by U_{max}, i.e., $U(\Re_{mix}) \geq U(\Re_{op}) - U_{\max}$. ■

Theorem 2.8: The $mixed\ allocation$ $\Re_{mix}$ is an optimal allocation for traffic mixed with identical unit-step functions and arbitrary concave functions.

Proof: Consider the first k QoS users in the sorted queue, where k is the maximum possible value satisfying $\sum_{i=1}^{k} r_i \leq r_{total}$. The set of all possible optimal allocations is then: $\Re_0, \Re_1, \cdots, \Re_i, \cdots, \Re_k$, where $\Re_i$ is the allocation in which the first i QoS users are allocated resource in an amount of $r_j = r_M/q_j$, $j = 1, 2, \cdots, i$, and the residual bandwidth is all allocated to best effort users. The utility gain ΔU_i, $i = 1, 2, \cdots, k$, is expressed by $\Delta U_i = U(\Re_i) - U(\Re_{i-1}) = U_{M_i} - \int_{r_{BE}-r_i}^{r_{BE}} u_\Sigma(r)dr$.

Since QoS users are sorted in decreasing order of their q_i, this leads to that r_i is allocated in increasing order of q_i, resulting in $\Delta U_i \geq \Delta U_{i+1}$ (as shown in Fig. 5). Thus, the allocation $\Re_i$ in $[\Re_0, \Re_1, \Re_2, \cdots, \Re_i, \cdots, \Re_k]$ is the optimal allocation, where i is the largest value satisfying $\Delta U_i > 0$. ■

E. Implementation Issue

In practice, we can avoid such operations as inverting nonlinear functions, inverting the summations of nonlinear functions, and inverting non-linear functions by a series of summations and simple interpolations.

Suppose u_{max} is the maximum marginal utility value allowed by the system. We can partition the range of the values in the y-axis for any (continuous) marginal utility function, say $u(.)$, into values spaced equally apart, i.e., $u_k = u_{\max} - k\frac{u_{\max}}{M}$, where M is a tunable input parameter which determines the number of steps in total, and $k = 0, 1, \cdots, M$. We then let $r_k = u^{-1}(u_k)$, $k = 0, 1, \cdots, M$, in a one dimensional fixed length array $\Gamma = [r_0, r_1, r_2, ...r_M]$. Since $u(.)$ is a decreasing function, there is a one-to-one correspondence between r_k and u_k. This method eliminates the inversion process and can obtain any required values (i.e., $u(r)$ and $u^{-1}(u)$) by interpolation. For a given value r, the system can find the value k satisfying $r_k \leq r < r_{k+1}$. By interpolation, $u(r) \cong u_k - \frac{u_{\max}}{M} \cdot \frac{r - r_k}{r_{k+1} - r_k}$. Similarly, for any given value u, $u^{-1}(u) \cong r_k + (r_{k+1} - r_k) \cdot \frac{u - u_k}{u_{k+1} - u_k}$, where k is the value satisfying $u_{k+1} < u \leq u_k$. This approach can also reduce the computational overhead of the summation function $u_\Sigma(u)$, which can be calculated by $R_\Sigma = [\sum_{i=1}^{j} u_i^{-1}(u_0), \sum_{i=1}^{j} u_i^{-1}(u_1), \cdots, \sum_{i=1}^{j} u_i^{-1}(u_M)]$ given that there are j utility functions $(u_1, u_2, \cdots, u_j)$ to be aggregated. Likewise, the overhead for the integral of $u_\Sigma(u)$ can also be reduced.

In this way, the computational complexity of finding $u(r)$ and $u^{-1}(u)$ is only $O(M)$ because it takes at most M steps to find the value k for $r_k \leq k < r_{k+1}$. The complexity of the summation is $O(jM)$ because for $k = 0, 1, \cdots, M$, it takes j steps to calculate $\sum_{i=1}^{j} u_i^{-1}(u_k)$.

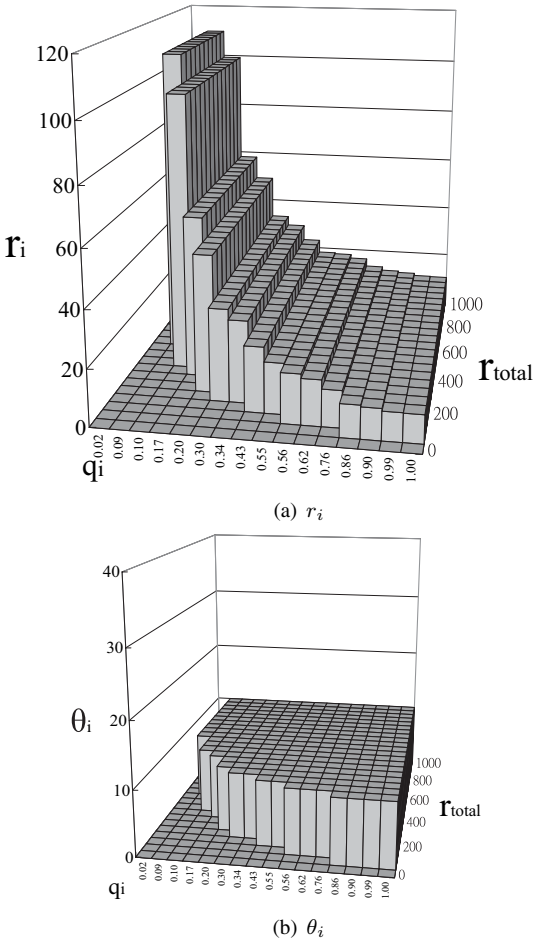
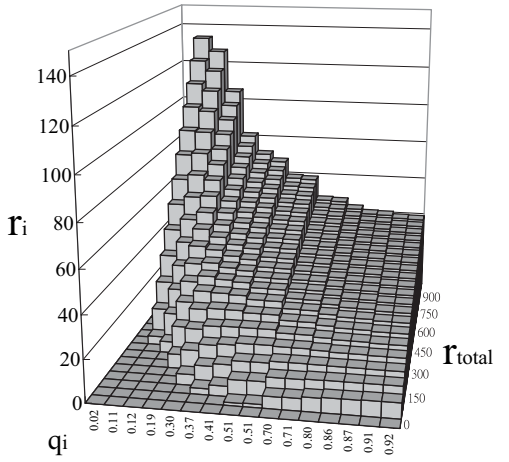
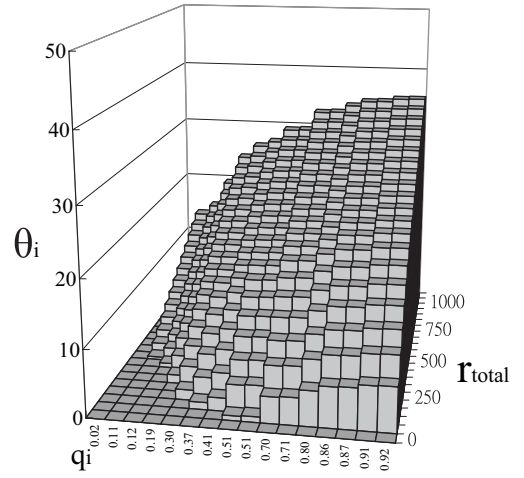


Fig. 6. Resource allocation in $\Re_{HQ} = \{r_1, r_2, \cdots r_n\}$.

Authorized licensed use limited to: National Taiwan University. Downloaded on January 20, 2009 at 23:15 from IEEE Xplore. Restrictions apply.



(a) r_i



(b) θ_i

Fig. 7. Resource allocation in $\Re_{elastic} = \{r_1, r_2, \cdots r_n\}$.

III. Performance Evaluation

In this section, we conduct simulations to evaluate the performance of our allocation algorithms. We consider QoS traffic, best effort traffic, and the con-existence of both. The simulation parameters are described as follows. For QoS traffic, the utility function is a unit-step function with $r_a = 10$ and $U_M = 1$, i.e., $U_{QoS}(r) = f_u(r-10)$; for best effort traffic, $U_{BE}(r) = 1 - e^{r/10}$. The value of q_i is randomly generated by a uniform distribution over [0, 1]. We then measure the distributions of r_i and θ_i under different values of r_{total}.

Fig. 6 shows the values of r_i and θ_i for QoS traffic in the $HQ\ allocation$. Fig. 6(a) depicts that when r_{total} is small, the system tends to allocate more resource to the users in better channel conditions; but as r_{total} increases, the amount of resource allocated to users is still fixed either at r_a or 0, because the utility function is a unit-step function. Fig. 6(b) shows that the actual amount of resource obtained by each user, if allocated, is identical, i.e., at their θ_i.

Fig. 7 shows the values of r_i and θ_i for best effort traffic in the $elastic\ allocation$. Fig. 7(a) depicts that when r_{total} is small, the system tends to allocate more resource to the users in better channel conditions. Since the utility function of best-effort traffic satisfies $u(r) > 0$ for all r, all users will demand resource all the time. Thus, the value of r_i for each user i increases with the value of r_{total}. Fig. 7(b) shows that users with larger q_i always result in larger θ_i.

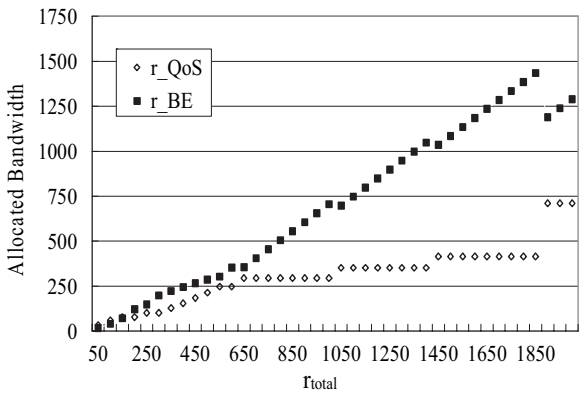


Fig. 8. An example of *mixed allocation*.

Fig. 8 shows the values of r_{QoS} and r_{BE}, as a function of r_{total} for mixed traffic in the *mixed allocation*. Since the QoS traffic is allocated resources in discrete amounts, when one more QoS user is served, the value of r_{BE} will drop. This effect becomes more pronounced when r_{total} is large, implying that the remaining unallocated QoS users have lower q_i and need more resources to compensate for their bad channel qualities. Due to the space limitations, we do not include the distributions of r_i and θ_i for the *mixed allocation* in this paper. The results can be found in [18], the characteristics of which are similar to those in Figs. 6 and 7.

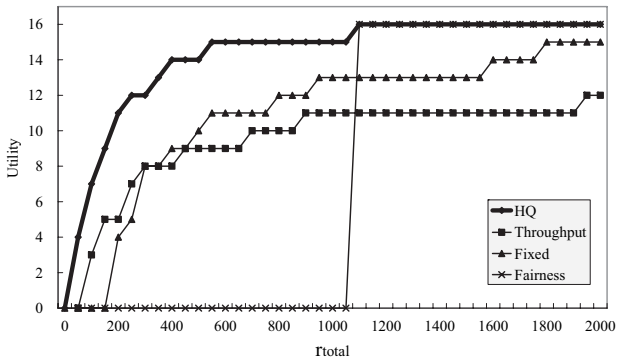
In Fig. 9, different resource allocation schemes are compared with the proposed allocation schemes. The comparison is based on the scheme propsoed in [4], which allocates radio resource proportionally based on factor q_i^α. Depending on the setting of the value α, the system can be tuned to work with different performance metrics. The curve denoted "throughput" is for $\alpha = 1$, which gives more resources to the users in better channel conditions, thereby leading to a larger system throughput. The curve denoted "fairness" is for $\alpha = -1$, giving all users an identical value of $\theta_i = r_i \cdot q_i$. The curve denoted "fixed" is for $\alpha = 0$, which provides the same amount of resource to all users. Note that the schemes proposed in [8-9] are the examples of the "fairness" scheme (i.e., $\alpha = -1$), and the GR+ scheme in [1] is an example of the "throughput" scheme.

Fig. 9(a) compares the proposed *HQ allocation* with different allocation schemes, and Fig. 9(b) compares the proposed *elastic allocation* with different allocation schemes. Note that the axis of r_{total} in Fig. 9(b) is in the logarithmic scale. The results show that the "throughput-first" scheme has a higher total utility when r_{total} is small, but the "fixed" allocation one is closer to the proposed scheme as r_{total} increases. Finally, when r_{total} becomes very large, the "fairness-first" scheme can achieve the highest utility.

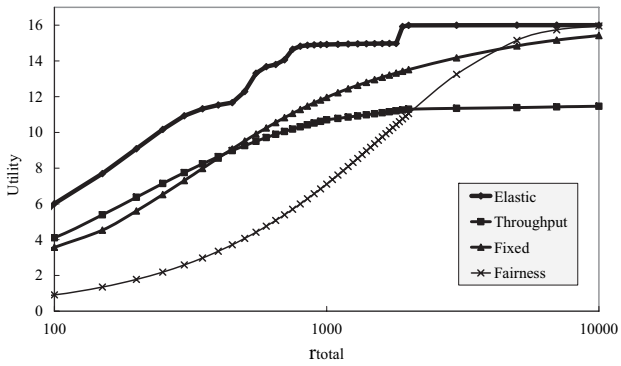
IV. Conclusion

In this paper, we study utility-based maximization for resource allocation in infrastructure-based wireless networks. We develop some essential theorems for utility-based resource

Authorized licensed use limited to: National Taiwan University. Downloaded on January 20, 2009 at 23:15 from IEEE Xplore. Restrictions apply.



(a) QoS traffic



(b) Best effort traffic

Fig. 9. Utility comparison with different resource allocation schemes.

management. Then, three polynomial time resource allocation algorithms are proposed for two types of utility functions. We prove that, in any case, the difference between the total utilities obtained by our proposed solutions and the optimal utility is bounded, and under certain conditions, all these three schemes can achieve the maximum total utility (i.e., optimal). From the simulation results, we find that different types of traffic require different kinds of schemes to achieve optimal allocation. In addition, when r_{total} is small, the system tends to allocate more resources to the users in better channel conditions, i.e., "throughput-oriented;" however, when r_{total} is abundant, the system becomes "fairness-oriented," meaning that even with the same traffic, the preference tendency between throughput and fairness can still differ. This leads us to conclude that existing channel-dependent-only resource schemes and schedulers cannot provide optimal allocation in wireless networks. To the best of our knowledge, this is the first work to address all the three issues together.

REFERENCES

[1] D. Angelini and M. Zorzi, "On the throughput and fairness performance of heterogeneous downlink packet traffic in a locally centralized CDMA/TDD system," in *Proc. IEEE VTC-Fall 2002*.

[2] A. Furuskar *et al.*, "Performance of WCDMA high speed packet data," in *Proc. IEEE VTC 2002-Spring*.

[3] Y. Cao, V. O. K. Li, and Z. Cao, "Scheduling delay-sensitive and best-effort traffic in wireless networks," in *Proc. IEEE ICC 2003*.

[4] Z. Jiang and and N. K. Shankaranarayana, "Channel quality dependent scheduling for flexible wireless resource management," in *Proc. IEEE Globecom 2001*.

[5] W.-T. Chen, K.-C. Shih, and J.-L. Chiang, "Flexible packet scheduling for quality of service provisioning in wireless networks," in *Proc. ICPADS*, Dec. 2002.

[6] M. Xiao, N. B. Shroff, and E. K. P. Chong, "A utility-based power-control scheme in wireless cellular systems," *IEEE/ACM Trans. Netw.*, vol. 11, no. 2, 2003.

[7] Y. Cao and V. O. K. Li, "Utility-oriented adaptive QoS and bandwidth allocation in wireless networks," in *Proc. IEEE ICC 2002*.

[8] G. Bianchi and A. T. Campbell, "A programmable MAC framework for utility-based adaptive quality of service support," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 2, pp. 244-255, 2000.

[9] R.-F. Liao and A. T. Campbell, "A utility-based approach to quantitative adaptation in wireless packet networks," *ACM WINET*, vol. 7, no. 5, pp. 541-557, Sept. 2001.

[10] X. Gao, T. Nandagopal, and V. Bharghavan, "Achieving application level fairness through utility-based wireless fair scheduling," in *Proc. IEEE Globecom 2001*.

[11] F. P. Kelly, "Charging and rate control for elastic traffic," *European Trans. Telecommun.*, Jan. 1997.

[12] V. A. Siris, B. Briscoe, and D. Songhurst, "Economic models for resource control in wireless networks," in *Proc. IEEE PIMRC* 2002, Lisbon, Portugal, Sept. 2002.

[13] V. A. Siris, "Resource control for elastic traffic in CDMA networks," in *Proc. IEEE MOBICOM 2002*.

[14] V. A. Siris and C. Courcoubetis, "Resource control for loss-sensitive traffic in CDMA networks," in *Proc. IEEE INFOCOM 2004*.

[15] P. Marbach and R. Berry, "Downlink resource allocation and pricing for wireless networks," in *Proc. IEEE INFOCOM 2002*.

[16] P. Liu, R. Berry, M. L. Honig, and S. Jordan, "Slow-rate utility-based resource allocation in wireless networks," in *Proc. IEEE Globecom*, Nov. 2002.

[17] S. Shenker, "Fundamental design issues for the future Internet," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1176-1188, 1995.

[18] W. H. Kuo and W. J. Liao, "Utility-based optimal resource allocation in wireless networks," in *Proc. IEEE Globecom 2005*.

[19] K.-D. Wu and W. Liao, "Flow allocation in multi-hop wireless networks: a cross-layer approach," to appear in *IEEE Trans. Wireless Commun.*, 2007.

[20] L. Chen, S. H. Low, and J. C. Doyle, "Joint congestion control and media access control design for wireless ad hoc networks," in *Proc. IEEE Infocom*, Miami, FL, March 2005.

Wen-Hsing Kuo was born on February 6, 1980, in Taichung, Taiwan. He received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, in 2002, where he is currently working toward the Ph.D. degree in the Graduate Institute of Electronics Engineering.

He joined Internet Research Laboratory, NTU, in 2002. His research interests include wireless resource management, network economics and 802.16 WiMAX networks.

Wanjiun Liao received the BS and MS degrees from National Chiao Tung University, Taiwan, in 1990 and 1992, respectively, and the Ph.D. degree in Electrical Engineering from the University of Southern California, Los Angeles, California, USA, in 1997. She joined the Department of Electrical Engineering, National Taiwan University (NTU), Taipei, Taiwan, as an Assistant Professor in 1997. Since August 2005, she has been a full professor. Her research interests include wireless networks, multimedia networks, and broadband access networks.

Dr. Liao is currently an Associate Editor of *IEEE Transactions on Wireless Communications* and *IEEE Transactions on Multimedia*. She served as the Technical Program Committee (TPC) chairs/co-chairs of many international conferences, including the Tutorial Co-Chair of IEEE Infocom 2004, the Technical Program Vice Chair of IEEE Globecom 2005 Symposium on Autonomous Networks, and the Technical Program Co-Chair of IEEE Globecom 2007 General Symposium. Dr. Liao has received many research awards. Papers she co-authored with her students received the Best Student Paper Award at the First IEEE International Conferences on Multimedia and Expo (ICME) in 2000, and the Best Paper Award at the First IEEE International Conferences on Communications, Circuits and Systems (ICCCAS) in 2002. Dr. Liao was the recipient of K. T. Li Young Researcher Award honored by ACM in 2003, and the recipient of Distinguished Research Award from National Science Council in Taiwan in 2006. She is a Senior member of IEEE.

Authorized licensed use limited to: National Taiwan University. Downloaded on January 20, 2009 at 23:15 from IEEE Xplore. Restrictions apply.