Utility-based Radio Resource Allocation for QoS Traffic in Wireless Networks

Wen-Hsing Kuo and Wanjiun Liao

***Abstract*—In this paper, we study utility-based resource allocation for soft QoS traffic in infrastructure-based wireless networks. Soft QoS traffic here refers to the traffic which demands certain amount of bandwidth for normal operation but allows some flexibility when the given bandwidth is close to the preferred value. The resource requirement of soft QoS traffic can be described with sigmoid utility function. Our objective is to maximize the total utility of all soft QoS flows without going through a wireless bidding process. We develop essential theorems as the design guidelines for this problem, and then propose a sub-optimal, polynomial time solution based on the developed theorems. We prove that the difference in the performance of our mechanism and the optimal solution is bounded. The performance of the proposed solution is evaluated via simulations. The results show that our solution can adapt to any types of soft QoS flows. Specifically, it acts like a hard QoS system and allocates resource in a fairness-oriented manner when the utility functions of flows are unit-step functions; on the other hand, when the utility functions are concave, it behaves like a best effort system and allocates resource in a throughput-oriented way.**

***Index Terms*—Utility optimization, resource allocation, soft QoS, wireless networks.**

I. Introduction

IN this paper, we study the utility-based resource allocation problem in infrastructure-based wireless networks. In wireless networks, radio resource is limited and scarce, and the channel quality of each user may vary over time. Given the channel condition of each user and the total available system resource, the amount of bandwidth assigned to each user may be guided by such performance metrics as throughput and fairness [1] or according to the type of traffic [2]. "Throughput" and "fairness," however, are conflicting performance metrics in scheduling. In this paper, we avoid such a "throughput-fairness" trade-off dilemma, and focus on "user satisfaction" for radio resource allocation.

The degree of user satisfaction can be described by the utility function of the traffic under consideration. A utility function $U(r)$ is a non-decreasing function with respect to the amount of allocated resource r. The more the resource is allocated, the more the user is satisfied. The wireless resource r here may refer to timeslots or radio frequency occupied by a single user. The marginal utility function $u(r)$ is the derivative of the utility function $U(r)$ with respect to r, i.e., $u(r) = \frac{dU(r)}{dr}$. The exact expression of a utility function may depend on the type of traffic, and can be derived from some metrics that reflect the user's perception or the content quality, such as Perceived Signal-Noise Ratio (PSNR) [3] or Mean Opinion Score (MOS) [4] of a multimedia stream. In this paper, we leave the work of finding utility functions to psychologists and economists, and focus on maximizing the total utility of given utility functions in the system.

Utility-based solutions have been widely used in wireless resource management. For example, a utility-based bandwidth allocation scheme is proposed in [5-6]; utility-based fairness allocation schemes are presented in [7-8], and a utility-based scheduler is proposed in [9]. Utility-based approaches have also been widely used for bandwidth pricing in wireless networks [10-12]. The idea of these schemes is to associate a price with each unit of radio resource and let each player maximize its welfare based on a bidding process. Bidding schemes, while useful for Internet pricing and congestion control [13], are less practical and feasible for centralized wireless networks [14]. This is because in wireless environments, the type of traffic, the number of flows, and channel conditions are time-varying. Consequently, the bidding process would be very costly as users would have to repeatedly exchange messages in a real-time bidding. We show in [14] that in the infrastructure-based wireless network, both flow information and channel condition are accessible at the base station. It follows that optimizing the total utility of all flows at the base station can achieve the same objective as in a bidding scheme but in a simpler way. Therefore, it is more desirable to implement an optimal radio resource allocation mechanism in centralized wireless schedulers without a bidding process.

In this paper, we attempt to maximize the total utility of all flows at the base station while eliminating the bidding process. We focus on soft QoS traffic. The soft QoS traffic here refers to the traffic which requires a preferred amount of bandwidth for normal operation but can tolerate certain flexibility when the given amount of resource is close to the traffic's preferred value. The bandwidth requirement of soft QoS traffic can be described by the sigmoid utility function with respect to the bandwidth resource [15]. The sigmoid utility function is a utility function whose $U(r) > 0$ and $u(r) > 0$, for all r, as shown in Fig. 1, where rc denotes the preferable amount of resource for the soft QoS traffic and U_c denotes the achieved utility value with a given amount of resource rc. We observe

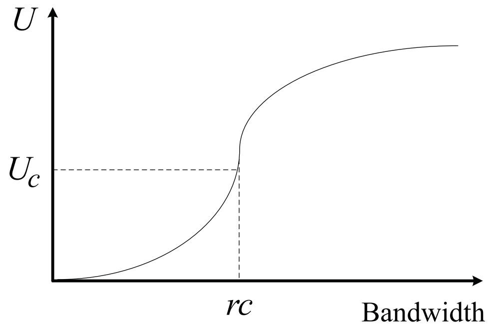
Manuscript received January 30, 2006; revised May 4, 2007; accepted June 22, 2007. The associate editor coordinating the review of this letter and approving it for publication was S. Shen. This work was supported by National Science Council (NSC), Taiwan, under a Center Excellence Grant NSC96-2752-E-002-006-PAE, and under Grant Number NSC96-2628-E-002-004-MY3.

W.-H. Kuo was with the Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan. He is now with the Department of Electrical Engineering, Yuan-Ze University (YZU), Taoyuan, Taiwan.

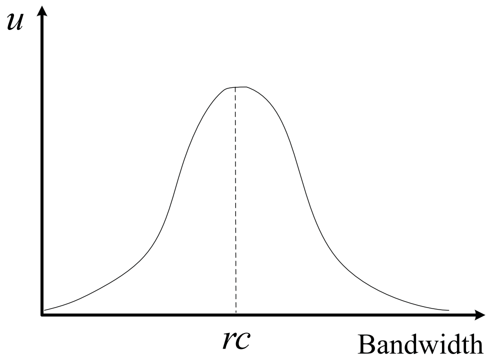
W. Liao is with the Department of Electrical Engineering and the Graduate Institute of Communication Engineering, National Taiwan University, Taipei, Taiwan (e-mail: wjliao@ntu.edu.tw).

Digital Object Identifier 10.1109/TWC.2008.070116.

1536-1276/08$25.00 © 2008 IEEE



(a) Utility function $U(r)$.



(b) Marginal utility function $u(r)$.

Fig. 1. The sigmoid utility function for soft QoS traffic.

that if $r < rc$, $u'(r) > 0$; if $r = rc$, $u'(r) = 0$; otherwise, $u'(r) < 0$. It can be interpreted as follows. When the amount of resource r is given insufficiently, it is useless for real-time traffic; as r approaches rc, the flow is gradually operational, and thus the marginal utility increases dramatically. Once the allocated resource r has exceeded rc, allocating more resource may not be helpful for operation, and thus the marginal utility drops hence forth.

In [14], we proposes a resource allocation algorithm which achieves utility maximization for both hard QoS and best effort traffic, where hard QoS traffic refers to the traffic with strict demand on resource requirement and best effort traffic refers to the traffic which does not have the minimum bandwidth requirement. In this paper, we extend the discussion to soft QoS traffic in centralized wireless networks (i.e., with base stations (BSs)). Allocating resource to soft QoS traffic is very challenging since unlike hard QoS traffic, which allocates resource discretely, soft QoS traffic demands certain amount of bandwidth for normal operation, but allows some flexibility in resource allocation, i.e., continuously. To better understand the characteristics of the optimal solution to resource allocation problem for soft QoS traffic, we develop some theoretical results as the design guidelines for this problem. Based on the developed theorems, we propose an efficient algorithm which assigns wireless resource to soft QoS traffic in polynomial time. The proposed algorithm considers the traffic type, the total available resource, and all users' channel qualities, rather than just considering the channel quality or traffic type as in most existing work [5-9]. More importantly, our algorithm can adapt to any kind of traffic with sigmoid utility functions. Specifically, when the slope of the utility function is very steep, it acts like a hard QoS system and allocates resource to flows in a fairness-oriented manner; on the other hand, when the slope of the utility function is relatively flat, it behaves like a best effort system and allocates resource in a throughput-oriented way. We also show that our algorithm can achieve a very tight bound to the optimum.

The rest of the paper is organized as follows. In Sec. II, the theoretical foundation for this problem is developed, the design guidelines for supporting soft QoS traffic are described, and a wireless resource allocation scheme is proposed. In Sec. III, the performance of the proposed scheme is evaluated via simulations. Finally, the paper is concluded in Sec. IV.

II. Radio Resource Allocation Based on Utility Maximization

A. System Model and Problem Statement

Consider a base station with a set of soft QoS flows, denoted by Γ, in the wireless network. The resource requirement of each flow is described with a sigmoid utility function $U(\cdot)$. Let r_{total} denote the total amount of radio resource available at the base station; let r_i denote the amount of resource allocated to flow i, $i \in \Gamma$, and q_i, channel quality of flow i, where $0 \leq q_i \leq 1$. The smaller the value of q_i, the worse the channel quality. Radio resource here is defined as the resource used to transmit data, and is allocated by the base-station's controller. Such resource can be the number of time-slots, the number of codes, and so on, depending on the type of the wireless network in use.

Advances in wireless networks allow the upper layers to access such information as user channel conditions, thus facilitating adaptive modulation and coding (AMC) schemes. As a result, it is reasonable to assume that 1) each base station knows about the channel conditions of all users and 2) users under different channel conditions can transmit data at different data rates. Let q_i denote the channel quality parameter, which refers to the ratio of the actual amount of resource received by the user to the amount of resource allocated by the system to the user, and thus it is in a range of [0, 1]. Given the channel quality q_i of flow i, the amount of resource actually beneficial to flow i is given by $\theta_i = r_i \cdot q_i$. Therefore, the utility function of flow i can be expressed as $U_i(\cdot) = U(r_i q_i)$, where $U(\cdot)$ is the utility function of the traffic under consideration and $U_i(\cdot)$ is the utility function for the type of traffic described by $U(\cdot)$ but taking into account the channel quality of flow i. To better distinguish $U(\cdot)$ from $U_i(\cdot)$, we refer to $U(\cdot)$ as the *traffic* utility function, and $U_i(\cdot)$, the flow utility function in the rest of the paper. The marginal utility function of $U_i(\cdot)$, denoted by $u(\cdot)$, is defined by $u_i(r_i) = \frac{dU(r_i q_i)}{dr_i} = q_i \cdot u(q_i \cdot r_i)$. Suppose that all flows in the system have the same traffic utility function $U(\cdot)$. Considering the channel condition of each flow may not be identical, we denote the preferable amount of resource for flow i by rc_i. For each flow i, $i \in \Gamma$, $U_i(rc_i) = U(rc) = U_c$, and thus $rc_i = \frac{rc}{q_i}$. Since a user who is allocated more radio resource is happier, any utility function must be an increasing function, i.e., $U(r_i) \leq U(r_j)$ holds when $r_i \leq r_j$ or $u(r) \geq 0$ for all r.

In this paper, we propose a wireless resource allocation scheme which maximizes the total utility $\sum_{i\in\Gamma} U_i(r_i)$, subject to $\sum_{i\in\Gamma} r_i \leq r_{total}$ and $\forall r_i \geq 0,\ i \in \Gamma$. We claim that an allocation solution $\Re^* = \{r_i, i \in \Gamma\}$ is optimal if for any allocation $\Re' = \{r_j', j \in \Gamma\}$, $U(\Re^*) \geq U(\Re')$, where $U(\Re^*) = \sum_{i\in\Gamma} U_i(r_i)$ and $U(\Re') = \sum_{j\in\Gamma} U_j(r_j')$. The optimal solution may not be unique.

Note that in a wireless network, the channel condition of each mobile node may vary over time. To reflect the latest channel condition, each base station may execute this proposed algorithm periodically or when some significant change to the wireless channel is detected. Since this algorithm operates in polynomial time, its computational overhead is acceptable and can be executed frequently.

B. Design Guidelines

For each sigmoid function $U(r)$, $U(r) > 0$ holds for all $r > 0$. This implies that to maximize the total utility of the system, the available radio resource at the base station must all be assigned to flows, i.e., $\sum_{i\in\Gamma} r_i \leq r_{total}$. Since sigmoid utility functions are continuous, we prove that in an optimal allocation $\Re^* = \{r_f, f \in \Gamma\}$, every flow f whose $r_f > 0$ (called allocated flow) has an identical marginal utility (Lemma 2.1). We further prove that in an optimal allocation $\Re^* = \{r_f, f \in \Gamma\}$, there is at most one flow, say, f which satisifies $r_f > 0$ and $u'_f(r_f) > 0$, where $u'_f(r_f) = \frac{du_f(r_f)}{dr_f}$ (Lemma 2.2).

Lemma 2.1: For an optimal allocation $\Re^* = \{r_f, f \in \Gamma\}$, if $r_i > 0$ and $r_j > 0$, $i, j \in \Gamma$, then $u_i(r_i) = u_j(r_j)$.

Proof: We can rewrite the optimization problem into the Lagrange multiplier form by letting constraint $g(r_1, r_2, ...r_n) = \sum_{i=1}^{n} r_i$ and objective $f(r_1, r_2, ...r_n) = \sum_{i=1}^{n} U_i(r_i)$. In optimum, the gradients of $[f(r_1, r_2, ...r_n) + \lambda(g(r_1, r_2, ...r_n) - c)] = 0$. Thus, for all $i \in \Gamma$, $\frac{d}{dr_i}[f(r_1, r_2, ...r_n) + \lambda(g(r_1, r_2, ...r_n) - c)] = u_i(r_i) + \lambda = 0$. Since $\lambda = -u_i(r_i)$ for all $i \in \Gamma$, all allocated users' $u_i(r_i)$ are identical to $-\lambda$. ■

Lemma 2.2: For an optimal allocation $\Re^* = \{r_f, f \in \Gamma\}$, there is at most one flow, say, f whose $r_f > 0$ and $u'_f(r_f) > 0$, where $u'_f(r_f) = \frac{du_f(r_f)}{dr_f}$.

Proof: This lemma can be proved by contradiction. Assume that more than one flow f satisfies $r_f > 0$ in $\Re^*$ and $u_f'(r_f) > 0$. Then, we can find two flows i and j in $\Re^*$ whose $u_i'(r_i) > 0$ and $u_j'(r_j) > 0$. Since $u(r)$ is continuous, we can always find a real number Δr, which satisfies both $\int_{r_i-\Delta r}^{r_i} u_i(r)dr < u_i(r) \cdot \Delta r$ and $\int_{r_j}^{r_j+\Delta r} u_j(r)dr > u_j(r_j) \cdot \Delta r$. Consider another resource allocation solution $\Re' = \{r_f', f \in \Gamma\}$ in which flows i and j are allocated an amount of $r_i - \Delta r$ and $r_j + \Delta r$, respectively. Since the sigmoid function is continuous and increasing, the difference in the total utilities of $\Re'$ and $\Re^*$ can be expressed by:

$$\begin{aligned}
&U(\Re') - U(\Re^*)\\
&= [U_i(r_i - \Delta r) + U_j(r_j + \Delta r)] - [U_i(r_i) + U_j(r_j)]\\
&= [U_j(r_j + \Delta r) - U_j(r_j)] - [U_i(r_i) - U_i(r_i - \Delta r)]\\
&= \int_{r_j}^{r_j+\Delta r} u_j(r)dr - \int_{r_i-\Delta r}^{r_i} u_i(r)dr
\end{aligned}$$

Since $\int_{r_j}^{r_j+\Delta r} u_j(r)dr > u_j(r_j) \cdot \Delta r > \int_{r_i-\Delta r}^{r_i} u_i(r)dr$, we obtain $U(\Re') - U(\Re^*) > 0$, violating the requirement that $\Re^*$ is optimal. Hence, for all allocated flows in $\Re^*$, there exists at most one such flow f whose $u'_f(r_f) > 0$. ■

Lemma 2.3: Consider two soft QoS flows i and j whose traffic utility functions are identical (i.e., with the same $U(.)$). For an optimal allocation $\Re^* = \{r_f, f \in \Gamma\}$, if $q_i \geq q_j$, $i, j \in \Gamma$, then inequality "$\theta_i \geq \theta_j$" must always hold.

Proof: This lemma can be proved by contradiction. We assume that there exists two flows i, j in $\Re^*$ satisfying $q_i \geq q_j$ and $\theta_i < \theta_j$. Then, θ_j can be expressed by $\theta_i + \Delta\theta$, where $\Delta r > 0$. It follows that the total amount of resource given to these two flows, i.e., $r_i + r_j$, can be expressed in terms of θ and q by $r_i + r_j = \frac{\theta_i}{q_i} + \frac{\theta_j}{q_j} = \frac{\theta_i}{q_i} + \frac{\theta_i + \Delta\theta}{q_j} = \theta_i(\frac{q_i+q_j}{q_iq_j}) + \frac{\Delta\theta}{q_j}$.

We can find another allocation $\Re' = \{r_f', f \in \Gamma\}$ in which the only difference from $\Re^*$ is that flows i and j are allocated resource differently. In $\Re'$, we let $\theta_i' = \theta_i + \Delta\theta$ and $\theta_j' = \theta_i + (1 - \frac{q_j}{q_i})\Delta\theta$. The total amount of resource allocated to flows i and j in $\Re'$ is given by $\frac{\theta_i'}{q_i} + \frac{\theta_j'}{q_j} = \theta_i(\frac{q_i+q_j}{q_iq_j}) + \frac{\Delta\theta}{q_j}$, equal to $r_i + r_j$ in $\Re^*$.

The aggregate utility contributed by flows i and j in $\Re'$ is given by $U(\theta_i') + U(\theta_j') = U(\theta_i + \Delta\theta) + U(\theta_i + \Delta\theta(1 - \frac{q_j}{q_i}))$, which is larger than $U(r_i) + U(r_j)$ (i.e., $U(\theta_i + \Delta\theta) + U(\theta_i)$). This leads to $U(\Re') > U(\Re^*)$, violating the assumption that $\Re^*$ is optimal. Thus, for any optimal allocation, if $q_i \geq q_j$, inequality $\theta_i \geq \theta_j$ must always hold. ■

Theorem 2.4: An optimal solution to this soft QoS allocation problem, denoted by $\Re^* = \{r_f, f \in \Gamma\}$, must satisfy (1) to (4).

$$\sum_{\forall i\in\Gamma} r_i = r_{total}; \tag{1}$$

$$\forall i, j \in \Gamma,\ u_i(r_i) = u_j(r_j), \text{ if } r_i > 0 \text{ and } r_j > 0; \tag{2}$$

$$\forall i, j \in \Gamma,\ q_i \geq q_j, \text{ if } r_i > 0 \text{ and } r_j = 0; \tag{3}$$

$$\forall i \in \Gamma, \text{ there is at most one flow } i \text{ whose } r_i > 0 \text{ and } u_i'(r_i) > 0. \tag{4}$$

Since an optimal solution to this problem is very hard to find and may be dependent on the channel qualities and utility functions of flows, we relax the constraint (4) as follows to further reduce the computational complexity.

Definition 2.5: A solution which satisfies (1), (2), (3), and (5), denoted by $\Re^\alpha = \{r_f, f \in \Gamma\}$, is a sub-optimal solution to this allocation problem.

$$\forall i \in \Gamma,\ u_i'(r_i) < 0 \text{ if } r_i > 0 \tag{5}$$

Since each allocated flow i in $\Re^\alpha$ satisfies $u_i'(r_i) < 0$, this implies that the amount of resource r_i allocated to flow i in $\Re^\alpha$ must exceed its preferable amount rc_i. Therefore, our objective is to find an allocation whose total utility is the highest among all sub-optimal solutions. This obtained solution, while sub-optimal, can greatly reduce the computational complexity. We will prove in the next section that this sub-optimal approach is tightly bounded to the optimum.

Next, we develop some lemmas to help the design of the algorithm. For simplicity, we assume that all flows in the queue are sorted in decreasing order of their channel qualities. Let $\Re_j^\alpha$ denote a sub-optimal allocation which allocates resource to a total of j flows. We claim that the solution $\Re_j^\alpha$ is unique. This leads to the following lemma.

Lemma 2.6: $\Re_j^\alpha$ is the one and the only one sub-optimal allocation in which a total of j flows are allocated resource.

Proof: We prove this lemma by contradiction. Assume that there are more than one allocation, say $\Re_{j_1}$ and $\Re_{j_2}$, which satisfies the same constraints as $\Re_j^\alpha$. According to (5), for any allocated flow i, $u_i^{'}(r_i) < 0$. Therefore, for all i, if its has a different marginal value from the others, it must has a different value of r_i. From (2), all allocated flows in $\Re_{j_1}$ (or in $\Re_{j_2}$) must have the same marginal value; from (3), the sets of allocated flows in $\Re_{j_1}$ and in $\Re_{j_2}$ are identical since they both have the same number of allocated flows, i.e., j. These facts dictate that given a marginal value and a total of j allocated flows, there exists only one allocation and thus only one total amount of allocated resource, because if these two allocations were different, they would have different marginal values. Since $\Re_{j_1}$ and $\Re_{j_2}$ are different by assumption, they each have a different marginal value for their respective allocated flows. However, a different marginal value leads to a different total amount of allocated resource. This violates the fact that the total amount of resource allocated to flows by $\Re_{j_1}$ or $\Re_{j_2}$ is equal to r_{total}, since both of them must satisfy (1). Therefore, it can be concluded that $\Re_{j_1}$ and $\Re_{j_2}$ are identical, and that $\Re_{j_1} = \Re_{j_2} = \Re_j^\alpha$. ■

Next we discuss the amount of resource allocated to each flow in $\Re_j^\alpha$. From (5), we obtain that in $\Re_j^\alpha$, $j = 1, 2, \cdots, n$, inequality "$r_i > rc_i$" holds for each allocated flow i, $i = 1, 2, \cdots, j$, where rc_i is given and determined by the traffic type (i.e., its traffic utility function $U(.)$) of flow i. Consider the marginal sigmoid utility function $u_i(r_i)$ of flow i, $i = 1, 2, \cdots, j$. Let $\hat{u}_i(r) = u_i(r - rc_i)$, $r > rc_i$. $\hat{u}_i(.)$ is a decreasing function with respect to the amount of resource r. Let $\hat{u}_i^{-1}(.)$ be the inverse function of $\hat{u}_i(.)$. We sum up the inverse functions $\hat{u}_i^{-1}(.)$ of all flows i, $i = 1, 2, \cdots, j$, i.e., $\hat{u}_{\Sigma_j}^{-1}(.) = \sum_{i=1}^{j} \hat{u}_i^{-1}(.)$, and find the aggregate marginal utility function $\hat{u}_{\Sigma_j}(.)$ by inversing $\hat{u}_{\Sigma_j}^{-1}(.)$. Denote the residual bandwidth after j allocations by rr_j, i.e., $rr_j = r_{total} - \sum_{i=1}^{j} rc_i$. We then find the aggregate utility $u_{a,j} = \hat{u}_{\Sigma_j}(rr_j)$. Based on $u_{a,j}$ and $\hat{u}_j^{-1}(.)$, we can obtain the amount of resource allocated to each flow, which is given by $r_i = rc_i + \hat{u}_i^{-1}(u_{a,j})$ for $i = 1, 2, \cdots, j$, and $r_i = 0$ otherwise. We claim that all flows i in $\Re_j^\alpha = \{r_i, i \in \Gamma\}$ must be allocated in such a way that $r_i = rc_i + \hat{u}_i^{-1}(u_{a,j})$, $0 \le i \le j$, and $r_i = 0$ otherwise.

Theorem 2.7: In $\Re_j^\alpha = \{r_i, i \in \Gamma\}$, all flows $i, \forall i \in \Gamma$, must satisfy (6).

$$r_i = \begin{cases} rc_i + \hat{u}_i^{-1}(u_{a,j}), 0 \le i \le j \\ 0, \text{otherwise} \end{cases} \tag{6}$$

Proof: We prove that an allocation (say $\Re_j^{'}$) in which the amount of resource allocated to each flow i satisfies (6) must be $\Re_j^\alpha$. We first show that the allocation $\Re_j^{'}$ satisfies (1), (2), (3), and (5). Since $\sum_{i=1}^{j} r_i = \sum_{i=1}^{j} [rc_i + \hat{u}_i^{-1}(\hat{u}_{\Sigma_j}(rr_j))] = \sum_{i=1}^{j} (rc_j) + rr_j = r_{total}$, it satisfies (1); since for $i = 1, 2, \cdots, j$, $u_i(r_i) = \hat{u}_i(\hat{u}_i^{-1}(u_{\Sigma_j}(rr_j))) = u_{\Sigma_j}(rr_j) = u_{a,j}$, all allocated flows must have the same marginal value (i.e., $u_{a,j}$), satisfying (2); since flows are sorted by their channel qualities and only the first j flows with the best channel qualities are allocated, we obtain (3); for all allocated flows i, the amount of allocated resource r_i is larger than rc_i, and therefore $u_i'(r_i) \le 0$, satisfying (5). Thus, we can conclude that this allocation $\Re_j^{'}$ must be a sub-optimal allocation. Since such sub-optimal allocation is unique (from Lemma 2.6), $\Re_j^{'}$ must be equal to $\Re_j^\alpha$. Thus, the resource allocated to each flow i in $\Re_j^\alpha$ must satisfy (6). ■

Let ΔU_j^α denote the difference between the total utilities of $\Re_j^\alpha$ and $\Re_{j-1}^\alpha$, i.e., $\Delta U_j^\alpha = U(\Re_j^\alpha) - U(\Re_{j-1}^\alpha)$. We show next that the value of ΔU_j^α decreases as the value of j increases.

Lemma 2.8: For $j = 1, 2, \cdots, n$, $\Delta U_j^\alpha > \Delta U_{j+1}^\alpha$, where $\Delta U_j^\alpha = U(\Re_j^\alpha) - U(\Re_{j-1}^\alpha)$.

Proof: Let $\Re_j^\alpha = \{r_i^{\alpha,j}, i \in \Gamma\}$, $j = 1, 2, \cdots, n$. ΔU_j^α can be expressed as follows.

$$\begin{aligned}
\Delta U_j^\alpha &= U(\Re_j^\alpha) - U(\Re_{j-1}^\alpha) \\
&= \sum_{i=1}^{j} U_i(r_i^{\alpha,j}) - \sum_{i=1}^{j-1} U_i(r_i^{\alpha,j-1}) \\
&= U_j(r_j^{\alpha,j}) + \sum_{i=1}^{j-1} U_i(r_i^{\alpha,j}) - \sum_{i=1}^{j-1} U_i(r_i^{\alpha,j-1}) \\
&= U_j(r_j^{\alpha,j}) - [\sum_{i=1}^{j-1} U_i(r_i^{\alpha,j-1}) - \sum_{i=1}^{j-1} U_i(r_i^{\alpha,j})] \\
&= U_j(r_j^{\alpha,j}) - \{(j-1)U_c + \hat{U}_{\Sigma(j-1)}(rr_{j-1}^\alpha) - [(j-1)U_c + \hat{U}_{\Sigma(j-1)}(rr_{j-1}^\alpha - r_j^{\alpha,j})]\} \\
&= U_j(r_j^{\alpha,j}) - \int_{rr_{j-1}^\alpha - r_j^{\alpha,j}}^{rr_{j-1}^\alpha} \hat{u}_{\Sigma(j-1)}(r)dr \ ,
\end{aligned}$$

where rr_j^α is the residual bandwidth after j allocations by $\Re_j^\alpha$, and $\hat{U}_{\Sigma_j}(rr_j^\alpha) + j \cdot U_c = \sum_{i=1}^{j} U_i(r_i^{\alpha,j})$.

Since for all j, both $\hat{u}_{\Sigma_j}(\cdot)$ and $\hat{u}_j(\cdot)$ are decreasing functions, we know that for all $a > b$,

$\hat{u}_{\Sigma_j}(a) \cdot (a-b) \ge \int_b^a \hat{u}_{\Sigma_j}(r)dr \ge \hat{u}_{\Sigma_j}(b) \cdot (a-b)$, and

$U_j(a) - U_j(b) = \int_b^a u_j(r)dr \ge u_j(b) \cdot (a-b)$.

We can further obtain that

$$\begin{aligned}
\Delta U_j^\alpha &= U(\Re_j^\alpha) - U(\Re_{j-1}^\alpha) \\
&= U_j(r_j^{\alpha,j}) - \int_{rr_{j-1}^\alpha - r_j^{\alpha,j}}^{rr_{j-1}^\alpha} \hat{u}_{\Sigma_{j-1}}(r)dr \\
&\ge U_j(r_j^{\alpha,j}) - \hat{u}_{\Sigma_{j-1}}(rr_{j-1}^\alpha) \cdot (r_j^{\alpha,j}) \ge U_j(r_j^{\alpha,j}) - u_{a,j-1} \cdot r_j^{\alpha,j} \\
&\ge U_j(r_j^{\alpha,j}) - u_{a,j} \cdot r_j^{\alpha,j} \\
&= U_j(r_{j+1}^{\alpha,j+1}) - u_{a,j} \cdot r_j^{\alpha,j} - (U_j(r_{j+1}^{\alpha,j+1}) - U_j(r_j^{\alpha,j})) \\
&\ge U_j(r_{j+1}^{\alpha,j+1}) - u_{a,j} \cdot r_j^{\alpha,j} - U_j(r_j^{\alpha,j}) \cdot (r_{j+1}^{\alpha,j+1} - r_j^{\alpha,j}) \\
&= U_j(r_{j+1}^{\alpha,j+1}) - u_{a,j} \cdot r_j^{\alpha,j} - u_{a,j} \cdot (r_{j+1}^{\alpha,j+1} - r_j^{\alpha,j}) \\
&= U_j(r_{j+1}^{\alpha,j+1}) - u_{a,j} \cdot r_{j+1}^{\alpha,j+1}
\end{aligned}$$

$= U_j(r_{j+1}^{\alpha,j+1}) - \hat{u}_{\Sigma_j}(rr_j^{\alpha}) \cdot r_{j+1}^{\alpha,j+1} \geq U_j(r_{j+1}^{\alpha,j+1}) - \int_{rr_j^{\alpha}-r_{j+1}^{\alpha,j+1}}^{rr_j^{\alpha}} \hat{u}_{\Sigma_j}(r)dr$
$= \Delta U_{j+1}^{\alpha}$.

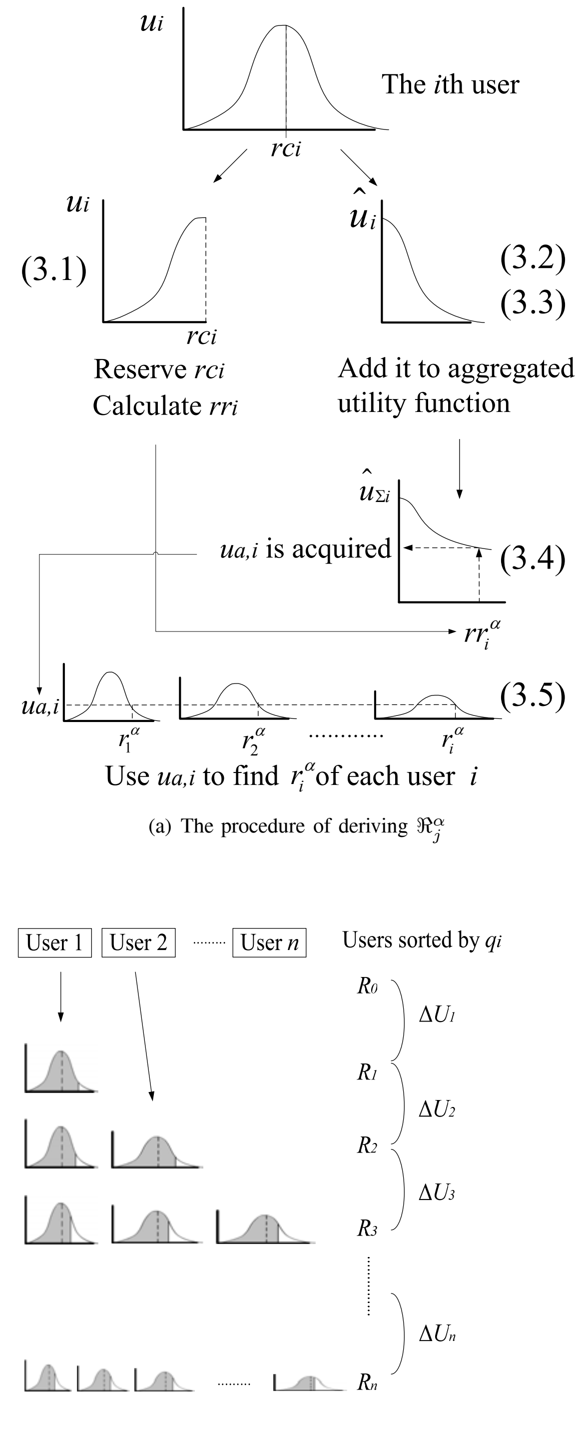
Therefore, for $j = 1, 2, \cdots, n$, $\Delta U_j^{\alpha} > \Delta U_{j+1}^{\alpha}$. ■

C. Resource Allocation Algorithm

Based on the lemmas and theorems developed in Sec. II-B, we propose a heuristic, called Utility-based allocation for Soft QoS (USQ), which finds an allocation with the highest total utility among all sub-optimal allocations (i.e., satisfying Definition 2.5). Suppose that there are n flows sorted in decreasing order of their channel qualities in the system, i.e., $|\Gamma| = n$. We determine the amount of resource to be allocated to each flow in the queue such that the total utility of the target allocation $\Re_{sigmoid}$ is maximized. Since an optimal solution to this problem is difficult to find, the target allocation $\Re_{sigmoid}$ we determine may be sub-optimal in order to reduce the computational overhead. In other words, we attempt to find all sub-optimal allocations $\Re_1^{\alpha}, \Re_2^{\alpha}, \cdots, \Re_n^{\alpha}$, each of which is unique (Lemma 2.6), and then determine $\Re_K^{\alpha}$ as the target allocation $\Re_{sigmoid}$, where $U(\Re_k^{\alpha}) = \max_{x=1,\cdots,n} U(\Re_x^{\alpha})$. We will prove shortly that the solution $\Re_{sigmoid}$ is tightly bounded to the optimum.

In Theorem 2.7, we have shown that the amount of resource allocated to each flow i by $\Re_j^{\alpha}$ must satisfy (6). Accounting for the number of flows being allocated resource in the system, we can determine the sub-optimal allocation $\Re_j^{\alpha}$ for j flows, $j = 1, 2, \cdots, n$. In Lemma 2.8, we have further shown that ΔU_j^{α} decreases as j increases given that all flows' traffic utility functions are identical. It follows that when $\Delta U_j^{\alpha} < 0$ and $\Delta U_{j-1}^{\alpha} > 0$ (i.e., $0 < \Delta U_{j-1}^{\alpha} < \cdots < \Delta U_2^{\alpha} < \Delta U_1^{\alpha}$ and $\Delta U_n^{\alpha} < \cdots < \Delta U_{j+1}^{\alpha} < \Delta U_j^{\alpha} < 0$), $\Re_{j-1}^{\alpha}$ has the highest total utility among all sub-optimal allocations $\Re_1^{\alpha}, \Re_2^{\alpha}, \cdots, \Re_n^{\alpha}$. In summary, our algorithm first sorts all flows by their channel qualities, and then finds all $\Re_j^{\alpha}$ and ΔU_j^{α}, $j = 1, 2, \cdots, n$. Once "$\Delta U_j < 0$" is determined, the algorithm stops and returns $\Re_{j-1}^{\alpha}$ as the target allocation $\Re_{sigmoid}$. Fig. 2(a) summarizes how $\Re_j^{\alpha}$ is obtained and Fig. 2(b) illustrates the relation between $\Re_j^{\alpha}$ and ΔU_j^{α}.

The detailed algorithm is summarized in Table I. Note that in Table I, since some steps are repeated as j increases from 1 to n, rr_j^{α} of $\Re_j^{\alpha}$ can be obtained more easily by subtracting r_j from the rc_{j-1} of $\Re_{j-1}^{\alpha}$ (Step 3.1.1). Similarly, we can obtain $\hat{u}_{\Sigma_j}^{-1}$ by adding $\hat{u}_j^{-1}(.)$ into the last $\hat{u}_{\Sigma_{j-1}}^{-1}$ of $\Re_{j-1}^{\alpha}$ (Step 3.1.2). Thus, the complexity of the algorithm can be further reduced. The worst-case complexity of the USQ algorithm is given by $O(n^2)$, where n is the number of flows in the system.



(a) The procedure of deriving $\Re_j^{\alpha}$

(b) $\Re_j^{\alpha}$ and ΔU_j^{α}

Fig. 2. An illustration of the proposed algorithm USQ

D. The Performance Bound to Optimum

Next we prove that the proposed USQ allocation algorithm is bounded to the optimal solution. In the previous sub-section, we have shown in Theorem 2.4 that an optimal solution must satisfy conditions (1), (2), (3) and (4). Then, we can classify all these optimal solutions into two mutually exclusive sets: α and β. All allocations in set α satisfy conditions (1), (2), (3) and (5) (i.e., Definition 2.5), while those in set β satisfies (1), (2), (3) and (6), where condition (6) is the complement set of (5) in (4), i.e.,

$\forall i \in \Gamma$, there is one and only one flow i whose $u_i^{'}(r_i) > 0$ if $r_i > 0$. (7)

Let S_j^{β} be a subset of β that has j allocated flows. Based on Lemma 2.3 and any given traffic utility function, all solutions

Input: For each flow i with channel qualities q_i in the queue, $i = 1, 2, \cdots, n$, a sigmoid utility function $U(r)$, and total available resource is r_{total};
Output: $\Re_{sigmoid} = \{r_1, r_2, \cdots r_n\}$;

Algorithm:
(1) Initialize $r_i = 0, i = 1, 2, \cdots, n, rr_0 = r_{total}$, and $U(\Re_0) = 0$.
(2) Sort all flows i in descending order of q_i, and store them in the queue.
(3) For $j = 1$ to n,
  (3.1) Derive the allocation $(\Re_j^{\alpha})$ as follows.
    (3.1.1) $rr_j^{\alpha} = rr_{j-1}^{\alpha} - rc_j$;
    (3.1.2) $\hat{u}_{\Sigma_j}^{-1}(.) = \hat{u}_{\Sigma_{j-1}}^{-1}(.) + \hat{u}_j^{-1}(.)$;
    (3.1.3) Invert $\hat{u}_{\Sigma_j}^{-1}$ to obtain $\hat{u}_{\Sigma_j}$;
    (3.1.4) $u_{a,j} = u_{\Sigma_j}(rr_j^{\alpha})$;
    (3.1.5) For $i = 1, 2, \cdots, j$, $r_i^{\alpha,j} = rc_i + \hat{u}_i^{-1}(u_{a,j})$;
    For $i = j+1, \cdots, n$, $r_i^{\alpha,j} = 0$;
    (3.1.6) Return $\Re_j^{\alpha}$
  (3.2) Calculate $\Delta U_j = U(\Re_j^{\alpha}) - U(\Re_{j-1}^{\alpha})$ as follows.
    (3.2.1) If $\Delta U_j < 0$, set $\Re_{j-1}^{\alpha}$ as $\Re_{sigmoid}$;
    exit this loop;
(4) Return $\Re_{sigmoid}$.

TABLE I
TABLE I. THE PROPOSED ALGORITHM USQ FOR SOFT QOS TRAFFIC

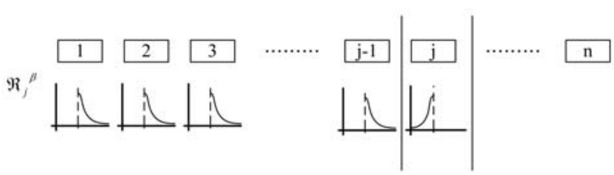


Fig. 3. An illustration of $\Re_j^{\beta}$ in S_j^{β}

in S_j^{β} satisfy the condition that the jth flow (which is the one with the least amount of allocated resource among all allocated flows) has $u'(r_j) > 0$. Let $\Re_j^{\beta}$ be the solution with the largest utility value in S_j^{β}. Fig. 3 gives an example allocation in S_j^{β}. We claim that if $\Re_j^{\beta}$ is the solution obtained by the proposed USQ allocation algorithm, the optimal solutions to this soft QoS allocation problem must only fall in the set $\{\Re_j^{\alpha}, \Re_j^{\beta}, \Re_{j+1}^{\beta}\}$.

Lemma 2.9: The optimal solution to this soft QoS problem must fall in the set $\{\Re_j^{\alpha}, \Re_j^{\beta}, \Re_{j+1}^{\beta}\}$, where $\Re_k^{\alpha}$ is the solution obtained by the USQ allocation algorithm.

Proof: We classify all possible optimal solutions in two mutually exclusive sets α and β. If an optimal solution falls in α, it must be $\Re_k^{\alpha} = \{r_f^{\alpha,k}, f \in \Gamma\}$, where $U(\Re_k^{\alpha}) = \max_{x=1,\cdots,n} U(\Re_x^{\alpha})$, i.e., the solution obtained by our proposed USQ algorithm. If it is in β, the optimal allocation, say, $\Re_k^{\beta} = \{r_f^{\beta,k}, f \in \Gamma\}$ must satisfy $U(\Re_k^{\beta}) = \max_{x=1,\cdots,n} U(\Re_x^{\beta})$.

If $\Re_j^{\beta}$ is optimal, $U(\Re_j^{\beta}) \geq U(\Re_{j-1}^{\alpha})$ holds, it follows that

$$U_j(r_j^{\beta,j}) + (k-1)\cdot U_c + \hat{U}_{\Sigma_{(j-1)}}(rr_j^{\beta})$$
$$\geq (k-1)\cdot U_c + \hat{U}_{\Sigma_{(j-1)}}(rr_{j-1}^{\alpha}),$$

which lead to $U_j(r_j^{\beta,j})$

$$\geq \hat{U}_{\Sigma_{(j-1)}}(rr_{j-1}^{\alpha}) - \hat{U}_{\Sigma_{(j-1)}}(rr_j^{\beta})$$
$$= \int_{rr_j^{\beta}}^{rr_{j-1}^{\alpha}} \hat{u}_{\Sigma_{(j-1)}}(r)dr$$
$$\geq \hat{u}_{\Sigma_{(j-1)}}(rr_{j-1}^{\alpha})\cdot(rr_{j-1}^{\alpha} - rr_j^{\beta})$$
$$= u_{a,j-1}\cdot(r_j^{\beta,j}).$$

Note that here we let $r_j^{\beta,j}$ denote the amount of resource allocated to the jth flow in $\Re_j^{\beta}$, and let rr_j^{β} denote $r_{total} - \sum_{i=1}^{j-1} rc_i - r_j^{\beta,j}$, which is the amount of residual resource allocated to the first $(j-1)$ flows (i.e., flows with $u_i'(r_i) < 0, i = 1, 2, \cdots, j-1$) in $\Re_j^{\beta}$.

Similar to Lemma 2.8, for $r < rc_i$, $i = 1, 2, \cdots, n$, $U_i(\cdot)$ is increasing, leading to

$$U_j(rc_j) - U_j(r_j^{\beta,j})$$
$$= \int_{r_j^{\beta}}^{rc_j} u_j(r)dr$$
$$\geq u_j(r_j^{\beta,j})\cdot(rc_j - r_j^{\beta,j})$$
$$= u_{\Sigma_{(j-1)}}(rr_j^{\beta})\cdot(rc_j - r_j^{\beta,j})$$
$$\geq u_{a,j-1}\cdot(rc_j - r_j^{\beta,j}).$$

Note that $u_{\Sigma_{(j-1)}}(rr_j^{\beta}) \geq u_{\Sigma_{(j-1)}}(rr_{j-1}^{\alpha}) = u_{a,j-1}$ since $u_{\Sigma(j-1}(.)$ is a decreasing function and $rr_j^{\beta} \leq rr_j^{\alpha}$.

Adding the two functions above, we obtain $U_j(r_j^{\beta,j}) + (U_j(rc_j) - U_j(r_j^{\beta,j})) = U_c \geq u_{a,j-1}\cdot rc_j$.

Then from Lemma 2.8, we have

$$U(\Re_{j-1}^{\alpha}) - U(\Re_{j-2}^{\alpha})$$
$$= U_{j-1}(r_{j-1}^{\alpha,j-1}) - \int_{rr_{j-1}}^{rr_{j-2}} \hat{u}_{\Sigma_{j-2}}(r)dr$$
$$\geq \left(U_c + \int_{rc_{j-1}}^{r_{j-1}^{\alpha,j-1}} \hat{u}_{j-1}(r)dr\right) - \hat{u}_{\Sigma_{j-2}}(rr_{j-1}^{\alpha})\cdot(rr_{j-2}^{\alpha} - rr_{j-1}^{\alpha})$$
$$\geq [U_c + \hat{u}_{j-1}(r_{j-1}^{\alpha,j-1})\cdot(r_{j-1}^{\alpha,j-1} - rc_{j-1}^{\alpha})] - \hat{u}_{\Sigma_{j-2}}(rr_{j-1}^{\alpha})\cdot(rr_{j-2}^{\alpha} - rr_{j-1}^{\alpha})$$
$$\geq [U_c + \hat{u}_{j-1}(r_{j-1}^{\alpha,j-1})\cdot(r_{j-1}^{\alpha,j-1} - rc_{j-1}^{\alpha})] - \hat{u}_{\Sigma_{j-1}}(rr_{j-1}^{\alpha})\cdot r_{j-1}^{\alpha,j-1}$$
$$= [U_c + u_{a,j-1}\cdot(r_{j-1}^{\alpha,j-1} - rc_{j-1}^{\alpha})] - u_{a,j-1}\cdot r_{j-1}^{\alpha,j-1}$$
$$= U_c - u_{a,j-1}\cdot rc_{j-1}^{\alpha} \geq [u_{a,j-1}\cdot rc_j^{\alpha} - u_{a,j-1}\cdot rc_{j-1}^{\alpha}]$$
$$\geq 0.$$

Since $\Re_j^{\beta}$ is optimal, $U(\Re_j^{\beta}) \geq U(\Re_j^{\alpha})$ holds. Then from Lemma 2.8, it follows that $U(\Re_{j-1}^{\alpha}) > U(\Re_{j-2}^{\alpha}) > \cdots > U(\Re_1^{\alpha})$ if $U(\Re_j^{\beta}) \geq U(\Re_{j-1}^{\alpha})$. Similarly, given that $U(\Re_j^{\beta}) \geq U(\Re_j^{\alpha})$, we can prove $U(\Re_j^{\alpha}) \geq U(\Re_{j+1}^{\alpha})$, and that $U(\Re_n^{\alpha}) < \cdots < U(\Re_{j+1}^{\alpha}) < U(\Re_j^{\alpha})$. Thus, if $\Re_j^{\beta}$ is an optimal allocation, either $\Re_{j-1}^{\alpha}$ or $\Re_j^{\alpha}$ is the solution obtained by USQ, i.e., $\Re_k^{\alpha}$ whose $U(\Re_k^{\alpha}) = \max_{x=1,\cdots,n} U(\Re_x^{\alpha})$. On the other hand, if $\Re_k^{\alpha}$ is the solution obtained by our algorithm, only $\Re_k^{\beta}$ and $\Re_{k+1}^{\beta}$ can be optimal in set β. Hence, given that $\Re_k^{\alpha}$ is the solution of our algorithm, the optimal solution must fall in the set $\{\Re_j^{\alpha}, \Re_j^{\beta}, \Re_{j+1}^{\beta}\}$. ■

Theorem 2.10: The difference in performance of the proposed USQ algorithm and the optimal allocation is bounded by the value U_c.

Proof: From Lemma 2.7, given that the proposed algorithm finds a solution of $\Re_k^{\alpha}$, the solutions that could be optimal include $\Re_k^{\alpha}$, $\Re_k^{\beta}$ and $\Re_{k+1}^{\alpha}$. If $\Re_{k+1}^{\alpha}$ is optimal, the difference in total utilities of the optimal and the one obtained by our algorithm is zero. If $\Re_{k+1}^{\beta}$ is optimal, the difference between $\Re_{k+1}^{\beta}$ and $\Re_k^{\alpha}$ is given by

$$U(\Re_{k+1}^{\beta}) - U(\Re_k^{\alpha})$$
$$= \sum_{i=1}^{k+1} u_i(r_i^{\alpha,k+1}) - \sum_{i=1}^{k} u_i(r_i^{\alpha,k})$$

$$= u_{k+1}(r_{k+1}^{\alpha,k+1}) + [\sum_{i=1}^{k} u_i(r_i^{\alpha,k+1}) - \sum_{i=1}^{k} u_i(r_i^{\alpha,k})]$$
$$\leq u_{k+1}(r_{k+1}^{\alpha,k+1}) < u_{k+1}(rc_{k+1}) = U_c.$$

Since $\Re_j^{\alpha}$ is the one with the largest utility value among all allocations in α, we have $\sum_{i=1}^{j} u_i(r_i^{\alpha,j}) \geq \sum_{i=1}^{j} u_i(r_i^{\alpha,j+1})$. From (6), we obtain $r_{j+1}^{\alpha,j+1} \leq rc_{j+1}$, and thus $u_{j+1}(r_{j+1}^{\alpha,j+1}) \leq u_{j+1}(rc_{j+1}) = U_c$.

Similarly, it can be easily proved that if $\Re_k^{\beta}$ is an optimal solution, then the difference in total utilities can also be bounded by U_c due to $U(\Re_k^{\beta}) - U(\Re_k^{\alpha}) \leq U(\Re_k^{\beta}) - U(\Re_{k-1}^{\alpha}) \leq U_c$. Therefore, the difference in the performance of our algorithm and the optimal solution is bounded by U_c, no matter if the optimal solution is $\Re_k^{\alpha}$, $\Re_k^{\beta}$ or $\Re_{k+1}^{\beta}$. ∎

In Table I, we observe that our proposed sub-optimal algorithm has a polynomial time computational complexity, and is feasible for a network BS. However, we cannot find an approach to solving the original optimal problem because it is difficult to find $\Re_j^{\beta}$ from S_j^{β}.

III. Simulation Result

In this section, we conduct simulations to evaluate the performance of the proposed USQ allocation algorithm. In the simulations, we have 16 user flows, all with the same sigmoid traffic utility function $U(r)$. The channel quality q of each flow is randomly selected from the range [0, 1]. The sigmoid utility function is composed of two exponential functions:

(1)$U(r) = q \cdot e^{p(r-rc)}$, if $r < rc$;

(2)$U(r) = 1 - (1-q) \cdot e^{-p(r-rc)}$, otherwise,

where q denotes the utility value when $r = rc$, and p determines the slope of the utility function. We fix q at 0.5 in the simulations. When $r < rc$, we let the utility increase with r. When $r > rc$, we let the utility saturate to 1.

In the first simulation, we vary the slope of the utility function by tuning the value of p so as to observe how the value of p affects the resource distribution. The value of r_{total} is fixed at 800 in this simulation. Fig. 4 plots the impact of p on the utility function $U(r)$. We observe that a different value of p results in a different function. The larger the value of p, the more similar it is to the hard QoS's unit-step function; when the value of p is 0.1, the shape is similar to the best effort concave utility function. The resource distribution among flows is depicted in Fig. 5. The result shows that when p is small, meaning that the marginal utility saturates more slowly as the given resource increases, the system favors flows with better channel qualities; on the other hand, when p becomes larger (i.e., the slope is sharper), meaning that the marginal utility saturates faster and the amount of resource each flow obtains is no more than rc_i, the system tends to give more resource to flows with bad channel qualities so as to let all flows i have an identical θ_i (i.e., $\theta_i = r_i \cdot q_i$). Therefore, by tuning the value of p, we can adjust the wireless system to behave in a throughput-oriented (i.e., p is small) or in a fairness-oriented (i.e., p is large) manner.

Next, we compare the resource distribution obtained by USQ with the algorithm proposed in our previous work [14]. In [14], we proposed a mechanism which allocates resource to both hard QoS and best effort traffic by maximizing the total utility. The utility functions of hard QoS and best effort traffic are based on a unit-step function and a concave function, respectively, as reported in [15]. We decrease the value of p in the proposed system (i.e., from fairness-oriented to throughput-oriented) and then compare the results obtained by USQ with the hard QoS traffic allocation and the best effort traffic allocation proposed in [12].

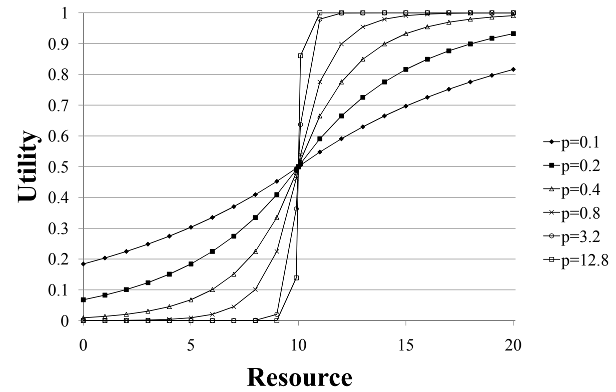


Fig. 4. Relationship between p and utility function $U(.)$

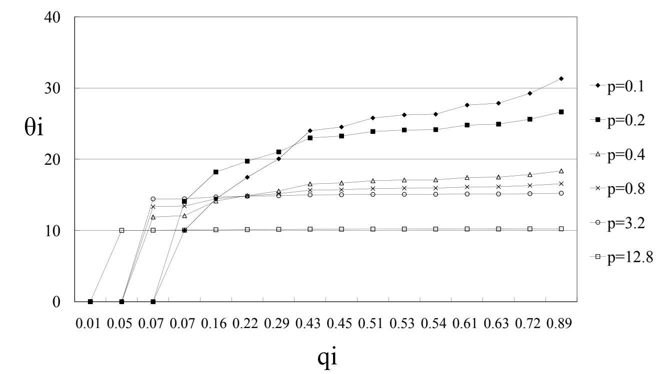
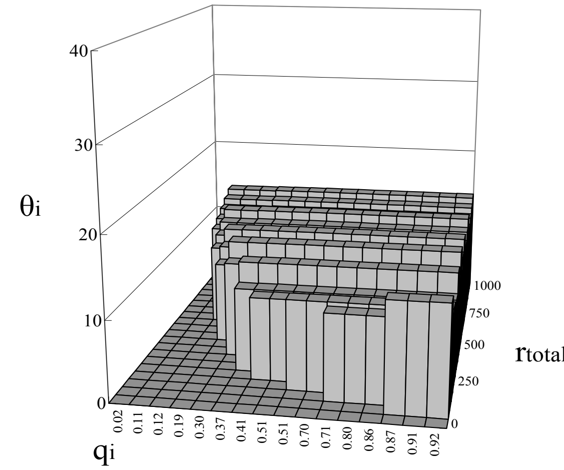
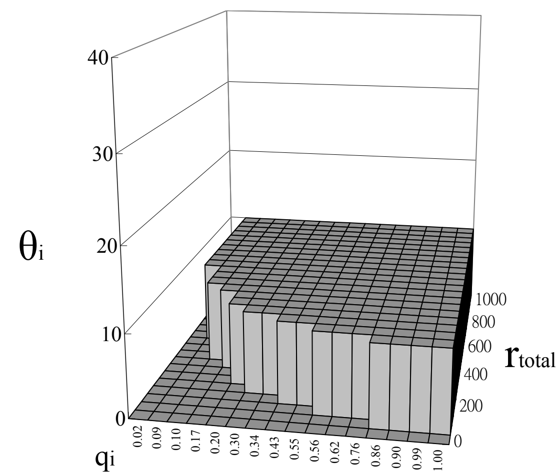


Fig. 5. Relationship between p and θ_i of each flow

We first set p to 12.8, i.e., a utility function with a very steep slope, and then compare the result with the hard QoS allocation. The utility function of hard QoS traffic is given by $f_u(r-10)$, where $f_u(\cdot)$ is a unit-step function. Since the step function for hard QoS traffic and the sigmoid function with $p = 12.8$ for soft QoS traffic are very similar (see. Fig. 6), and both mechanisms allocate resource based on utility maximization, the rc_i of each allocated flow i in the proposed USQ algorithm (Fig. 6(a)) is very close to that in the hard QoS allocation in [14] (Fig. 6(b)). However, there are still some subtle differences. In the hard QoS allocation, the rc_i of each allocated flow i is guaranteed to be $10/q_i$, but in the soft QoS allocation, slightly more resources are assigned to flows. The reason is that with the sigmoid utility function, all available resources are allocated to flows for maximizing the total utility (i.e., (1)), while the resource assigned to hard QoS is allocated in a discrete way and only with the requested amount, leading to some residual resource left unused. By setting p to a very

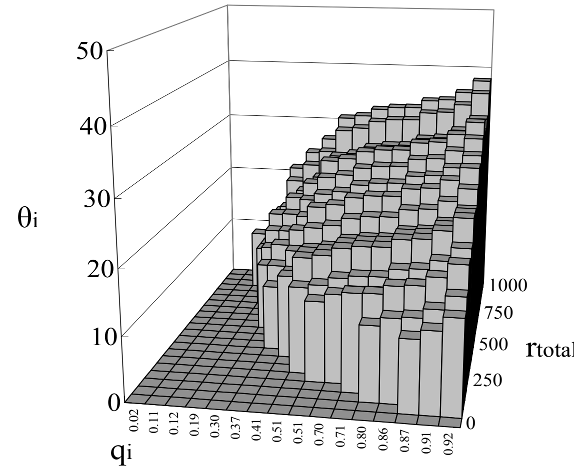


(a) Soft QoS allocation ($p = 12.8$)

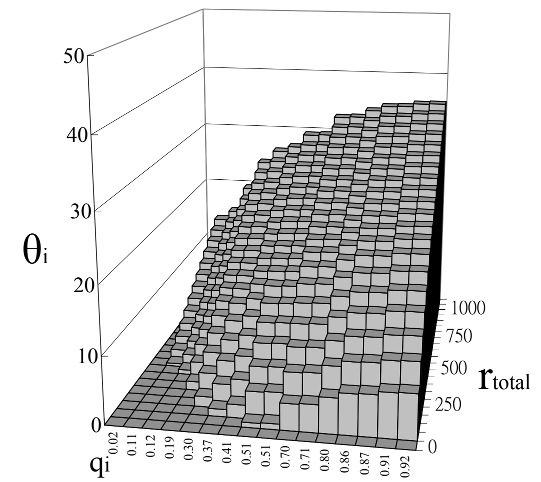


(b) Hard QoS allocation

Fig. 6. Resource distribution for the soft QoS allocation ($p = 12.8$) and the hard QoS allocation in [12].



(a) Soft QoS allocation ($p = 0.1$)



(b) Best effort allocation

Fig. 7. Resource distribution for the soft QoS allocation ($p = 0.1$) and the best effort allocation in [12].

large value (i.e., rendering a very steep function similar to a unit-step function), the proposed soft QoS algorithm act like the hard QoS algorithm in [14].

Next, we set p to 0.1, which results in a relatively flat utility function, and compare the allocation result with the best effort allocation in [14]. Again, both systems are given similar settings and the results are shown in Fig. 7. The utility function of best-effort traffic is based on a concave function $U_{BE}(r) = 1 - e^{r/10}$. The figure shows that when p is small, the result of the soft QoS allocation (Fig. 7(a)) is very close to that of the best effort allocation (Fig. 7(b)). Unlike in Fig. 7, USQ gives much more resource to flows with better channel qualities as r_{total} increases, i.e., being more throughput-oriented, as shown in Fig. 7. However, in Fig. 7(a), each allocated flow's minimal rate can be guaranteed not less than 10, while in Fig. 7(b), allocated flows cannot be ensured to obtain a minimal value.

IV. Conclusion

In this paper, we study the utility maximization problem for resource allocation to soft QoS traffic in infrastructure-based wireless networks. We describe the design guidelines for this problem, and propose a polynomial time algorithm which is proved to be tightly bounded to the optimal solution. The performance of the proposed USQ mechanism is evaluated by simulations. The results show that this mechanism can not only allocate resource according to both users' channel qualities and total network resource, but also adapt to different traffic types (i.e., throughput-oriented or fairness oriented). We also show that the mechanism proposed in our previous work in [14], which considers the co-existence of hard QoS traffic (e.g., VoIP traffic) and best effort traffic in the network, is a special case of USQ. Thus, the proposed USQ mechanism is applicable to the scenarios in which traffic with QoS requirements (including hard QoS and soft QoS) and without QoS requirements (e.g., best effort traffic) co-exist in the system.

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

[3] J.-R. Ohm, "Description of core experiments in MPEG-4 video," ISO/IEC JTCI/SC29/WG11, N2554, 1998.

[4] ITU-T recommendation P.800.

[5] K. D. Wu and W. Liao, "Flow allocation in multi-hop wireless networks: a cross-layer approach," *IEEE Trans. Wireless Commun.*, vol. 7, no. 1, pp. 269–276, Jan. 2008.

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

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

[8] 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.

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

[10] 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.

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

[12] 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.

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

[14] W. H. Kuo and W. Liao, "Utility-based resource allocation in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 6, no. 10, pp. 3600–3606, Oct. 2007.

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

Wen-Hsing Kuo received his BS and the PhD degrees in Electrical Engineering from National Taiwan University, Taipei, Taiwan, in 2002 and 2008, respectively. He joined the Department of Electrical Engineering, Yuan-Ze University, Taoyuan, Taiwan, as an Assistant Professor in 2008. His research interests include network resource management, network economics and wireless access 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.