# Delay Sensitive Scheduling Schemes for Heterogeneous QoS over Wireless Networks

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

***Abstract*— Future wireless networks will support the growing demands of heterogeneous and delay sensitive applications. In this paper, a users' satisfaction factor (USF) is defined to quantify quality of service (QoS) for different types of services such as voice, data, and multimedia, as well as for different delay constraints. This USF not only predicts the final delivered QoS during transmission, but also take advantages of the fact that different packets can be decoded at different time in the receivers. Based on this USF, four types of scheduling schemes considering tradeoffs between system performance and individual fairness are proposed. These schemes explore the time, channel, and multi-user diversity to guarantee quality of service and enhance the network performance. From the simulation results, the proposed scheduling schemes achieve different tradeoffs between individual fairness and high system performance for the heterogeneous and delay sensitive applications, compared with the weighted round-robin and the modified proportional fairness scheduling schemes.**

***Index Terms*— Resource management, scheduling, multimedia communication, integrated voice-data communication, adaptive modulation.**

## I. INTRODUCTION

TO satisfy the growing demand for heterogeneous applications of wireless networks, it is critical to deliver flexible, variable rate services with high spectral efficiencies. Different Quality of Service (QoS) provisioning over time-varying channel conditions is a key challenge for system design. Resource allocation provides a strategic means to address this challenge. Current wireless systems use signal to noise ratio (SNR) as a QoS measure. Some popular resource allocation schemes in this context are power control and rate adaptation [1]–[3]. However with the increasing demand for multimedia transmissions, their delay sensitive QoS cannot be satisfied by the SNR measure itself. So it is necessary to define new QoS measures for multimedia transmission services.

Opportunistic scheduling [4] [5] has attracted a lot of research attentions as a special kind of resource allocation over time. The basic idea of "opportunistic scheduling" is allocating resources to links experiencing good channel conditions while avoiding allocating resources to links experiencing bad channel conditions, thus efficiently utilizing radio resources. This is also referred as channel-aware scheduling which explores time/multi-user diversity. On the other hand, opportunistic scheduling introduces an important tradeoff between system performance and fairness among users. For example, allowing only users with good channel conditions to transmit may result in high throughput, but meanwhile sacrifice the transmissions of other users. In the literature, some long term fairness has been considered, such as proportional fairness [6], [7] and time average fairness [8]. A class of dynamic fair scheduling schemes based on the generalized processor sharing (GPS) fair service discipline, under the generic name code-division GPS (CDGPS), is proposed for a CDMA cellular network to support multimedia traffic [9]. However, there are few works studying the short-term fairness [10] with delay constraints.

In this paper, our goals are to define a new QoS measure for multimedia transmissions and to develop new scheduling schemes, so as to improve system performance and maintain the individual fairness among users. First, we quantify the QoS measure as a user satisfaction factor (USF) for different delay constraints and different applications such as voice, video, and data. The proposed USF is a function of both the number of received bits and the delay sensitivity profiles. Since some layered decoders can decode the layers such as base layer earlier than the other layers, the USF measure can take this advantage for better resource allocation. Moreover, the USF predicts the final delivered QoS during the transmission. Then based on this QoS measure, we develop four types of scheduling schemes, namely the maximin approach, the overall performance approach, the two-step approach, and the proportional approach. Simulation results show that these four scheduling schemes achieve different tradeoffs between the system performance and individual fairness, compared with the weighted round-robin and the modified proportional fairness scheduling schemes.

This paper is organized as follows: System model is given in Section II. In Section III, USF is defined for heterogeneous QoS with different delay requirements. In Section IV, four types of scheduling schemes are proposed and some analysis is conducted. Simulation results are shown in Section V and conclusions are drawn in Section VI.

## II. SYSTEM MODEL

We consider a downlink scenario of a single-cell system. The system has $K$ users randomly located within the cell. The transmission bandwidth is $W$. Slow fading is assumed such that the channel gain is stable within each packet. For simplicity, the channel parameters from different users are

Manuscript received April 8, 2005; revised September 5, 2005; accepted January 29, 2006. The associate editor coordinating the review of this paper and approving it for publication was X. Shen. This work is supported in part by MURI AFOSR F496200210217.

Z. Han is with the Department of Electrical and Computer Engineering, Boise State University, Boise, ID 83725 (e-mail: zhuhan@boisestate.edu).

X. Liu is with the Department of Computer Science, University of California, Davis, CA 95616, USA (e-mail: liu@cs.ucdavis.edu).

Z. J. Wang is with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: zjanew@ece.ubc.ca).

K. J. R. Liu is with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA (e-mail: (kjrliu@glue.umd.edu).

Digital Object Identifier 10.1109/TWC.2007.05221.

TABLE I
REQUIRED SNR FOR DIFFERENT BER OF DIFFERENT ADAPTIVE MODULATION AND CONVOLUTIONAL CODING RATES

| $k$ | Rate $\omega_k$ | Modulation | Coding Rate | SNR $\rho_k$ (dB) BER $\leq 10^{-6}$ |
|---|---|---|---|---|
| 1 | 1W | QPSK | 1/2 | 4.65 |
| 2 | 1.33W | QPSK | 2/3 | 6.49 |
| 3 | 1.5W | QPSK | 3/4 | 7.45 |
| 4 | 1.75W | QPSK | 7/8 | 9.05 |
| 5 | 2W | 16QAM | 1/2 | 10.93 |
| 6 | 2.66W | 16QAM | 2/3 | 12.71 |
| 7 | 3W | 16QAM | 3/4 | 14.02 |
| 8 | 3.5W | 16QAM | 7/8 | 15.74 |
| 9 | 4W | 64QAM | 2/3 | 18.50 |
| 10 | 4.5W | 64QAM | 3/4 | 19.88 |
| 11 | 5.25W | 64QAM | 7/8 | 21.94 |

assumed perfectly estimated, and the channel information is reliably feed back from mobiles users to the base station without any delay. Denote $\Gamma_k$ as the $k^{th}$ user's SNR as: $\Gamma_k = \frac{B_k G_k}{\sigma^2}$, where $G_k$ is the channel gain and $B_k$ is the transmit power for the $k^{th}$ user. The thermal noise power for different users is assumed to be the same and represented as $\sigma^2$. For the downlink system, because of the practical constraints in implementation, such as the limitation of power amplifier and consideration of co-channel interferences to other cells, the transmit power is bounded by $B_{max}$.

We also assume that with appropriate channel modulation and coding, the packet loss rate is sufficiently low. In other words, Bit Error Rate (BER) of the channel transmission is kept lower than some desired threshold. Adaptive modulation and adaptive channel coding provide each user with the ability to adjust his/her data transmission rate $r_k$, according to the channel condition, for achieving the desired BER. We focus on MQAM modulation and convolutional codes as they provide high spectrum efficiency and strong forward error protection, respectively. We select the BER threshold as BER $\leq 10^{-6}$. The required SNR's for BER $= 10^{-6}$ with different modulations and convolutional coding rates using bit interleaved coded modulation (BICM) are listed in Table I, based on the results in [2]. To facilitate discussion, denote the number of combinations in Table I of different modulation and convolutional coding rates as $Q$. Define the feasible set of the rate as $\omega = \{\omega_0, \omega_1, \omega_2, \ldots, \omega_Q\}$, and the corresponding set of the required SNR for BER $\leq 10^{-6}$ as $\rho = \{\rho_0, \rho_1, \rho_2, \ldots, \rho_Q\}$, where $\omega_0 = 0$ and $\rho_0 = 0$. Thus, all $r_k$ should be selected from the set $\omega$. For the scheduling scheme, the rate should be selected such that the maximal transmission power $B_{max}$ is able to maintain the BER lower than $10^{-6}$. Specifically, the following inequalities should be held if user $i$'s transmit rate $\omega_j$ is selected:

$$\rho_j \leq \frac{B_{max} G_i}{\sigma^2} < \rho_{j+1}, \text{ for } j = 0, \ldots, Q-1; \quad (1)$$
$$\text{and } \rho_j \leq \frac{B_{max} G_i}{\sigma^2}, \text{ for } j = Q.$$

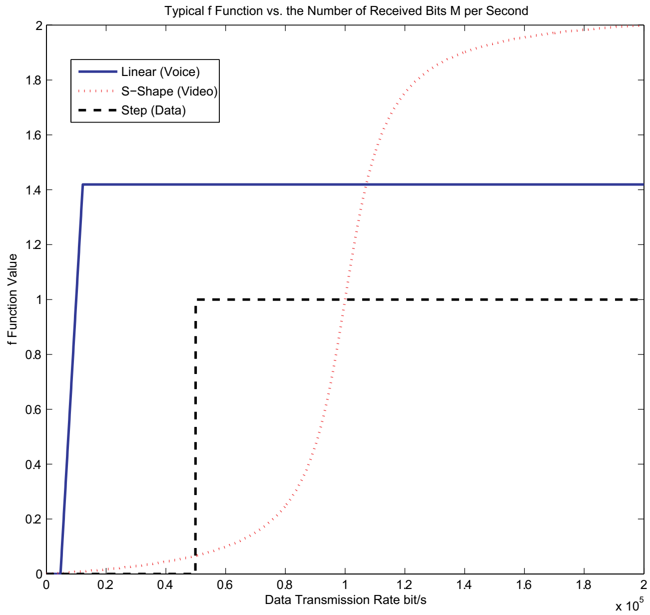


Fig. 1. Typical $f$ functions vs. received bits $M$ per second.

## III. User Satisfaction Factor

In this section, we define USF that represents multimedia users' QoS considering both received bits and delay sensitivity profile. First, we consider QoS representation when all transmitted bits are received. We define $f$ to quantify QoS as a function of $M$, number of received bits. For different types of services, $f$ can have different definitions. For example, for data transmission, $f$ can be a step function, i.e., only when all information is received can QoS be ensured. For most embedded voice coders, the mean square error (MSE) can be approximated by an exponential function of $M$. So the corresponding reconstructed peak signal to noise ratio (PSNR) can roughly be a linear function of $M$. For the layered video encoders such as MPEG4-FGS, the base layer packets should be transmitted first to maintain the basic QoS, and then the enhanced layer packets are transmitted with the higher priority packets first. The resulting PSNR is a kind of S-shape. Some typical $f$ functions are illustrated in Figure 1. Here we normalize the QoS for the different users, so that users' QoS is exactly satisfied when $f = 1$. When $f > 1$ users are over satisfied, and when $f < 1$ users are under satisfied. For the different applications, to achieve the same $f$, the necessary numbers of the received bits $M$ per second can be different.

For different types of multimedia transmissions, different packets can have different delay sensitivity profiles. For example, for voice services, the packets should be received with a strict delay. For layered video services, base layer can be transmitted and decoded earlier than the enhanced layers. So some proportion of data can be transmitted earlier than the other. We define $N$ as the packet transmission time with the strictest delay constraint, and $n'$ as the packet transmission finishing time. The delay sensitivity profile can be presented by $P(n')$, the weight of finishing the transmission at time $n'$, when $n' \geq N$. This weight represents different delay tolerance levels for different packets.

In this paper, we consider the geometric distribution for delay profiles as an example. Other types of distributions can be applied in a similar way. A parameter $\alpha$ depicting the delay tolerance is assigned when each user is admitted, where $0 \leq \alpha < 1$. Suppose a user finishes the transmission at time $n'$, where $n' \geq N$, the packet is associated with a weight $P(n' = N + i) = (1 - \alpha)\alpha^i,\ i = 0, 1, \cdots$. For notation simplicity, here we assume $0^0 = 1$, if $\alpha = 0$ and $i = 0$. The value of $\alpha$ represents the delay tolerance of the user. Different payload types in terms of delay tolerance can be categorized as:

1) **Strict Delay Constraint:** $\alpha = 0$, $P(n' = N) = 1$, and $P(n') = 0, \forall n' > N$. In this case, the packet must be transmitted before time $(N + 1)$. This fits such applications as voice payloads.
2) **Soft Delay Constraint:** $0 < \alpha < 1$. This fits the video/image payloads, where the higher priority bits can be transmitted and decoded earlier than the lower priority bits.
3) **No Delay Constraint:** $\alpha \to 1$, so $P(n' = N + i) \to 0, \forall\ i \geq 0$. Users can tolerate arbitrary delays. This fits time insensitive payloads such as emails.

In traditional networks, the system assigns the resources to users, according to current QoS and channel conditions. There are few works that considers the final delivered QoS, even though the wireless channels may fluctuate. Therefore, we need to define USF for QoS measurement such that the system can adapt its resource allocation strategy by predicting the final delivered QoS. Suppose there is an estimator $g$ that estimates the number of transmitted bits at time $n'$, based on the transmission history and channel conditions using the information no later than time $n$. The most straight-forward method is the linear prediction. At the current time $n$, the resulting prediction of the number of bits received at a future time $n'$ is

$$g(n, n') = \frac{n' \sum_{j=1}^{n-1} T_i(j)}{n - 1},\ n' \geq N, \tag{2}$$

where $T_i(j)$ is the $i^{th}$ user's transmitted bits at time $j$. Other types of function $g$ can also be explored, such as moving window, Kalman filter, etc.

We define USF at time $n$ as:

$$\mathrm{USF}_i(n) = \sum_{n'=N}^{\infty} P(n') f(g(n, n')). \tag{3}$$

The physical meaning of USF is the estimated weighted QoS according to the delay sensitive profile [1]. For example, for the voice services, $\alpha = 0$, then any components of $f$ with $n' > N$ are weighted by zero in (3). So the delayed transmission has no impact on USF. For the data transmissions with $\alpha \to 1$, USF is always 1. For the soft delay case, USF is affected by both QoS function $f$ and delay distribution profile of $P$. Note that this USF defines the short term QoS, since the delay profile is related to each packet's transmission time. The value of USF represents the user's QoS measurement as follows:

1) $\mathrm{USF}(n) > 1$: the user is over satisfied at the current time $n$. For the rest of transmission time, he/she may be less aggressive for acquiring resources.
2) $\mathrm{USF}(n) = 1$: user's QoS is exactly satisfied. If he/she acquires the resource at the current rate, the final QoS $f$ is equal to 1.
3) $0 \leq \mathrm{USF}(n) < 1$: when USF becomes smaller, the user becomes more unsatisfied and has to transmit more aggressively in the rest of packet transmission time.

[1]USF is not the estimation by the distribution of the real packet finishing time but by the delay sensitive weight instead.

## IV. Four Scheduling Schemes with USF

In this section, based on USF, we develop four types of scheduling schemes to improve both individual fairness and system overall performance. Then, to demonstrate the performance of the proposed scheduling schemes, a simple analysis regarding how different parameters affect the value of the USF change is given.

**Scheduling Scheme 1: Maximin Approach**: The first scheduler selects the user with the minimal USF for transmissions so as to improve the minimal performance of the system. The scheduling policy is expressed as:

$$\arg\min_i\ w_i \cdot \mathrm{USF}_i(n)$$

where $w_i$ is the weight value for user $i$. In this paper, we assume $w_i = 1, \forall i$. This approach is fair among users but will generate inferior system performance because the users in the bad channel conditions might consume most of scheduled transmission times.

**Scheduling Scheme 2: Overall Performance Approach**: This type of scheduler tries to maximize the overall USF. Define $\Delta\mathrm{USF}_i(n)$ as the USF change if the scheduler assigns the current time slot to user $i$. Since each time only one user occupies the channel, the scheduling policy is

$$\arg\max_i\ w_i \cdot \Delta\mathrm{USF}_i(n).$$

By doing this, a user with a good channel condition has a good chance for transmissions. Consequently the scheduler can take advantages of multiuser diversity and channel diversity to optimize the system performance. However the fairness among users is not considered. Notice that this type of scheduling scheme is very unfair for the users with no delay constraint, since $\Delta$USF is always a small number for them. In implementation, when the other users have USF's larger than 1, the users with no delay constraint are selected.

**Scheduling Scheme 3: Two-Step Approach**: The philosophy of this approach is to maintain the users' minimal QoS first, then to try to maximize the overall system performance. The approach is the combination of the first two approaches, shown as:

$$\begin{cases} \arg\min_i\ w_i \cdot \mathrm{USF}_i(n), & \text{if } \mathrm{USF}_i < 1, \exists i; \\ \arg\max_i\ w_i \cdot \Delta\mathrm{USF}_i(n), & \text{otherwise.} \end{cases}$$

This approach can satisfy the basic fairness among users and then optimize the system performance. So it is a tradeoff between the first two approaches.

**Scheduling Scheme 4: Proportion Approach**: In [6], the proportional fair scheduling is proposed, where the scheduler selects the user with the largest ratio of current possible maximal rate over the sum of rates in his/her transmission history. By doing this, the scheduler takes into account of the current channel condition and maintains the *long-term* *proportional* fairness. In this paper, we employ the similar idea. Compared with the second approach, the scheduler weights the user with the inverse of his/her current USF value plus a small positive constant $\varepsilon$. This constant is to prevent the error of "divided by zero". If the user has a better channel, the rate adaptation can transmit more information to improve his/her USF and the user has the advantage to transmit. On the other hand, if the USF is small, the user also has the advantage for transmissions. The scheduler can be expressed as:

$$\arg\max_i \frac{w_i \cdot \Delta\mathrm{USF}_i(n)}{\mathrm{USF}_i(n) + \varepsilon}.$$

This scheduler provides another tradeoff between the performance and fairness.

In the rest of this section, by using some assumptions to simplify the problem, we try to provide some theoretical analysis on how different parameters influence the value of $\Delta$USF. By using Rayleigh fading assumption as an example, the channel of any link is modelled as a zero mean circularly symmetric complex Gaussian random variable with unit variance. From (1), the probability that a user selects constellation of $2^j$ is

$$P_r(T_i = j|G_i) = \begin{cases} \exp^{-\frac{\sigma^2\rho_j}{B_{max}G_i}} - \exp^{-\frac{\sigma^2\rho_{j+1}}{B_{max}G_i}}, & j = 0, \ldots, Q-1, \\ \exp^{-\frac{\sigma^2\rho_Q}{B_{max}G_i}}, & j = Q. \end{cases} \quad (4)$$

Assume a linear assumption of $f$ as $f = \mu g(n, n')$ and $\alpha > 0$. The larger $\mu$, the higher demand for transmission rate. From (2) and (3), at each time the estimated $\Delta\mathrm{USF}_i$ with known $\mathrm{USF}_i$ and $G_i$ is given by

$$E(\Delta\mathrm{USF}_i|G_i, \mathrm{USF}_i) = C\left(\frac{\mu}{\alpha}\sum_{j=1}^{Q}\exp^{-\frac{\sigma^2\rho_j}{B_{max}G_i}} - \mathrm{USF}_i\right) \quad (5)$$

where $C$ is a constant. The estimated $\Delta\mathrm{USF}_i$ is large, when the demand of transmission rate $\mu$ is large, the delay requirement is strict (small $\alpha$), or the channel gain $G_i$ is large. On the other hand, when the current USF is large, the user has less advantage for $\Delta\mathrm{USF}_i$. Notice that (5) is obtained based on several assumptions such as linear assumption of $f$ and Rayleigh Fading. In practice, it is very difficulty to obtain the close form solution of $\Delta\mathrm{USF}_i$. In the next section, we will conduct numerical study to evaluate the proposed schemes.

## V. Simulations Results

We conduct simulations with the following settings to demonstrate performances of the proposed schemes based on the new QoS measure USF. The channel gain [11] is given by

$$G = \frac{C_0}{r^a(1 + r\lambda_c/(4h_b h_m))^b} \quad (6)$$

where $C_0$ is a constant that depends on the antenna gain (here $C_0 = 10^{-7}$), $r$ is the distance between the mobile and the base station, $a$ is the basic path loss exponent (we select two), $b$ is the additional path loss component (we select one), $h_b$ is the base station antenna height, $h_m$ is the mobile antenna height, and $\lambda_c$ is the wavelength of the carrier frequency. The mobile

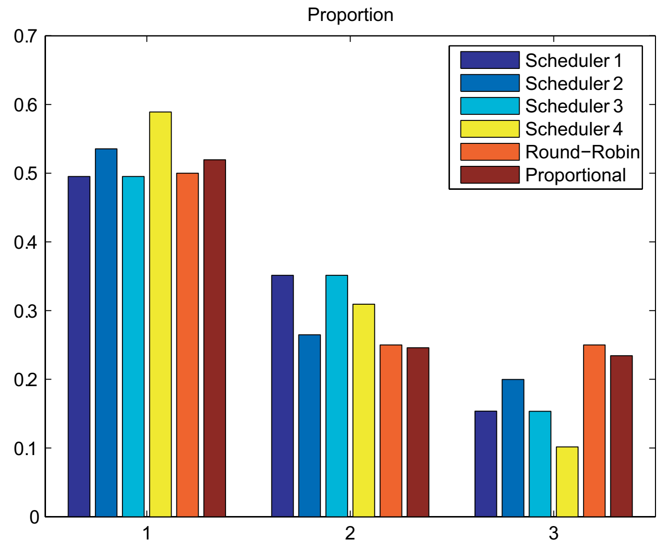


Fig. 2. Overall transmission proportions of different payloads.

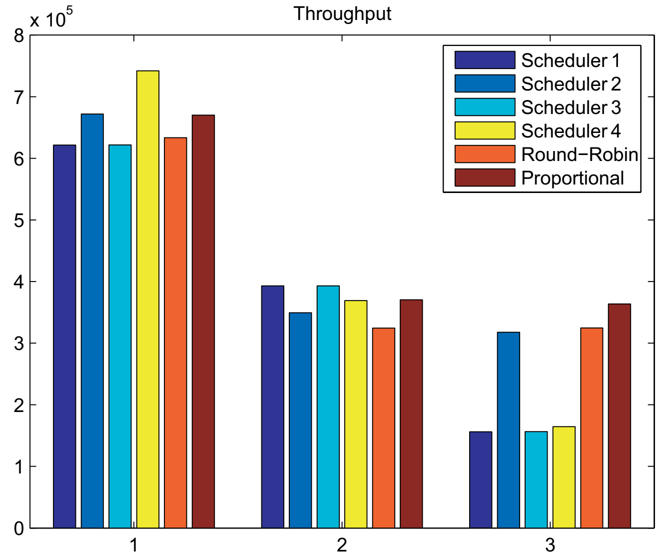


Fig. 3. Overall throughput of different payloads.

antenna height is 2m, and the base station antenna height is 50m. The carrier frequency is 1845-MHz. The total bandwidth is 90KHz. The maximal base station transmission power is 10W. Shadowing has standard deviation of 10dB. Background noise power is -90dBmW. All users are uniformly distributed within the range of $[r_0, r_1]$ with $r_0 = 20$m being the closest distance and $r_1 = 500$m being the cell radius. Each user has the minimum speed of 2km/h and maximum speed 100km/h. The period for direction change is 127 steps. The scheduling updates 2000 times per second and frame transmission time is equal to scheduling interval. In the simulations, we have 100 different settings of users' locations with 60 seconds for each setting. $\varepsilon = 0.001$. There are 13 users in the system. 10 users have voice payloads, 1 user has video payloads, and 2 users have data payloads. Three types of heterogenous payloads are shown in Figure 1 and described in details as follows:

1) **Type 1 User, Voice Payloads:** To simulate the real-time wireless voice communications, we employ the

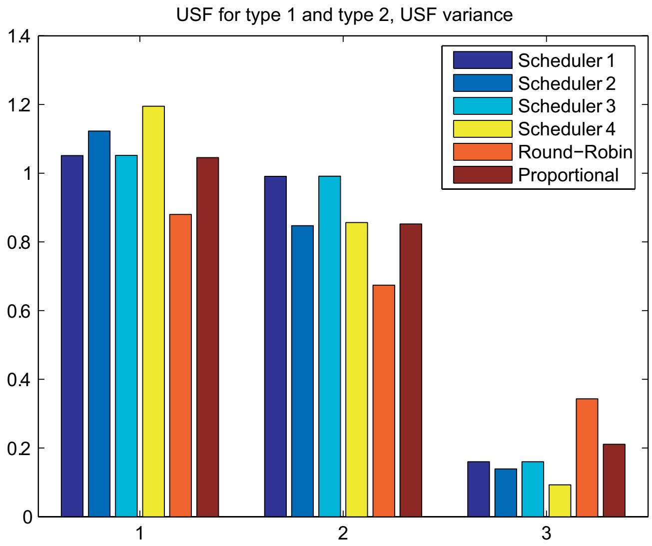


Fig. 4. Average USF of different payloads.

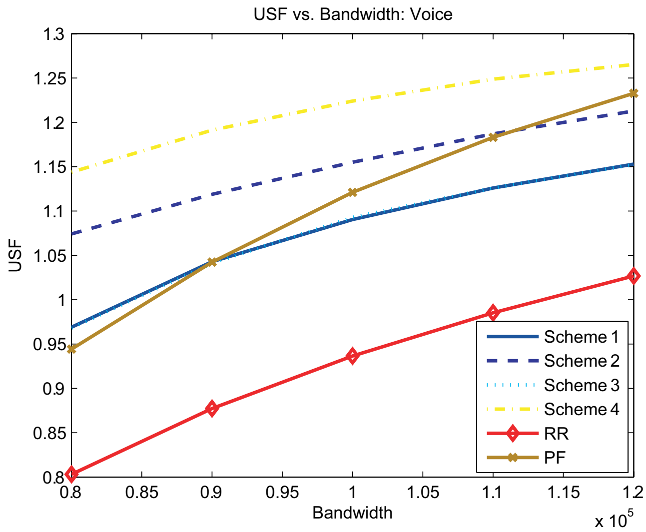


Fig. 5. USF vs. bandwidth for voice users.

GSM AMR (Advance Multi-Rate) Narrow-band Speech Encoder. This encoder operates with 20ms per packet with the possible encoding rates range from 4.75kbps to 12.2kbps. So we assume $\alpha = 0$ and delay time is 20ms. We use the linear approximation and assume $f = 0$ at 4.75kbps, $f = 1$ at 10kbps, and $f = 1.421$ when the rate is larger than 12.2kbps.

2) **Type 2 User, Video Payloads:** We consider the embedded video encoder that generates constant rates over time for simplicity. For general variable bit rate video encoder, the parameters for function $f$ and $M$ are changed for different video contents and different I, B, and P frames. We assume there are 10 frames per second. The strictest transmission time is 50ms and delay factor $\alpha = 0.99$. We approximate the QoS function as an $S$ shape function $f = \text{arctangent}(0.05 * (M - 5000)) + 0.5 * \pi$, where $f = 1$ at rate 100kbps and $f = 2.9442$ at rate 200kbps. This fits typical transmission of QCIF video with resolution of $176 \times 144$.

3) **Type 3 User, Data Payloads:** We assume the packet length is 50kbit. Each packet is appended with the CRC check. So we have $f = 1$, if all bits are decoded correctly and $f = 0$, otherwise. The arrival packets are modelled as Poisson arrival with rate equal to 1 packet/s.

Since the voice user's data rate is smaller than that of video or data users, each frame can carry up to 10 voice users' data, but only 1 video or data user's data. The above quantifications of QoS can be generalized. More practical considerations can be taken into accounts. For example, for voice coders, the talk spurt and silence spurt can have different rate requirements; for video encoders, different scenes and frames may have different parameters; other QoS measurement can also be considered for QoS function $f$.

In order to compare the performances, we also study the weighted round-robin and proportional fair scheduling schemes to the above settings. Since short term fairness is considered, we modify the proportional fair scheme [6], [7] using the following criteria:

$$\arg\max_i \frac{c_i T_i(n)}{\sum_{j=1}^{n-1} T_i(j)}, \tag{7}$$

where $T_i(n)$ is the maximal rate that can be achieved for user $i$ at time $n$, and $c_i$ is the weight factor. The video payloads have 5 times more weight and the data payloads have 2.5 times more weight than the voice payloads for both round-robin and modified proportional fair scheduling schemes.

In Figure 2-4, we show the overall transmission proportions, overall throughput, average USF, variance of USF of the different payloads, respectively. From the figures, we can see that the proposed four schemes give more proportions of time to the first type of users. Consequently, the first type of users have higher throughput and higher USF than those of the second type of users. This is because the proposed scheduling schemes take consideration of the strict short term delay requirement of the first type of users. In addition, for type 3 users, USF is always equal to 1 for all schedulers, because there is almost no delay requirement. The proposed schemes allocate the transmission time only if the other delay sensitive types of users' QoS's are satisfied. We quantify the fairness as the variance of different users' USF and show the variances for different schemes in the third column in Figure 4.

Compared with the proposed four schemes, we have the following observations. Scheme 1 has the least differences of USF among type 1 and type 2 payloads. This is because scheme 1 always tries to improve the performance of user with the least value of USF. However this scheme has the lowest throughput. Scheme 2 has the highest overall USF and the overall throughput is also high. However the difference of USF for two types is large. This is because scheme 2 only tries to maximize the overall USF without considering the fairness among users. Scheme 3 and scheme 4 provide certain tradeoffs between the system performance and individual fairness. Scheme 3 finds the tradeoff closer to the user fairness while scheme 4 finds the tradeoff more close to the system performance. The variances of USF for the proposed

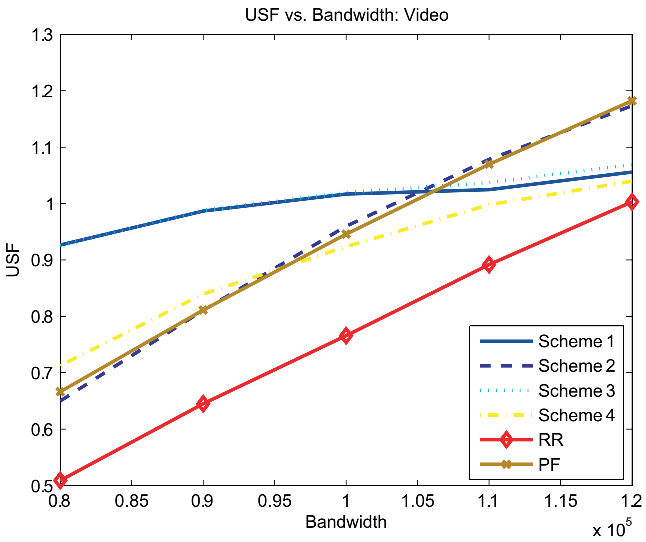


Fig. 6. USF vs. bandwidth for video users.

schemes are much less than those of the round robin and the proportional fairness schemes.

In Figure 5 and Figure 6, we show the USF of the different proposed schemes versus the overall bandwidth for voice and video users, respectively. We modify the bandwidth from 80kHz to 120kHz so that the system changes from being crowed to being less crowed. From the figure, we can see that all schemes perform better when the bandwidth is increasing. All proposed schemes have better performances than round robin schemes. When the bandwidth is small and the network is crowded, the proposed schemes have better performances than proportional fairness scheme. The increasing rates as function of bandwidth for scheme 2 and scheme 4 are larger than those of scheme 1 and scheme 3, since the schemes find different tradeoffs to performance and fairness.

## VI. Conclusion

In this paper, we define a new QoS measurement for multimedia and data transmissions considering heterogeneous and delay sensitive applications. Based on this measurement, four scheduling schemes are proposed. From the simulation results, compared with the weighted round-robin and modified proportional fair schemes, the proposed schemes find the different tradeoffs between individual fairness and system performance, while the heterogeneous nature and delay sensitivity of payloads are considered. The schemes can be further generalized to facilitate the design of the future wireless networks.

## References

[1] J. Zander and S. L. Kim, *Radio Resource Managment for Wireless Networks*. Artech House Publishers, 2001.

[2] M. H. Ahmed *et al.*, "Performance enhancement of joint adaptive modulation, coding and power control using cochannel-interferer assistance and channel reallocation," in *Proc. WCNC'03*.

[3] X. Qiu and K. Chawla, "On the performance of adaptive modulation in cellular systems," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 884–895, June 1999.

[4] X. Liu, E. Chong, and N. Shroff, "A framework for opportunistic scheduling in wireless networks," *Computer Networks Journal* (Elsevier), vol. 41, no. 4, pp. 451-474, 2003

[5] A. Jalali, R. Padovani, and R. Pankaj, "Data throughput of CDMA-HDR: a high efficiency-high data rate personal communication wireless system," in *Proc. IEEE Vehicular Technology Conference 2000-Spring*, vol. 3.

[6] A. Jalali, R. Padovani, and R. Pankaj, "Data throughput of CDMA-HDR: a high efficiency-high data rate personal communication wireless system," in *Proc. IEEE Vehicular Technology Conference 2000-Spring*, volume 3.

[7] F. P. Kelly, A. K. Maulloo and D. K. H. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *J. Operational Research Society*, vol. 49, pp. 237-252, 1998.

[8] Z. Han and K. J. R. Liu, "Power minimization under throughput management over wireless networks with antenna diversity," *IEEE Trans. Wireless Commun.*, vol. 3, no. 6, pp. 2170-2181, Nov. 2004.

[9] L. Xu, X. Shen, and J. W. Mark, "Dynamic fair scheduling with QoS constraints in multimedia W-CDMA cellular networks," *IEEE Trans. Wireless Commun.*, vol. 3, no. 1, pp. 60-73, Jan. 2004.

[10] Z. Han, Z. J. Wang, and K. J. R. Liu, "A resource allocation framework with credit system and user autonomy over heterogeneous wireless networks," in *Proc. IEEE Global Telecommunications Conference* 2003.

[11] P. Harley, "Short distance attenuation measurements at 900MHz and 1.8 GHz using low antenna heights for microcells," *IEEE Trans. Sel. Areas Commun.*, vol. 37, pp. 220-222, Nov. 1988.