

# Efficient Scheduling Algorithm for QoS Support in High Speed Downlink Packet Access Networks

MohammadReza HeidariNezhad, Zuriati Ahmad Zukarnain, Nur Izura Udzir, and Mohamed Othman

**Abstract**—In this paper, we propose APO, a new packet scheduling scheme with Quality of Service (QoS) support for hybrid of real and non-real time services in HSDPA networks. The APO scheduling algorithm is based on the effective channel anticipation model. In contrast to the traditional schemes, the proposed method is implemented based on a cyclic non-work-conserving discipline. Simulation results indicated that proposed scheme has good capability to maximize the channel usage efficiency in compared to another exist scheduling methods. Simulation results demonstrate the effectiveness of the proposed algorithm.

**Keywords**—Scheduling Algorithm, Quality of Service, HSDPA.

## I. INTRODUCTION

HIGH Speed Downlink Packet Access (HSDPA) improves on WCDMA by using different techniques for modulation and coding. The first phase of HSDPA has been specified in the 3rd Generation Partnership Project (3GPP) [1] release 5. It introduces new basic functions and is aimed to achieve peak data rates of 14.4 Mbps and 20 Mbps for MIMO systems.

In order to improve user and system performance for high speed IP traffic, HSDPA introduces new features such as a reduction of the Transmission Time Interval (TTI) to 2ms, link adaptation through and Adaptive Modulation and Coding (AMC) scheme, fast retransmission through a fast physical layer hybrid ARQ mechanism and multi-user diversity fast scheduling.

The Scheduler is a key element of HSDPA, which determines the overall behavior of the system and, to a certain extent, its performance.

One of the main goals of the HSDPA scheduler can be specified to maximize cell throughput while satisfying the QoS of different users. Since wireless bandwidth is the scarce resource and always the bottleneck for network throughput,

the scheduler should focus on improving the spectral efficiency of wireless resources. In the other words, the wireless scheduler should try to transmit as much data per slot as possible. From a system's perspective, enhancing channel usage efficiency is more important than assuring users Optimal QoS.

With the Channel Quality Indicator (CQI)[2] feedback, the scheduler is required to track the user's channel conditions quickly and adapt the data allocation accordingly. Due to the time-shared nature of HS-DSCH, users with good channel quality will get higher selection priority to achieve the optimal rate allocation and benefit for system efficiency. In addition, the design of the scheduling algorithm should take the fairness into account by giving the ones who are having bad channel conditions more priorities to increase their chance to being served and avoid the problem of starvation.

In this paper, a novel scheduling scheme with QoS support for hybrid services in HSDPA networks has been introduced. This method that is called Anticipated Proportionally Optimal (APO) algorithm, implemented based on a cyclic non-work-conserving discipline.

The rest of this paper is organized as follows. Section II reviews existing wireless packet scheduling algorithms. Section III presents a link-layer channel model that is used in the scheduling scheme. Section IV presents the proposed system framework and details of the cyclic APO algorithm. Section V illustrates results of simulations and analytical discussion. Section VI concludes this paper.

## II. RELATED WORKS

Existing QoS-support scheduling algorithms in wired networks cannot be used in wireless networks, in particular HSDPA, because they do not consider radio link conditions. In wireless networks, the users' channel capacities vary with time and in an asynchronous manner. For example authors in [3] show that the Earliest-Deadline-First (EDF) algorithm which provides optimal QoS-aware scheduling in wired networks, is not always optimal in the wireless case.

Recent research trend in scheduling algorithms is to consider following two conditions: user's traffic types or traffic class-based (QoS-based) and channel conditions dependent. Traffic class-based scheduling algorithms are designed to guarantee QoS for end users. They are classified in Real-Time (RT) and Non-Real-Time (NRT) services. Popular algorithms

Authors are with the Faculty of Computer Science and Information Technology, University Putra Malaysia. 43400, Serdang Selangor (phone: +603-89466555; fax: +603-89466576). MohammadReza HeidariNezhad, Zuriati Ahmad Zukarnain and Mohamed Othman are with Department of Communication Technology and Network (e-mail: {heidari, zuriati, mothman}@fsktm.upm.edu.my). Nur Izura Udzir is with Department of Computer System (e-mail: izura@fsktm.upm.edu.my).

for NRT packet scheduling include Maximum Carrier to Interference (Max C/I), Round Robin (RR), Proportionally Fair (PF) and Fast Fair Throughput (FFTH) [4].

PF and its extensions provide a good balance between the system throughput and fairness, but they do not take the QoS provision into account. Al-manthari *et al.* in [5] proposed a Fair and Effective Channel Dependent (FECD) scheduling algorithm. As another instance, a fair transmission scheduling algorithm proposed by Berggren and Jantti in [6]. Jiang [7] presents a utility-based approach for best-effort traffic.

The popular RT scheduling algorithm in wireless networks is Max-Weight based algorithms including Largest Weighted Delay First/Modified Largest Weighted Delay First (LWDF/MLWDF)[8]. Exponential Rule (EXP)[9], Modified Exponential Rule [10] and its extensions are another well defined instance in this scope.

Many researchers have been adopted and tested existing QoS schemes in wired networks for HSDPA wireless environment. For example, Delay-Sensitive Dynamic Fair Queuing (DSDFQ) [11] is based on a sorted priority queue mechanism which uses virtual finish time in weighted fair queuing algorithm that is generally used in wired networks. As another example, channel state aware EDF scheduler in wireless environment is studied by Chaporkar and Sarkar in [12].

Shakkottai and Stolyar in [9] present preliminary results of token-based scheduling scheme for mixed RT and NRT users. They use EXP scheduling for RT services. For case RT users are not present, it allocates leftover capacity to NRT users in a PF manner. However, in this scheduling scheme, channel usage efficiency does not take into account.

A numerous of proposed scheduler for HSDPA networks consider stationary channel modeling that is not a valid assumption due to movement of users. Nevertheless, in the existing scheduling methods, popularly a stationary channel model has been used. For example a effective capacity (EC) in [13] is designed for efficient bandwidth allocation and QoS provisioning. However, this model has not considered the dynamic link rate that happens in HSDPA. Moreover, stationary stochastic Markov process with finite state channel (FSC) models have been widely accepted and used by most of existing fast scheduling algorithms. For example Gilbert-Elliott channel, the finite state Markov channel model (FSMC) [14], the general hidden Markov models [15] and the K-th order Markov model [16] governed by this assumption. To increase the channel usage, our packet scheduling algorithm attempt to approximate future channel states in order to make a suitable decision for slot allocation. In addition, the packet scheduler requires queuing analysis of the wireless link to provide the QoS support for RT traffic.

### III. LINK LAYER CHANNEL MODEL

In HSDPA, the wireless transport capability can be different for different users and will change over time due to user mobility and channel fading. Furthermore, wireless transport capability is bursty. The transport segment per slot is much larger than the packet size in wired networks. One transport segment can transport a lot of incoming packets. All arriving packets ought to be put into transmission queues waiting for

scheduling. Thus, a work-conserving scheduler is not always optimal in contrast to the wireline case.

Under ideal conditions, right after there are enough packets in the transmission queue to send, the channel will be in the best condition. The subscriber is selected and packets will be transported at the best rate.

Our model is a statistical segment channel model. We assume that channel conditions of different users are independent. We divide the time of each user  $i$ , into coherent intervals, called *time segment*  $s$ , similar to the physical layer channel model used in [17], therefore:

$$T = \{[0, T_i], [T_i, 2T_i], [2T_i, 3T_i], \dots, [s \cdot T_i, (s+1)T_i]\}$$

The  $T_i$  is the size of one time segment or one scheduling cycle. The  $T_i$  of each user can be different and is depends on the characteristics of the user's physical channel. Every time segment  $s$  contains  $T_i$  time slots. We also use  $r_i(t)$ , in our model as amount of data can be transmitted to the RT user  $i$  at time  $t$  or feasible data rate at time  $t$ , if user  $i$  is chosen. In the other word,  $r_i(t)$  is the transport segment size per slot that is determined by the CQI report of user  $i$  at time  $t$ . It is distinct from physical layer parameters such as signal to noise ratio that usually used in the channel-fading model. The mapping between reported CQI values and feasible rates for different user categories are defined in the 3GPP standard. In each time segment  $s$ , the feasible data rate of each user  $i$  varies over time which is denoted by a  $T_i \times I$  vector called  $r_i^s(t)$ , where:

$$t = s \cdot T_i, s \cdot T_i + 1, \dots, s \cdot T_i + T_i - 1$$

In the other word,  $r_i^s(t)$  is the feasible rate vector of user  $i$  in time segment  $s$ . Statistical parameters, segment mean  $\mu_i^s$  and standard deviation  $\sigma_i^s$ , are used to represent the property of the feasible rate vector  $r_i^s(t)$  in time segment  $s$ .

$$\mu_i^s = E[r_i^s(t)] ; t \in [s \cdot T_i, (s+1) \cdot T_i] ; \forall s \geq 0$$

In proposed statistical segment channel model, the non-stationary property of fading channels can be characterized as follows. Within a time segment  $s$ , instant rate  $r_i(t)$ , ( $t = s \cdot T_i, s \cdot T_i + 1, \dots, s \cdot T_i + T_i - 1$ ), is fluctuating around a mean value  $\mu_i^s$  by a variance of  $(\sigma_i^s)^2$ . On the other hand, for a long term, the mean values  $\mu_i^s$  ( $p=0,1,2,\dots$ ) are time-varying which represents non-stationary channel conditions. Therefore, the segment standard deviation  $\sigma_i^s$  represents short-term channel fluctuations and variations of segment mean  $\mu_i^s$  represent long-term spatial non-stationary channel fluctuations.

### IV. PROPOSED SCHEME

A base station serving  $N$  real-time and  $M$  non-real-time users is considered. The base station transmits in slots of some fixed duration. It is assumed that the base station transmits to exactly one user in each slots (2 ms in HSDPA), although the base station can transmit to multiple users at the same time in HSDPA. However, it is easy to extend the scheme to multiple users. Fig. 1 illustrates the proposed architecture. NRT service includes *Interactive* and *Background* class. The goal is to support QoS to RT users, while NRT users are provided with best

effort services. RT users in 3GPP have been classified to *Streaming* and *Conversational*.

Scheduler component in proposed framework implement a cyclic APO algorithm, which will be described in the following section. To support QoS for RT services, scheduler gives higher priority to  $N$  real-time traffic. After serving real-time users, leftover capacity is allocated to  $M$  non-real-time users using the PF scheduling algorithm

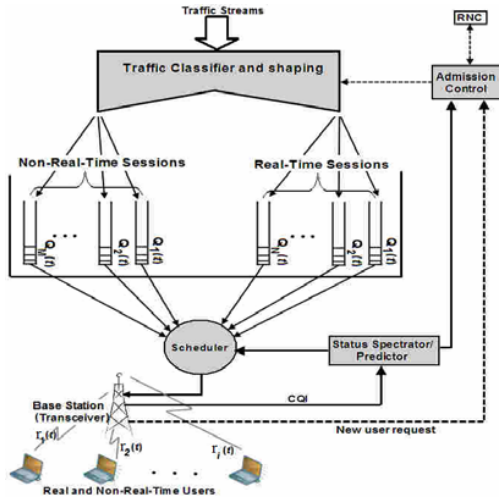


Fig. 1 Proposed scheme with Hybrid RT and NRT services in HSDPA

Channel status spectator unit receive the subscriber's Channel Quality Indicator (CQI) response in RT. In addition, it observes instantaneous channel status and anticipates near future channel capability. Channel state information will be used in the scheduling scheme and admission control module. Incoming traffic stream are separated and placed into different packet queues by traffic classifier. Each session has a queue that is fed by an arrival process. In a fair manner, the scheduler can delete waiting packets if they are over their delay due time so to keep input queues stable.

The duty of admission control is to avoid overload situations where the QoS contracts of RT services are broken. With AMC, the modulation type, the coding rate and number of spreading codes are adapted to instantaneous channel quality instead of adjusting power to control transmission rate. However, typical link admission control algorithms in WCDMA are power-based schemes. The introduction of HS-DSCH results in a new situation, where admission control must be able to handle multiple services on shared channels. Authors in [18] discusses the ineffectiveness of power-based admission control when HS-DSCH is developed.

After the resource reservation procedure at the admission control module, the base station knows the assured amount of bandwidth  $B_i$  for RT sessions.  $B_i$  is a constant value during the session. Due to the variation of bursty wireless data service, users can be served at the higher assured bandwidth when they are close to the base station. Therefore, an adaptive multimedia service is a better choice for HSDPA. If adaptive

multimedia service is supported,  $B_i$  may not be constant and be negotiated to be adjusted during the session.

#### A. Cyclic APO Scheduling

Based on the link-layer channel segmentation model proposed in the Section 2, a cyclic scheduling scheme is designed of the global scheduling approach used in traditional schemes. All arriving streaming packets are backlogged in transmission queues at the base station waiting for scheduling. The downlink packet schedule for the real-time user  $i$  can be considered as a cyclic task, which is characterized by a cycle  $T_i \in \mathbb{N}$  and bandwidth requirement  $b_i$ . User  $i$  expects to be allocated bandwidth  $b_i$  bits in every interval

$\{t \mid s.T_i \leq t < (s+1).T_i\}$  for each  $s$ . Here, the scheduling cycle  $s$  is correspondent with one time segment whose size is  $T_i$ . Since the slot length  $TTI$  is 2ms, in order to provide the assured amount of bandwidth  $B_i$  bps to the RT user  $i$ , the bandwidth requirement  $b_i = TTI \cdot T_i \cdot B_i$  in each cycle. For each user  $i$ , incoming packets arriving during the previous cycle  $s-1$  are backlogged in the buffer at the base station and will be scheduled to transport in the current cycle  $s$ . (Figure 2)

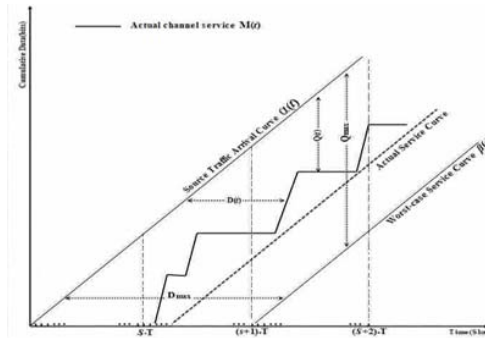


Fig. 2 Cycling scheduling

In the worst case, data may be scheduled to transmit in several first slots in the cycle  $s-1$  and in several last slots in the cycle  $s$ . Thus the maximum jitter delay of user  $i$  is:

$$J_{\max}(t) = (s+1) \cdot T_i - (s-1) \cdot T_i = 2 T_i$$

To guarantee jitter delay bound,  $\lambda_i$ , the cycle length of cyclic schedule  $T_i$  should be less than half of the bound  $\lambda_i$ , i.e.  $T_i \leq \lambda_i/2$ . Thus, if the required bandwidth  $b_i$  can be sufficed in each cycle and  $T_i \leq \lambda_i/2$ , the cyclic scheduling can provide user  $i$  assured bandwidth  $B_i$  and guaranteed jitter delay  $\lambda_i$ . Furthermore, the packet delay  $D_i(t)$  (Fig. 2) can be bounded by:

$$D_{\max}(t) = (s+1) \cdot T_i - (s-1) \cdot T_i = 2 T_i$$

At the beginning of every scheduling cycle  $s$  ( $\forall s$ )

$$Q_i(s.T_i) = b_i = B_i.TTI.T_i$$

The buffer size  $Q_i(t)$  in each cycle can be bounded by:

$$Q_{\max}(t) = Q_i(s.T_i) + B_i.TTI.T_i = 2B_i.TTI.T_i$$

If the required bandwidth  $b_i$  cannot be satisfied in the cycle  $s$ , the residual data  $\tau_i^s$  of the unsatisfiable cycle  $s$  will be left over to the next cycle  $s+1$  for transmission.

In each scheduling cycle, APO, is designed to enhance channel usage efficiency.

### B. Anticipated Proportionally Optimal (APO)

In this paper, we assume statistical properties of channel conditions in the upcoming time segment  $s_i$  can be estimated by the history channel information. Hence, segment mean  $\mu_i^s$  and standard deviation  $\sigma_i^s$  can be anticipated at the beginning of every time segment  $s$ . The channel anticipation requires to be deployed cyclically in order to track the non-stationary channel fluctuations for each user  $i$ .

Consider the scheduling in a cycle  $s$  where, for any instantaneous time  $t$  in the user  $i$ 's time segment  $s$ , where  $s_i \cdot T_i \leq t < (s_i+1) \cdot T_i$ , the feasible rate vector  $\vec{r}(t)$  is available by mapping from instantaneous reported CQI values. At instantaneous time  $t$ , the APO prefers the user

$$i^*(\vec{r}(t)) = \arg \max \frac{r_i(t) - \mu_i^s}{\sigma_i^s}; \quad 1 \leq i \leq N$$

$$\text{i.e. } x_i(t) = \{1, \text{ if } i = i^*; 0, \text{ if } i \neq i^*\}$$

$\sigma_i^s$  and  $\mu_i^s$  are anticipated statistical standard deviation and segment mean of the feasible rate vector  $V_{i,s}(t)$  of user  $i$  in its time segment  $s_i$ , which are constant within its time segment  $s$ .  $W(t) = \frac{r_i(t) - \mu_i^s}{\sigma_i^s}$  is a standardized score or normalized-

weight of instantaneous feasible rate  $r_i(t)$  which measures the number of standard deviations that  $r_i(t)$  falls from the expected mean rate of time segment  $s$ . Thus, the scheduling decision is made from instantaneous channel conditions and long-term statistically anticipated channel state in each user's time segment  $s$ . Higher  $W$ -weight of the instantaneous feasible rate represents that the user has anticipated proportionally better channel quality. When the feasible rate segment  $r_i^s(t)$  is regarded as a set of Gaussian samples, in future slot:

( $t < \text{future slot} < (s_i+1) \cdot T_i$ ). Within time segment  $s$ , the probability that the user  $i$  has a proportionally better channel quality than that at instantaneous time  $t$  is given by:

$$P(r > r_i(t)) = P(W > \frac{r_i(t) - \mu_i^s}{\sigma_i^s}) =$$

$$1 - P(W \leq \frac{r_i(t) - \mu_i^s}{\sigma_i^s})$$

The relationship between  $W$ -weight and their probabilities can be expressed as : (" $\approx$ " denotes "proportional to"):

$$P(W \leq \frac{r_i(t) - \mu_i^s}{\sigma_i^s}) \approx \frac{r_i(t) - \mu_i^s}{\sigma_i^s}$$

Therefore,

$$P(r > r_i(t)) \approx (-1) \frac{r_i(t) - \mu_i^s}{\sigma_i^s}$$

Finally, we can conclude:

$$i^*(\vec{r}(t)) = \arg \min P(r > r_i(t)); \quad 1 \leq i \leq N$$

Thus the APO algorithm prefers user  $i^*$  who has the least probability to get better channel condition at a future slot than that at the current slot. The consequence is that each user transmits only on a good instantaneous rate, which is close to the maximum channel access in its current time segment. To analysis the scheduling strategy for one user, the feasible rate modeled as Gaussian random variable. The user  $i$  can be scheduled for transmission only when its feasible rate is proportionally highest in its current time segment. During the scheduling cycle  $s_i$ , RT user  $i$  can compete for the service at each slot until all its required amount of data  $b_i$  has been transported.

The scheduling criterion to make a decision between multiple users in *anticipated proportionally optimal* which compare each user's current channel quality with its own future conditions. For example, at slot 5 and slot 10 in Figure 4, user1 has a better absolute channel condition than user2. However, while comparing the instantaneous channel quality to their long-time conditions, user2 has proportionally optimum channel condition than user1 at these two slots. Thus, user2 is chosen at slot 5 and slot 10 by the APO algorithm. The APO prefers the user who has the least probability to get a better channel condition at a future slot than that the current a lot to achieve higher channel utilization.

From RT users' point of view, APO is a non-work-conserving scheduling scheme. The RT user  $i$  will be waiting for proportionally optimal channels within its time segment  $s$  for transmission.

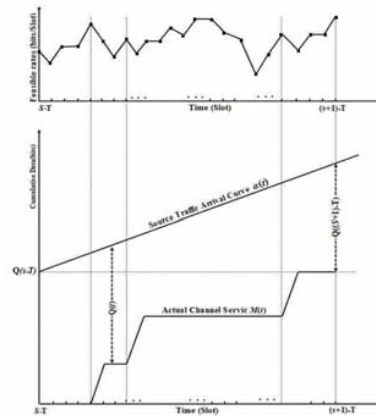


Fig. 3 APO non-work-conserving scheduling for one user

From the system's point of view, the scheduler selects to serve the one with proportionally highest feasible rate. In this way, the scheduler will assign as few slots as possible to one RT user to satisfy its bandwidth requirement so as to serve more users and improve system efficiency.

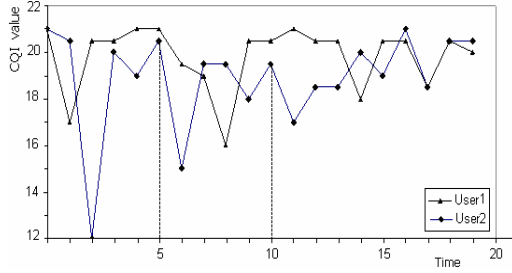


Fig. 4 Anticipated Optimal scheduling for multi users

### C. Effective Channel Rate

In order to improve the effectiveness of APO algorithm, the effective channel rate scheme is deployed for scheduling between RT users and NRT users. Each RT user  $i$  has defined an effective channel rate  $\eta_i^s$  for each time segment  $s$ . Only when the instantaneous feasible rate  $r_i(t)$  is above its effective channel rate, i.e.  $r_i(t) \geq \eta_i^s$ , can RT user  $i$  compete for the slot allocation. If instantaneous feasible rates of all RT users are less than their effective channel rates, the slot will be left over to serve NRT users. For each user  $i$ , at the beginning of its time segment  $s$ , the scheduler needs to anticipate the effective channel rate vector  $\eta_i^s$  by tracking the previous optimal assignment vectors  $\vec{x}(t)$  ( $\forall t < s \cdot T_i$ ). Previous optimal assignment vectors can be computed by using some offline scheduling algorithms. Estimation schemes such as the ML method [17] can be adopted to estimate the effective channel rate. Here we give a simple solution for slow fading channels by tracking  $h_i$  previous time segments. First, we calculate the effective channel rate  $\eta_i^s$  in  $h_i$  previous time segments.

$$\eta_i^l = \text{Min}_{l-T_i \leq t < (l+1) \cdot T_i} (r_i(t) \cdot x_i(t)), \text{ for } (s - h_i) \leq l < s; \text{ where } x_i(t) = 1;$$

means the time slot  $t$  is assigned to the user  $i$  to transmit data.

Next, we approximate  $\eta_i^s$  by (17):

$$\eta_i^s = \sigma_i^s \cdot \text{Min}_{s-h_i \leq l \leq s} \frac{\eta_i^l - \mu_i^l}{\sigma_i^l} + \mu_i^s$$

Where  $h_i$  depends on the long-term channel variation of user  $i$ . The larger is the variation, the smaller is  $h_i$ .

Note that the APO algorithm cannot guarantee that all users' time segments are satisfiable, because estimations of effective channel rate  $\eta_i^s$  and anticipated  $\eta_i^s$ ,  $\sigma_i^s$  are not accurate due to the unanticipated variation of future channel conditions. If the segment is unsatisfiable under overload situations, the residual data  $\tau_i^s$  will be left over to next time segment  $s+1$  for transmission. Proportionally smaller  $\eta_i^s$  will decrease  $\tau_i^s$  but will also reduce the channel usage efficiency.

## V. SIMULATION RESULTS

The performance of cyclic APO algorithm is compared with two popular scheduling algorithms. MLWDF is a common RT scheduling algorithm for EDGE/HSDPA/HDR and PF is prevalent algorithm for NRT services. A HSDPA base

station consists of twelve attached users, six users with RT and six NRT users is implemented. Twelve CQI sets are created by the CQI generator for all users, each set contains 2000 CQI values which represent the variation of channel conditions. The non-stationary channel condition for each user, i.e.  $r_i(t)$  ( $\forall i$ ), is simulated by time-varying centered complex Gaussian random variables. The initial mean CQI values  $\mu_i^0$  in time segment 0 are as follows:

$\mu_1^0$	$\mu_2^0$	$\mu_3^0$	$\mu_4^0$	$\mu_5^0$	$\mu_6^0$
2191	1480	4179	3559	14580	15438
bits/slot	bits/slot	bits/slot	bits/slot	bits/slot	bits/slot

To simplify the channel estimation function in the simulation, the mean feasible rate  $\mu_i^s$  of generated CQI sets increases or decreases in a random linear function.

We compare the channel usage efficiency of proposed scheme in sample slot assignment for one user at one scheduling cycle with MLWDF and optimal offline scheduling algorithms. For the sake of simplicity and clarity, the result for one user is shown in Fig. 5 but similar comparison can be generated for other users at all scheduling cycle. Fig. 5 presents the assignment vector  $\vec{x}(t)$  ( $100 \leq t < 200$ ) of the real-time session in time segment by these three algorithm. The feasible rates in the time segment are simulated by Gaussian random variables.

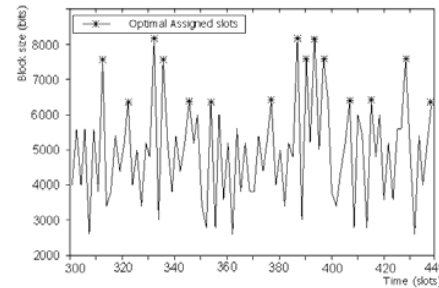


Fig. 5 (a) Assigned slots in OOS

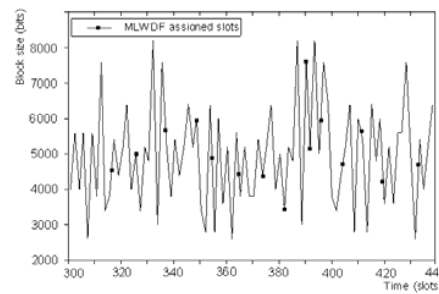


Fig. 5 (b) Assigned slots in MLWDF

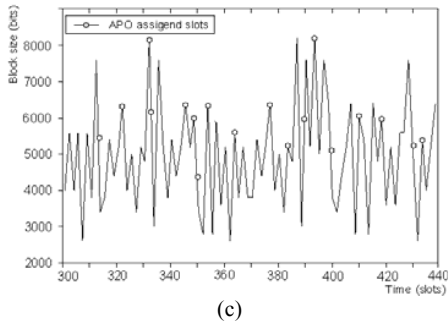


Fig. 5 (c) Assigned slots in APO

As it is indicated, the Optimal Offline Scheduling (OOS) provides the best channel usage for assigned serving slots, i.e. the RT user is always scheduled at its highest feasible rates.

Fig. 6 shows the comparison of average feasible rates of assigned slots for all RT users during the entire simulation time by these three algorithms.

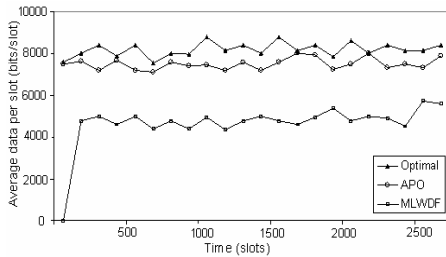


Fig. 6 Channel efficiency comparison for RT users

As it can be observed from Fig. 5 and Fig. 6, the APO algorithm provides sub-optimal channel usage by choosing the proportionally optimal channel condition that is a corollary of non-work-conserving feature of our real-time APO algorithm. In the APO algorithm, the history channel conditions are used to estimate the anticipated mean feasible rates  $\mu_i^s$ , the standard deviation  $\sigma_i^s$  and the effective channel rate  $\eta_i^s$  in the current time segment  $s$ . The inaccurate anticipate of these parameters can lead to non-optimal slot assignment.

In addition, due to the unanticipated variation of future channel conditions, the on-time scheduler cannot wait for future optimum channel conditions to serve users as with the optimal off-line algorithm, but can serve users in anticipated proportionally optimal channel conditions.

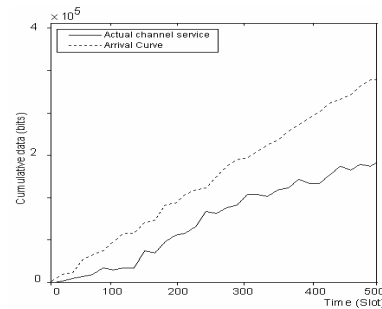
We analyze the QoS performance through the comparison of APO, PF and WLWDF algorithms using *Service Curve*. The actual channel service curves of the RT users have been shown in Fig. 7. It can be observed that non-real-time PF algorithm cannot provide the bounded service curve. Namely, it cannot provide the bounded delay and buffer size that increase with time. In the other word, the PF algorithm does not consider QoS assurance metric. The user with better channel has a higher chance to get the service so that the delay and bandwidth of the users cannot be guaranteed.

The WLWDF algorithm provides less delay and a smaller buffer size due to its work-conserving characteristic. How-

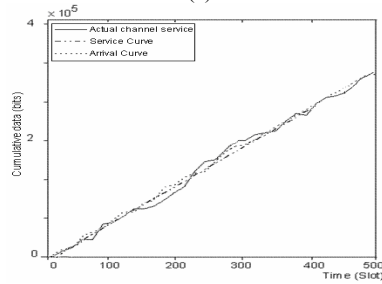
ever, its real-time traffic load  $\frac{Total_{RT-slot}}{T}$  is always above

97%, which results in the starvation of the NRT users. This is due to its low channel usage efficiency.

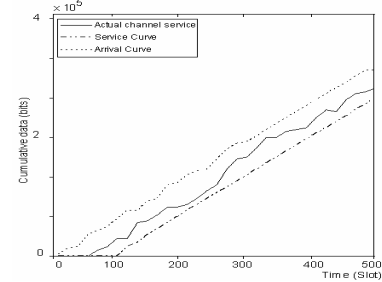
Fig. 7(c) shows the combination of the service curve guarantee with the *Arrival Curve* constraint forms deterministic bounds on the delay and buffer size of our cyclic algorithm's channel service. The maximum delay, jitter and buffer size are usually much less than  $D_{max}(t)$ ,  $J_{max}(t)$  and  $Q_{max}(t)$  that were analysis in the previous section. Therefore, even if there exist some unsatisfiable time segments under overloaded situations, the delay and the buffer size are still in the range of bounded values.



(a)



(b)



(c)

Fig. 7 QoS performance comparison in (a) PF, (b) MLWDF and (c) APO

## VI. CONCLUSION

The quality of packet scheduling is a crucial issue that affects directly the operation of HSDPA since it controls the distribution of scarce radio resources among mobile users. In this article, a QoS support wireless packet scheduling scheme for hybrid real and non-real-time services for HSDPA networks has been developed. Proposed method is a time segmentation-based algorithm that implements a non-



conserving scheduling scheme. The key idea behind cyclic APO is that it prefers the user who has the least probability of getting an optimal channel state at a future slot than the current slot.

Our experimental results demonstrate that employing cyclic APO can maximize channel capacity with QoS support for users with RT service. The inefficiency of the channel anticipation model will decrease the performance of the APO scheduling algorithm. Applying further method to improve channel estimation is a topic for the future research.

#### REFERENCES

- [1] Online Available: <http://www.3gpp.org>.
- [2] Nokia. White Paper. Nokia HSDPA Solution. Online Available at: <http://www.nokia.com>. 2005.
- [3] Shakkottai, S. and R. Srikant, Scheduling Real-Time Traffic With Deadlines over a Wireless Channel, *Wireless Networks*, vol. 8, no.1, pp. 13-26, 2002.
- [4] Ameigeiras, P., Packet Scheduling and Quality of Service in HSDPA, Ph. D. Thesis, University of Aalborg (Denmark), October 2003.
- [5] Al-Manthari, B., N. Nasser, and H. Hassanein. Fair Channel Quality-Based Scheduling Scheme for HSDPA System, *Proceedings of the IEEE International Conference on Computer systems and applications: IEEE Computer Society Washington, DC, USA*, 2006.
- [6] Berggren, F. and R. Jantti, Asymptotically fair transmission scheduling over fading channels, *Wireless Communications, IEEE Transactions on*, vol. 3, no.1, pp. 326-336, 2004.
- [7] Jiang, Z., Y. Ge, and Y. Li, Max-utility wireless resource management for best-effort traffic, *Wireless Communications, IEEE Transactions on*, vol. 4, no.1, pp. 100-111, 2005.
- [8] Andrews, M., K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar, and P. Whiting, SCHEDULING IN A QUEUING SYSTEM WITH ASYNCHRONOUSLY VARYING SERVICE RATES, *Probability in the Engineering and Informational Sciences*, vol. 18, no.02, pp. 191-217, 2004.
- [9] Shakkottai, S. and A. Stolyar. Scheduling algorithms for a mixture of real-time and non-real-time data in HDR, *Proceedings of 17th International Teletraffic Congress (ITC-17)*, 2001.
- [10] Chang, K. and Y. Han. QoS-based adaptive scheduling for a mixed service in HDR system, *Personal, Indoor and Mobile Radio Communications 2002. The 13th IEEE International Symposium on*, 2002.
- [11] Shao, H. R., C. Shen, J. Zhang, D. Gu, and P. Orlik, Dynamic resource control for high-speed downlink packet access wireless channels, *Proceeding of 23rd International Conference on Distributed Computing Systems Workshop*, pp. 838-843, 2003.
- [12] Chaporkar, P. and S. Sarkar. Providing stochastic delay guarantees through channel characteristics based resource reservation in wireless network, *Proceedings of the 5th ACM International Workshop on Mobile Multimedia WoWMoM*, Atlanta: ACM Press New York, NY, USA, 2002.
- [13] Wu, D. and R. Negi, Effective capacity: a wireless link model for support of quality of service, *Wireless Communications, IEEE Transactions on*, vol. 2, no.4, pp. 630-643, 2003.
- [14] Arauz, J., P. Krishnamurthy, and M. A. Labrador, Discrete Rayleigh fading channel modeling, *Wireless Communications and Mobile Computing*, vol. 4, no.4, pp. 413-425, 2004.
- [15] Turin, W. and R. Van Nobelen, Hidden Markov modeling of flat fading channels, *Selected Areas in Communications, IEEE Journal on*, vol. 16, no.9, pp. 1809-1817, 1998.
- [16] Pimentel, C., T. H. Falk, and L. Lisboa, Finite-state Markov modeling of correlated Rician-fading channels, *Vehicular Technology, IEEE Transactions on*, vol. 53, no.5, pp. 1491-1501, 2004.
- [17] Dogandzic, A. and J. Jin, Estimating statistical properties of MIMO fading channels, *Signal Processing, IEEE Transactions on* vol. 53, no.8 Part 2, pp. 3065-3080, 2005.
- [18] Hiltunen, K., M. Lundevall, S. Magnusson, E. Res, Olme Ab, and F. Jorvas. Performance of link admission control in a WCDMA system with HS-DSCH and mixed services, *15th IEEE International Symposi-*

um on Personal, Indoor and Mobile Radio Communications, PIMRC 2004. , 2004.

- [19] Marzetta, T. L., EM algorithm for estimating the parameters of a multivariate complex Rician density for polarimetric SAR, in *Acoustics, Speech, and Signal Processing, ICASSP-95*, vol. 5, pp. 3651-3654, 1995.



**MohammadReza HeidariNezhad** obtained his B.Sc and M.Sc in Computer Engineering field in 1996 and 1999, respectively. He is currently Ph.D. candidate in department of Communication Technology and Network, Faculty of Computer Science and Information Technology in the University Putra Malaysia (UPM). His research interests include resource and traffic engineering, mobility management and 3G/4G wireless network.