

# MODIFIED QUEUE-BASED EXPONENTIAL RULE SCHEDULER FOR IMPROVED QOS IN OFDMA SYSTEMS

C.Kalyana Chakravarthy<sup>1</sup> and Prof. P.V.G.D. Prasad Reddy<sup>2</sup>

<sup>1</sup>Department of CSE, M.V.G.R.College of Engineering, Vizianagaram, India  
[kch.chilukuri@gmail.com](mailto:kch.chilukuri@gmail.com)

<sup>2</sup>Dr. P.V.G.D.Prasad Reddy, Department of CS&SE, Andhra University College of Engineering, Visakhapatnam, India  
[prof.prasadreddy@gmail.com](mailto:prof.prasadreddy@gmail.com)

## ABSTRACT

*In our paper, we propose a modification of queue based exponential scheduler that dynamically improves the channel gain of far away users (by increasing their credits) at the cost of a small reduction of the same for the nearby users (by reducing the credits), since users nearby anyway have a fairly high gain. Results clearly suggest that the proposed method is able to maintain a constant average throughput (always greater than that of the P.F scheduling) while it far outperforms the P.F and Queue-Based Exponential Rule Scheduler in the delay characteristics. The throughput fairness index was interestingly observed to improve with increasing number of users while it worsens for the other two schemes.*

## KEYWORDS

*scheduling; quality of service; latency;*

## 1. INTRODUCTION

In a multi-user variable channel model, a scheduling algorithm can take advantage of channel variations by giving some form of priority to users with (temporarily) better channels. Since channel capacities (service rates) of different users vary in time in an asynchronous manner, the QoS of all users can be improved compared to scheduling schemes that do not take channel conditions into account. However, there exists a trade off between average throughput and fairness of scheduling between users. Users who are closer to the base station continuously experience better channel conditions while users who are far away are constantly experience lower channel gains. On the other hand, improving fairness, results in increasing scheduling opportunities to users with lower channel gains which in turn results in a degradation of throughput of the System. Thus a proper resource allocation mechanism needs to strike a balance between the two factors. Several classifications of scheduling algorithms are available in literature.

Scheduling falls into two categories - throughput optimal scheduling and delay optimal scheduling. While throughput optimal scheduling policies concentrate on maintaining the stability of the queuing system, they do not attempt to minimize the queue length or the delay. M-LWDF, EXP scheduling are examples of this category. Delay optimized scheduling policies optimize the rate and power so as to minimize the queue length. Policies that attempt to control rate and power separately may fall into this category.

Popular algorithms for NRT packet scheduling include Maximum Carrier to Interference (Max C/I), Round Robin (RR), Proportionally Fair (PF) and Fast Fair Throughput (FFTH) [1].

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)[2]. Exponential Rule (EXP)[3], Modified Exponential Rule [4].

A non real time content scheduling algorithm has been proposed in [5] for elastic applications with no per-packet delay constraints. Non-real time content is differentiated and scheduled using the concept of stretch factor. The stretch factor for content is defined in terms of the increase in the download time of content in a loaded system with respect to the download time in an ideal situation where the entire capacity is allocated to that content.

Further, scheduling algorithms can either be long term fair or short term fair, depending on whether it provides a fair share of throughput to all users over a long period of time or over few time slots respectively. The fairness index of a user 'i' can thus be defined as  $FI = (th_{max} - th_{min}) / th_{max}$  where  $th_{min}$  and  $th_{max}$  are the minimum and maximum throughput values over an interval of time 't'.

Various long-term fairness criteria, such as proportional fairness, temporal fairness, and utilitarian fairness, have been studied for scheduling problems in wireless networks. However, there is a need for general short-term fairness criteria tailored to wireless networks and dealing with the short-term performance in depth. References related to the subject include [6], [7], [8], [9], [10], where queueing status is a part of scheduling decisions.

## 2. PREVIOUS WORK

### 2.1 CHANNEL INDEPENDENT SCHEDULING ALGORITHMS

Round Robin (RR): The Round Robin scheduler periodically serves all terminals in a cyclic manner. It is the prototype of an algorithm equally distributing the transmission times among all terminals. In a homogeneous setup with constant data rate demands, it is expected to provide the best timing performance in terms of a relatively low and constant delay. However, the disadvantage is unfairness.

Power Based Round Robin (PBRR): The Power Based Round Robin scheduler is an adaption of the RR principle aiming at a fair distribution of the transmission power. For this purpose, it sums up the transmission power attributed to each terminal and serves the one that has currently received the smallest share of power. This method has better throughput because a terminal with poor conditions draws more power during a particular scheduling cycle. The disadvantage is a poor timing performance.

The earliest deadline-first(EDF) Scheduler

Let  $\mu_i$  be the state of the channel of user  $i$  at time  $t$ , i.e, the actual rate supported by the channel. This rate is constant over one slot. Let  $\mu_i$  be the rate corresponding to the mean fading level of user  $i$  and  $W_i(t)$  be the amount of time the HOL packet of user  $i$  spends in the base station. The FIFO discipline schedules the user whose HOL packet has been at the base station for the largest time first..Thus, at time  $t$ , then user  $u = \arg \max W_i(t)$  is scheduled first. However, this method is channel unaware and its performance is very poor under variable channel conditions. The maximum rate rule schedules the user whose channel supports the largest data rate  $\mu_i$  in the next slot. The user  $u = \arg \max \mu_i(t)$  is scheduled first.

Wireless fair queueing schemes [11, 12, 13] are aimed at applying wired fair queueing to wireless networks. The objective of these schemes is to provide fairness, while providing loose QoS guarantees. Although these schedulers make decisions based on the channel state information (i.e., good or bad channel), they do not exploit asynchronous channel variations to improve efficiency since packets destined to different users are transmitted at the same bit-rate.

## 2.2 CHANNEL AWARE SCHEDULING ALGORITHMS

Channel-aware schedulers, also referred to as opportunistic algorithms, take into account the channel quality in order to increase the performance of a wireless system. In opportunistic scheduling, users with good channel are given priority over users with poor channel conditions. This is because the number of retransmissions on a poor channel may be more resulting in a wastage of resources. Terminals omitted for being in bad channel conditions may be compensated later.

### Maximum SINR (maxSINR)

This scheduling algorithm serves in every TTI the user with best channel conditions and, therefore, the highest instantaneous supportable data rate. The serving principle has obvious benefits in terms of cell throughput. Consequently, under idealized conditions it is the system throughput optimal scheduler. Mathematically seen, it schedules user

$J = \arg \max_i \{R_i(t)\}$  at time  $t$ .  $R_i(t)$  is the instantaneous data rate experienced by user  $i$  if it is served by the packet scheduler. The main disadvantage of this approach is the inherent unfairness. For instance, when a User Equipment (UE) is far away from the base station and its mobility is low, it may never be scheduled.

The M-LWDF algorithm [14] and the opportunistic transmission scheduling [15] optimize a certain QoS parameter or utility index. They both exploit asynchronous channel variations and allow different user to transmit at different bit-rate or signal-to-interference-noise ratio (SINR), so that higher efficiency can be achieved. However, they do not provide the explicit QoS guarantees such as data rate, delay bound, and delay-bound violation probability.

### Modified Largest Weighted Delay First

The Modified Largest Weighted Delay First (M-LWDF) algorithm was proposed by Andrews et al. in [16]. This algorithm attempts to keep the probability of the queuing delay exceeding the due time below a certain ratio all the while trying to utilize the wireless channel efficiently. The M-LWDF computes the priority of user  $i$  at every TTI as follows:

$$SF_i(t) = -\log(p_i) \cdot \frac{R_i(t) \cdot Q_i(t)}{s_i(t) \cdot T_i}$$

where  $T_i$  expresses the discard timer parameter for user  $i$  and  $p_i$  denotes a QoS parameter that allows to differentiate between users with different QoS requirements (such as end-to-end delay). Moreover,  $p_i$  can be considered as the probability of a user to exceed their delay requirements:

$$\Pr(Q_i > T_i) \leq p_i$$

A user with a higher value of  $p_i$  has a higher probability of exceeding its delay requirement.

A problem with this algorithm is that the packet with the highest priority receives the largest possible data rate and is transmitted as fast as possible. The data rates may therefore vary widely, depending on how much bandwidth is available at that moment. The scheduler prioritizes the packets in the same manner as the M-LWDF algorithm does, but the bandwidth is distributed using a General Processor Sharing (GPS) approach.

### Proportional Fairness Scheduling

The PF scheduler [17,18] is designed to take advantage of multiuser diversity, while maintaining comparable long-term throughput for all users. Let  $R_k(t)$  denotes the instantaneous

data rate that user k can achieve at time t, and  $T_k(t)$  be the average throughput for user k up to time slot t. The proportional fairness scheduler selects the user, denoted as k, with the highest

$$R_k(t)/T_k(t) \text{ for transmission. } T_k(t+1) = \begin{cases} (1-\frac{\epsilon}{t_c})T_k(t) + \frac{\epsilon}{t_c}R_k(t) & k=k^* \\ (1-\frac{\epsilon}{t_c})T_k & k \neq k^* \end{cases}$$

However, the presented scheduling methods do not take into account the delay experienced by each individual user. As a result, they are not suitable for scheduling of RT services.

Score Based (SB): The Score Based scheduler [19] has been conceived in order to overcome two issues of the basic PF algorithm: The preferential treatment of flows with small data rates, and the bias against variable radio channels in asymmetric conditions observed in [20]. The scheduling decision is based on the rank of the currently achievable data rate among the values in a recent time window of size N. TTTI:

$$tag_i = 1 + \sum_{t=1}^{N-1} \begin{cases} X_t & \text{if } R_k(t) > R_k(t-1-TTTI) \\ X_t & \text{if } R_k(t) = R_k(t-1-TTTI) \end{cases}, 0 \text{ otherwise}$$

where  $X_t$  are i.i.d. random variables on {0,1} with  $P\{X_t = 0\} = 1/2$ . The tags are therefore independent of the history of scheduling decisions.

#### TCP – Proportional Fair (TCP-PF)

The TCP Proportional Fair scheduler [21] is an adaption of the PF algorithm suitable to TCP traffic on wireless links. Generic channel aware schedulers provide no bounds on ISTs, possibly resulting in strongly varying packet delays. Excessive delays may erroneously provoke TCP timeouts eventually triggering the protocol's slow start mechanism, and therefore reduce the achievable throughput. To avoid such spurious timeouts, the TCP-PF scheduler introduces a correction factor limiting inter scheduling gaps. The scheme aims at improving the performance on the TCP layer by estimating transport layer metrics from measurements available on the link layer. Applying the PF update function (Eq. (2)), the scheduling tag is calculated as:

$$tag_k = \frac{R_k(t)}{\bar{R}_k(t)} \left( \frac{1}{2} \frac{\sum_k \bar{B}_k(t)}{\bar{\Phi}_k(t)} + \frac{\bar{B}_k(t)}{\bar{R}_k(t)} \right)$$

PF tag

where  $B_k(t)$  is the sliding average of the transmission buffer level of flow k at time t.  $\bar{B}_k(t)$  is the estimation of the average of the IST in a sliding time window,  $\sum_k \bar{B}_k(t)$  its second moment.

#### Expo-Linear (EL)

To avoid the intrinsic delay in M-LWDF, some other algorithms have been proposed. One of the examples is the Expo-Linear algorithm proposed in [22]. It schedules user

$$J = \arg \max_i \left\{ a_i \cdot \frac{R_i(t)}{\lambda_i(t)} \cdot e^{-a_i D_i(t)} \right\} \times \text{ where the } a_i, R_i(t), \lambda_i(t) \text{ and } D_i(t) \text{ have the same meaning as}$$

in M-LWDF mentioned above. This algorithm introduces an exponential term to better equalize the weighted delay.

#### Exponential Rule

The Exponential Rule (ER) scheduler [23] is a modified version of the PF scheme which has been customized for scheduling RT applications.

The ER algorithm prioritizes users based on the following formula:

$$SF_i(t) = a_i \cdot \frac{R_i(t)}{s_i(t)} \cdot \exp\left(\frac{a_i Q_i(t) - aQ(t)}{1 + \sqrt{aQ(t)}}\right) \text{ where } aQ(t) = \frac{1}{n} \sum_{i=1}^n a_i Q_i(t)$$

Where  $a_i$  is the priority value used to characterize the desired QoS and n is the total number of users. In order to better understand the ER scheme, let us consider the exponent term. According to the exponent, if the difference between a users' prioritized delay and the prioritized delay of

all users is greater than  $\sqrt{aQ(t)}$  the exponent term will dominate and exceed the impact of channel variations. If the difference in delays is less, then  $SF_i(t) = a_i \cdot \frac{R_i(t)}{S_i(t)}$

#### Queue-Based Exponential Rule Scheduler

Users having higher variance in queue length compared to average value are given more preference.

$$SF_i(t) = a_i \cdot \frac{R_i(t)}{S_i(t)} \cdot \exp\left(\frac{a_i Q_i(t) - \overline{aQ(t)}}{1 + \sqrt{aQ(t)}}\right) \cdot \exp\left(\frac{q_i(t) - \overline{q(t)}}{1 + q(t)}\right)$$

where  $\overline{aQ(t)} = \frac{1}{n} \sum_{i=1}^n a_i Q_i(t)$   
and  $\overline{Q(t)} = \frac{1}{n} \sum_{i=1}^n q_i(t)$

Where  $q_i(t)$  is the queue length of user  $i$  at the beginning of the  $t$ th time interval. The second exponent term in the above equation is used to balance the service queue length among multiple users as it gives a higher priority to users that have a high variance in queue length when compared to the average value. The authors argue that the magnitude of the second exponent term does not exceed that of the first exponent term as the denominator of the second exponent term does not comprise of a square root term. Simulation results in [24] show that this scheme achieves a higher fairness index when compared to the ER scheme at the cost of slightly degraded performance in terms of throughput and queuing delay.

### 3.MODIFIED QUEUE-BASED EXPONENTIAL RULE SCHEDULER

However, a problem inherent with the Queue-Based Exponential Rule Scheduler is that when there is a large variance in queue length, users who have higher priority are always benefited while other users are not given an opportunity. One solution to this problem could be to reduce some credits of top 'n' users whose variance is consistently high and increase the credits of the 'n' users who continuously suffer lack of opportunities due to poor channel conditions. In other words, we reduce the number of subcarriers allocated to some users regularly and increase the same for deprived users. To prevent some users from continuously depriving other users of subcarriers, there is a credit limit beyond which subcarriers cannot be borrowed or lent. This has inherent fairness built into it since the channel conditions are varying and a channel that is good for some time may later turn out to be a bad one, on the long run.

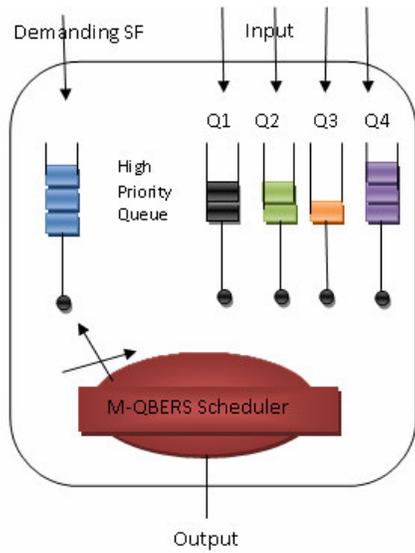


Figure 1. Modified Queue-Based Exponential Rule Scheduler

The algorithm consists of two parts, subcarrier allocation which determines the number of subcarriers to be assigned to every user 'i' and subcarrier assignment algorithm which finds the subcarrier assignment that improves fairness between the users. The subcarrier allocation is assumed to be naïve following the first available first with higher priority for real time flows compared to non-real flows, while the subcarrier assignment we propose is based on modification of the Queue-Based Exponential Rule Scheduler. Both these procedures have been combined to a single algorithm.

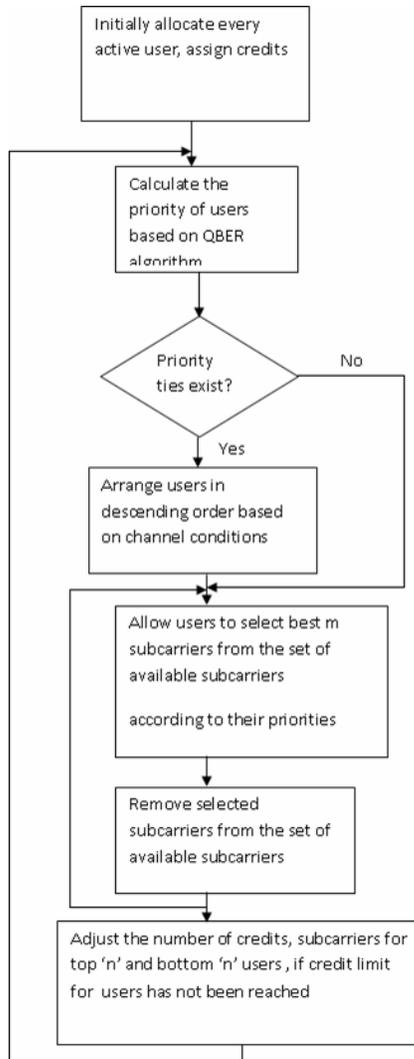


Figure 2. Scheduling in Modified Queue-Based Exponential Rule Scheduler

#### 4. THE SIMULATION MODEL

We consider the downlink of an OFDMA system with  $N$  subchannels and  $K$  users. The time axis is divided into frames. A frame is further divided into  $S$  time slots, each of which may contain one or several OFDM symbols. The duration of a frame is set to be 5 ms, thus we can assume that the channel quality remains constant within a frame, but may vary from frame to frame. Two separate queues, a RT queue and a NRT queue, are maintained for each user at the base station (BS). Each queue has a finite length of  $L$  packets. On arriving at the BS, the packets are buffered in the corresponding queue and are served by the BS scheduler. We assume that packets are of fixed length by fragmentation in the MAC layer and each packet contains  $d$  information bits. For the purpose of simulation, 1024 subcarriers, 1024 particles, fading channel with 50 users has been considered. The output Bandwidth is assumed to be 10 MHz and the QPSK  $\frac{1}{2}$  type modulation scheme is used. The inter-arrival time for different kinds of traffic was distributed exponentially with different mean values, while the packet size has a geometric distribution.

## 5. SIMULATION RESULTS

We have compared our algorithm with the existing PF scheduling algorithm and the Queue Based Exponential Rule Scheduler. Results clearly suggest that the proposed method is able to maintain a constant average throughput (always greater than that of the P.F scheduling) with increase in the number of users while clearly. Though throughput fairness is initially poor compared to the other two schemes, it improves with the number of users, unlike in the other two schemes, until from some point it betters them. This may be attributed to the fact that the credits accumulated by users with poor channel conditions and the credits reduced for the users with good channel conditions show a significant effect only on reaching certain lower threshold value beyond which an improvement in fairness is observed. However, the most significant improvement can be observed in terms of the difference between the maximum and minimum delays, thus the proposed scheduler has particular advantage in handling real time traffic with reduced latency making it suitable for applications like real time video-streaming.

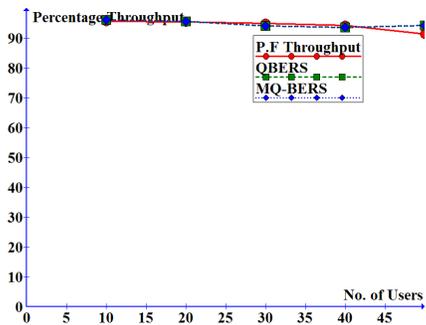


Figure 3. Number of users VS percentage Throughput utilization

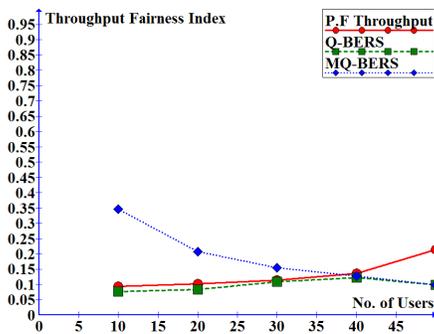


Figure 4. Number of users VS Throughput fairness Index

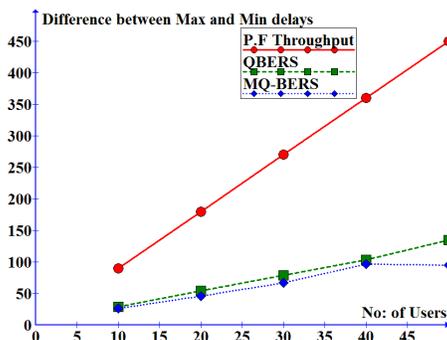


Figure 5. Number of users VS Difference between Max. and Min. Delays

## 6. CONCLUSIONS

In this article we discuss some key resource allocation and scheduling questions in IEEE 802.16 based wireless networks. More specifically, we consider the QoS-constrained scheduling problem and the joint scheduling and resource allocation problem for an 802.16, and present an algorithm that guarantees a reduced latency for high priority flows in the system. The area of QoS constrained scheduling and resource allocation is relatively new and has many applications in the next generation wireless networks. In the current work, satisfying Per-user demand constraints in the joint subcarrier and power allocation problem have been considered. Scheduling and power allocation problems in multi-cell scenarios remain open for future investigation.

## REFERENCES

- [1] Ameigeiras, P., Packet Scheduling and Quality of Service in HSDPA, Ph. D. Thesis, University of Aalborg (Denmark), October 2003.
- [2] Andrews, M., K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar, and P. Whiting, SCHEDULING IN A QUEUEING SYSTEM WITH ASYNCHRONOUSLY VARYING SERVICE RATES, *Probability in the Engineering and Informational Sciences*, vol. 18, no.02, pp. 191- 217, 2004.
- [3] 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.
- [4] 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.
- [5] Samrat Ganguly, Mainak Chatterjee, and Rauf Izmailov. Non-Real-Time Content Scheduling Algorithms for Wireless Data Networks, *IEEE TRANSACTIONS ON COMPUTERS*, VOL. 55, NO. 7, JULY 2006.
- [6] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *IEEE Communications Magazine*, vol. 39, no. 2, pp. 150–153, February 2001.
- [7] M. Andrews, S. Borst, F. Dominique, P. Jelenkovic, K. Kumaran, K. Ramakrishnan, and P. Whiting, "Dynamic bandwidth allocation algorithms for high-speed data wireless networks," *Bell Labs Technical Report*, 2000.
- [8] S. Shakkottai and A. Stolyar, "Scheduling algorithms for a mixture of real-time and non-real-time data in HDR," *Bell Laboratories Technical Report*, 2000.
- [9] S. Shakkottai and A. Stolyar, "Scheduling for multiple flows sharing a time-varying channel: The exponential rule," *Translations of the AMS*, 2001, A volume in memory of F. Karpelevich.
- [10] S. Shakkottai and A. Stolyar, "Scheduling of a shared a time-varying channel: The exponential rule stability," in *INFORMS Applied Probability Conference*, New York, July 2001.
- [11] S. Lu, V. Bhargavan, and R. Srikant, "Fair scheduling in wireless packet networks," *IEEE/ACM Trans. on Networking*, vol. 7, no. 4, pp. 473–489, Aug. 1999.
- [12] T. S. E. Ng, I. Stoica, and H. Zhang, "Packet fair queueing algorithms for wireless networks with location-dependent errors," in *Proc. IEEE INFOCOM'98*, pp. 1103–1111, San Francisco, CA, USA, March 1998.
- [13] P. Ramanathan and P. Agrawal, "Adapting packet fair queueing algorithms to wireless networks," in *Proc. ACM MOBICOM'98*, Oct. 1998.
- [14] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *IEEE Communications Magazine*, vol. 39, no. 2, pp. 150–154, Feb. 2001.
- [15] X. Liu, E. K. P. Chong, and N. B. Shroff, "Opportunistic transmission scheduling with resource-sharing constraints in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 10, pp. 2053–2064, Oct. 2001.
- [16] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing Quality of Service over a Shared Wireless Link," *IEEE Communications Magazine*, vol. 39, no. 2, pp. 150–154, 1996.
- [17] E. F. Chaponniere, P. Black, J. M. Holtzman, and D. Tse, "Transmitter directed Multiple Receiver System using Path Diversity to Equitably Maximize Throughput," Sept. 2002. U.S. Patent No. 6449490.
- [18] F. Kelly, "Charging and Rate Control for Elastic Traffic," *European Transactions on Telecommunications*, vol. 8, pp. 33–37, 1997.
- [19] T. Bonald, "A score-based opportunistic scheduler for fading radio channels," in *Proc. European Wireless 2004*, Barcelona, Spain, February 2004.
- [20] J. Holtzman, "Asymptotic analysis of proportional fair algorithm," in *Proc. 12th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2001)*, vol. 2, San Diego, CA, October 2001, pp. F.33 . F.37.
- [21] T. E. Klein, K. K. Leung, and H. Zheng, "Enhanced scheduling algorithms for improved TCP performance in wireless IP networks," in *Proc. IEEE GlobeCom*, Dallas, TX, December 2004.
- [22] A. Gougousis and M. Paterakis, "Scheduling with QoS support for multirate wireless systems with variable channel conditions," 2005, technical University of Crete, Greece.
- [23] S. Shakkottai and A. Stolyar, "Scheduling Algorithms for a Mixture of Real-Time and Non-Real-Time Data in HDR", *Proceedings of the 17th International Teletraffic Congress (ITC-17)*, Salvador de Bahia, Brazil, September 2001.
- [24] L. C. Wang and M. C. Chen, "Comparisons of link-adaptation-based scheduling algorithms for the WCDMA system with high-speed downlink packet access", *Canadian Journal of Electrical and Computer Engineering*, volume 29, no. 1/2, pp. 109-116, 2004.

## Authors

C.Kalyana Chakravarthy has a teaching experience of over nine years and is currently working as Associate prof. in M.V.G.R. College of Engineering, Vizianagaram. He has been actively working on diverse areas of network caching, MANETs routing protocols, resource allocation and scheduling in WiMAX, Mesh networks



Dr. P.V.G.D Prasad Reddy has a teaching experience of over twenty years. He is currently serving as the Registrar at the Andhra University. He has 16 publications in International Journals and 19 papers in conferences. His Research areas include Soft Computing, Software Architectures, Knowledge Discovery from Databases, Image Processing , Number theory & Cryptosystems.

