Channel- and delay-aware scheduling and packet dropping for real time traffic over WiMAX networks

Rudzidatul Akmam Dziyauddin*, Dritan Kaleshi, Angela Doufexi
Centre for Communication Research, Department of Electrical Engineering
University of Bristol, UK
{Rudzi.Dziyauddin, Dritan.Kaleshi, A.Doufexi@bristol.ac.uk}

Abstract—**In cases when system load approaches capacity (>90% of capacity), scarce radio resources are wasted when a scheduler serves video packets that exceed a required latency, as the decoder will ultimately discard any such packets. In this paper, we introduce a combination of a packet-dropping policy with a modified Greedy-based scheduler to guarantee meeting maximum latency requirements under heavy load conditions, whilst optimising the system goodput for the WiMAX real-time Polling service (rtPS) class. Results of the comparison with existing schedulers show that they all can successfully guarantee the required maximum latency after they have been extended by the same packet-dropping policy, with average latency values well below the required maximum latency. The proposed channel- and delay-aware scheduler results in the lowest number of packet drops. Results also show that mobility does not have significant impact on the system performance across the schedulers compared to the number of users in the cell.**

Keyword: scheduling; real-time Polling Service; packet dropping; Greedy; WiMAX

I. INTRODUCTION

Multimedia services have become popular these days with the emerging of wireless technologies and computing. One of the most popular multimedia-enriched applications is video, which is categorised as a delay-sensitive application. For the video to be of a good quality, it requires several hundred kilobits per second per user [1], with the latency and the packet losses less than 100 ms and 10% respectively [2] depending on the decoder type. A scheduler is one of the key components for designing QoS-oriented wireless networks [3]. However, the standard does not define a specific QoS scheduler to be employed at the Base Station (BS) and Subscriber Station (SS). Therefore, considerable work remains in new QoS scheduling that takes diverse QoS requirements into consideration for either uplink or downlink scenarios[4, 5]. In this paper, we focus a smart scheduling strategy for the real-time Polling service (rtPS) class, one of the five classes defined in the WiMAX standard [6].

To guarantee the required maximum latency of the rtPS class, a high complexity of the scheduling framework is required as found in [7-9]. This inspired a research, as in this paper, to investigate a simple technique by combining a packet dropping policy with the scheduling for heavy load scenarios. Only a few WiMAX scheduling studies [4, 5, 10] have combined the packet dropping policy with the scheduler, but [4, 5] do not consider for heavy load environments where the Quality of Service (QoS) becomes a problem. An interesting observation found by Nicolaou *et al.* [11] is that a Greedy scheduler yields a shorter average delay and optimises the utilization of radio resources. Our previous studies [12] show a similar observation for another scenario. This motivates us to improve the Greedy scheduler by taking into account a packet lifetime parameter besides the packet dropping mechanism, which differ from the study in [10].We also adopt the packet dropping policy to other present schedulers, namely, Proportional Fair (PF), Weighted Fair Queuing (WFQ) and pure Greedy since there is no previous WiMAX scheduling study has combined a packet dropping rule with schedulers in particular for the rtPS class. We also explore the impact of stationary and mobile users on the system performance when this smart scheduling strategy is used.

This paper is organised as follows: Section II presents the channel- and delay-aware scheduler. Section III shows the simulation model. Section IV discusses the simulation results and analysis. Section V concludes the paper.

II. CHANNEL- AND DELAY-AWARE SCHEDULER

Two types of attributes, namely, *packet latency* and *packet deadline* are often taken into consideration in techniques and methods designed to provide packet latency assurance. The packet latency, often referred to as packet delay, is the time difference between when a packet is sent from the source and when it arrived at the destination. Previous work also exploits the queuing time of a packet or using the EDF (Earliest-Deadline First) approach to compute the packet latency[5]. Meanwhile, the packet deadline is often based on the required maximum latency or the sum of the required maximum latency and other possible latency parameters such as packet arrival time at the BS [4, 5, 13]. When the packet approaches the maximum latency value, the chosen dropping policy (which packets are dropped from the queues) can make the radio resources to be used efficiently.

Another challenge for serving users in the wireless network is the time-varying channel where the user's channel can either be in good or poor quality at an instantaneous time. Every user is more likely to undergo different channel states during the time of transmission. These users' channel variations can be beneficial in

* Rudzidatul Akmam is also with Universiti Teknologi Malaysia International Campus Kuala Lumpur (UTMKL), Jalan Semarak, 54100 Kuala Lumpur, Malaysia. (e-mail: rudzi@ ic.utm.my).

978-1-4673-1881-5/12/$31.00 ©2012 IEEE

maximising the system throughput by serving the users with good channel conditions at every scheduling time. In real communication systems, packets arrive at the BS with a certain probability distribution that will lead to a variable queue length. It is thus very important for the scheduling algorithm to consider the queues status as well as the users' channel conditions, so that an optimal scheduler is designed [14]. However, very few studies [5, 10] consider all the three aspects, namely, channel-aware scheduling, packet latency, and the dropping policy, together in the WiMAX system. Therefore, the Greedy-based scheduler proposed in the next section takes all these three into account, and can then be classified as a channel- & delay-aware optimal scheduler. Modified Largest Weighted Delay First (MLWDF) [15] and Exponential Rule (EXP Rule) [15] are also known as a channel- & delay- aware schedulers, however, both schedulers do not consider packet dropping policy and also the required maximum latency parameter as mandated in the WiMAX standard. From here on we refer our proposed scheduler as Greedy-Latency.

A. Greedy-Latency Scheduler

The utility function of the Greedy-Latency scheduler is formulated to guarantee the required maximum latency, T_k, whilst optimising the system throughput. Assuming that there are *N* active users who have packets awaiting in the queues at the BS, the utility function of the scheduler is given as

$$U(d,T,\gamma,k) = \begin{cases} \arg\max_k\left(\frac{d_k}{T_k}\gamma_k\right)\ for\ \frac{d_k}{T_k} < 1 & \textbf{(1)} \\ Packet\ Drop\ \ for\ \frac{d_k}{T_k} \geq 1 & (2) \end{cases}$$

$$and\ T_k > 0, d_k > 0, \forall\ k\ \ \mathrm{k}\ =\ 1,2,3,..,\mathrm{N}$$

where γ_k and d_k denote the instantaneous SNR and the Head-of-Line (HOL) packet latency of user *k* respectively. The latter is computed either as

$$d_k(ms) = PacketScheduled\ Time - Packet\ Arrival\ Time\ in\ Queue \quad (3)$$

or,

$$d_k(ms) = Packet\ Scheduled\ Time - Packet\ Sent\ Time \quad (4)$$

In the first approach as in Equation (3), the parameter d_k measures the delay of the HOL packet waiting in the queue at the BS until it is served. In Equation (4), the delay is measured from the time when the HOL packet is sent from the sender (i.e. sending Subscriber Station (SS) /node) until when it is scheduled at the BS. The latter is the more accurate approach to measure the HOL packet latency, because it includes both the propagation time (from sender node to BS) and the queuing time in the BS. However, because it requires a good time synchronisation between the BS and the sender, it is less practical in real systems. In a single cell environment, which is the scope of this paper, the former approach is more realistic in measuring the HOL packet latency and is easier to implement. Therefore, in this study, the measurement of d_k follows Equation (3).

Let the ratio of d_k to T_k represent the *packet latency ratio*, α. Using this term, the pseudocode of the Greedy-Latency algorithm is summarised in Table I. Firstly, the parameters d_k is computed using Equation (3) and subsequently the parameter α equivalent to d_k/T_k is calculated. If the computed α is less than 1 (α <1 for non expired packets), the utility function is computed using Equation (1). Otherwise, the packet is considered to have expired and it is dropped from the queue. Next, the calculated utility values for all non-dropped HOL packets are sorted in a descending order. Finally, the packet that has the maximum utility value is then served by the scheduler. This process is repeated for the next HOL packets. Note that we assume FIFO queues in this study.

TABLE I. PSEUDOCODE OF GREEDY-LATENCY SCHEDULER

Begin
1. **for** *every head-of-line packet(or active user)* **do**
   - i. Compute d_k as per Equation (3);
   - ii. Compute $\alpha = d_k/ T_k$;
   - iii. **If** α < 1, **then**
     compute the utility function using Equation (1);
     else
     drop the packet from the queue; which is Equation(2);
     end if

   end for
2. Sort the computed utility values in descending order;
3. Serve the HOL packet with maximum utility value;

End

When a user with the highest SNR has a packet latency below 50% of the required maximum latency, the probability for other users to be scheduled highly rises when their instantaneous SNRs above half of that the user with the highest SNR. This is the key benefit of Greedy-Latency since the users with good channel conditions often being considered for scheduling when their packets are closely expired. Nevertheless, the user with the highest SNR will be scheduled when its packet latency is close to 1 (> 0.5). As an example, if there are two users in the cell, the first user has an instantaneous SNR of 30 dB with the packet latency ratio of 0.6 whilst the second user has an instantaneous SNR of 26 dB with the packet latency ratio of 0.9, the second user demonstrates the maximum utility function value and thus it is scheduled to be served. Despite the lower instantaneous SNR value of the latter, its packet is actually closer to expiring compared to the former that makes the second user to have a higher priority. If both users have the same packet latency ratio, the first user who has the highest instantaneous SNR will be scheduled. This then leads to the Greedy scheduler behaviour. When both users have the same instantaneous SNRs, the Greedy-Latency prefers the user with the tighter packet latency; hence Greedy-Latency offers better service than the pure Greedy for such condition.

B. Combination of packet dropping with other schedulers

We apply the same proposed packet dropping (PD) policy above with other present schedulers to perform a fair comparison, which is explained as follows. Firstly, the parameters d_k and α are computed using Equation (3) and α

$=d_k/T_k$ accordingly for all HOL packets in the scheduler. Secondly, if the computed α is less than 1 (α <1 for non expired packets), the existing scheduling algorithm (e.g. PF) is computed. Otherwise, the packet is considered to have expired and it is dropped from the queue. We represent the combination of the dropping policy with PF, Greedy and WFQ as PF+PD, Greedy+PD and WFQ+PD respectively. The PF+PD and Greedy+PD are classified as the channel-& delay-aware scheduler, whilst WFQ+PD is known as a queue-&delay-aware scheduler.

III. SIMULATION MODEL

A system-level simulation using Qualnet [16] is conducted to investigate the performance of the joint packet dropping and scheduling policy including Greedy-Latency in satisfying the required maximum latency of 50 ms. Table II summarises the configured system parameters used specifically for downlink transmission. Since these studies focus on the rtPs class, we consider Variable-Bit-Rate (VBR) packets with size of 300 bytes [17] to represent video streaming traffic. The system load of ~11.5 Mbps generated is assumed to be close to the congested network, ~ 92% of the total system capacity[18], aiming to examine the QoS issues under real conditions. A BS employs WFQ + PD, PF+PD, Greedy+PD and Greedy-Latency independently as an outbound scheduler to serve the rtPS users. The BS is also assumed to establish video connections with 10 and 25 uniformly distributed users in a single WiMAX cell. Three types of users' profiles are used for different number of users above: stationary, low mobility (4 to 18 km/h) and high mobility (32 to 119 km/h). Users move with a random speed between the specified ranges and have the same maximum latency requirement (i.e. 50 ms). The BS starts streaming the video traffic 15s after the simulation begins and it lasts for 100s. Every simulation runs for four independent seeds.

IV. SIMULATION RESULTS AND ANALYSIS

A. Total Achievable Goodput

The channel-&delay-aware schedulers (Greedy-Latency, PF+PD, Greedy+PD), as expected, achieve higher goodput than WFQ+PD, as Figure 1 shows. These channel-&delay-aware schedulers are able to achieve approximately between ~84% and ~92% of the system capacity across the stationary and mobility users despite the fact that the packets are pre-emptively dropped from the queues. In addition, the channel-&delay-aware schedulers exploit the users' time varying channel to achieve some goodput gains. Greedy-Latency, still exploits multi-user diversity, but only partially since the *packet latency ratio* (i.e. d_k/T_k) is introduced. The packets destined for the SS with best channel conditions are sometimes scheduled later in order to serve an SS with a good quality channel but with tighter packet latency. The difference of packet drop performance (see Figures 2 and 3) in the 10 and 25 users significantly affects the achievable goodput trend as shown in Figure 1. The WFQ+PD generally achieves the worst goodput performance due to the highest packet drop percentage in both 10 and 25 users scenarios, and also it blindly schedules users upon the smallest finish number[19] at a lower SNR and neglects the users with a higher SNR. In the 25 high mobility case, all channel-&delay-aware schedulers lose some goodput due to the out-of-date channel feedback (3s) received by the BS. When the packets reach the high mobility users, the users' channel conditions have varied rapidly and the channel is different from the feedback received due to the small coherence time (i.e. 2ms to 7 ms), depending on the users' speed. This then results to inaccuracies in the link adaptation and scheduling which causes some packets to be received in burst errors. Having a more frequent feedback will lead to goodput degradation due to considerable overheads in the system.

TABLE II. SIMULATION PARAMETERS FOR BOTH 10 AND 25 CASES

Configuration	Parameters	Value
PHY	Operating Frequency	2.5 GHz
	PHY Mode	OFDMA
	Duplexing Mode	TDD
	Channel Bandwidth	10 MHz
	FFT Size No of Used Subcarriers Sampling Factor Guard Interval	1024 720 8/7 1/8
	Propagation model Propagation Limit Shadowing Model Shadowing Mean Fading	Two Ray Ground -111.0 dBm Log-normal 8.9 Rayleigh
MAC	Frame Size	5 ms
	DL: UL symbol ratio	24:24
	ARQ	Disabled
Others	Transport and IP	UDP and IPv4
	Simulation Duration	100 s
Traffic Model	Traffic Type	Variable-Bit-Rate (VBR)
	Packet Size	300 bytes
	Mean Packet Interval (25 users)	3.64 ms
	Mean Packet Interval (10 users)	1.20 ms

B. Packet Drop Percentage

All schedulers apart from WFQ+PD show a low packet drop percentage, which is lower than 5% for all scenarios, as can be seen in Figures 2 and 3. This implies that the adopted drop policy ensures that the video is of a good quality as a packet loss is less than 10%[2] for the channel-&delay-aware schedulers. However, this is not possible for the queue-&delay-aware scheduler, WFQ+PD. This is because the WFQ algorithm has a dependency on the number of queues in the scheduler, random arrival packets and current queue length to compute the finish number for its scheduling, which ultimately worsens the efficiency of the radio resource allocation. On the contrary, all the channel-&delay-aware schedulers rely on the user's channel condition and not on the other users' states. Both figures indicate that Greedy-Latency offers a proportionate service amongst the users and hence the lowest packet drop percentage is achieved. The figures then also show that the packet drop percentage increases significantly with the number of users in the network. The schedulers have more choices to serve 25 users compared to 10 users and thus in the former case. a considerable number of users are waiting for their turn to be served This consequently results to their packets being delayed longer than the required latency and hence they have to be dropped. Both figures, in general, show that the high mobility scenarios usually achieve the lowest packet drop across the schedulers. The high mobility users have higher probability of experiencing good channel conditions at different time and frequencies. That makes the schedulers to serve the users more fairly and not give access only to certain users, like in the stationary case.

C. *Average Delay*

Figure 4 shows that all schedulers are able to achieve an average delay below 25ms; hence the joint packet dropping and scheduling together can satisfy the required maximum latency for a high system load environment. It can be shown that, a scheduler alone without a packet dropping policy for high load scenarios, the average delay is more than 50 ms[12, 13]. Greedy-Latency offers a large amount of packets to be served (see Figures 2 and 3 - less dropped packets) at the expense of a higher average delay, as demonstrated in Figure 4. Greedy-Latency, in general, still achieves better performance than the required latency. Figure 4 also shows that the high mobility scenarios achieve the shortest average delay because the packets spend less time in the queues to be served. For the channel- & delay-aware schedulers, the high mobility profile offers the users higher probability to be scheduled due to multiuser diversity. When there are more users in the network such as 25 users, a longer average delay can be seen in Figure 4, because 25 users have a lower probability to be served compared to 10 users.

D. *Goodput and Delay Min-Max Fairness Indices*

Table III summarises the min-max fairness indices for 25 stationary, low mobility and high mobility cases across all schedulers employed. The values in bold indicate the highest values for the corresponding metrics. Greedy-Latency is the fairest scheduler amongst the 25 low mobility users in terms of goodput distribution, and also amongst the 25 high mobility users in terms of delay. It also improves the delay and goodput fairness indices over Greedy+PD but results to a longer average delay. Nevertheless, PF+PD is typically the fairest scheduler for most cases. The combination of WFQ with the dropping policy as given in this paper is the least fair in all cases. This finding is in contrast to the case when the BS employs WFQ without the dropping policy [20]. This is because the computation of the finish number as a scheduling metric penalizes the overall performance including its fairness, as described before.

V. CONCLUSION

This paper introduces a joint channel- and delay-aware scheduling and packet dropping, Greey Latency, to satisfy the required maximum latency whilst optimising the system goodput for WiMAX networks. The performance of Greedy-Latency is compared with other schedulers (PF, Greedy and WFQ); the same simple drop policy is applied to maintain a fair comparison. Simulation results show that all schedulers can guarantee the maximum latency requirement (50ms) for both stationary and mobility cases after the packet dropping is introduced; however, this is not the case for the schedulers

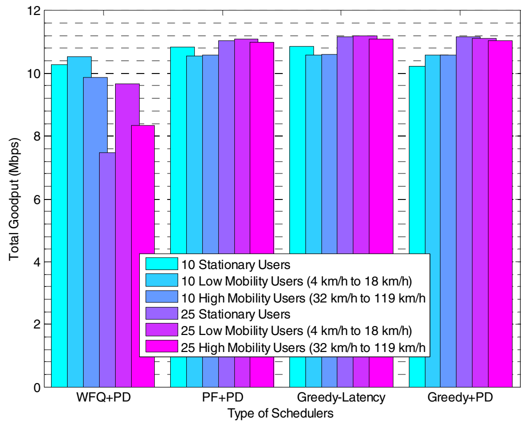

Figure 1. Total Achievable Goodput across 10 and 25 users

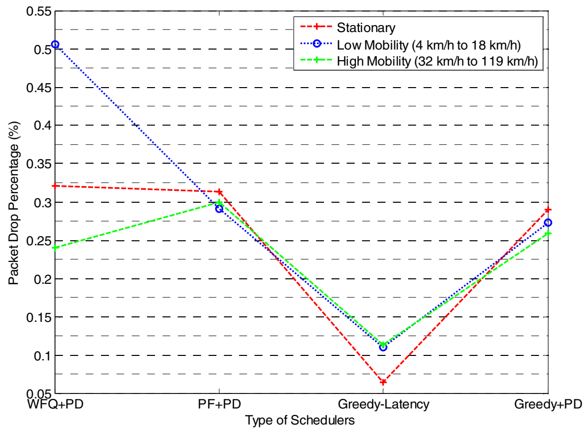

Figure 2. Packet drop percentage across 10 users

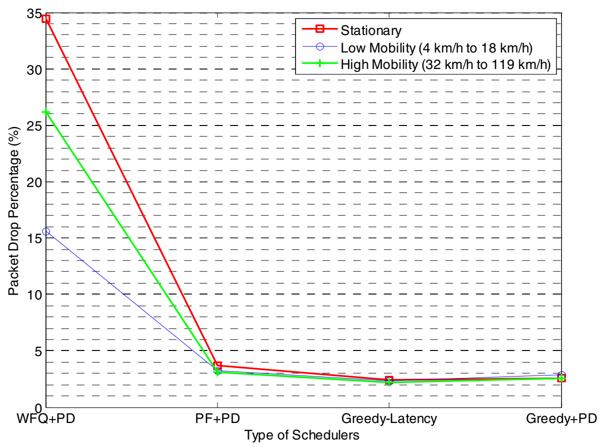

Figure 3. Packet drop percentage across 25 users

TABLE III. MIN-MAX FAIRNESS INDICES FOR GOODPUT AND DELAY METRICS ACROSS 25 STATIONARY, LOW MOBILITY AND HIGH MOBILITY USERS

Mobility Profiles	**Stationary**		**Low Mobility**		**High Mobility**	
Schedulers and Metrics	***Delay***	***Goodput***	***Delay***	***Goodput***	***Delay***	***Goodput***
WFQ-plus-packet dropping	0.2018	0.8343	0.3026	0.8712	0.2452	0.6278
PF-plus-packet dropping	**0.8311**	**0.8961**	**0.8395**	0.9442	0.7878	**0.7268**
Greedy-plus-packet dropping	0.4061	0.8055	0.6661	0.9344	0.7076	0.7126
Greedy-latency-plus-packet dropping	0.6542	0.8055	0.7812	**0.9452**	**0.8294**	0.7265

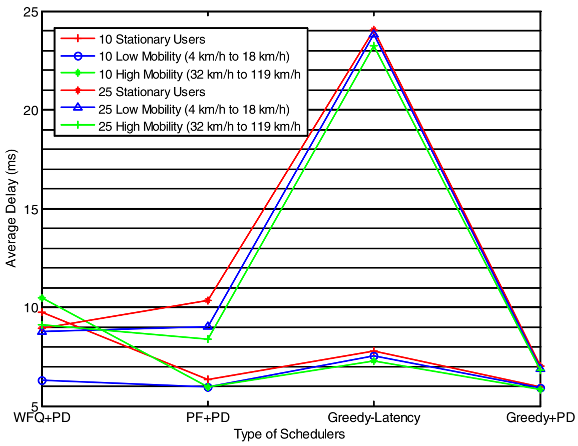

Figure 4. Average delay across 10 users

without the dropping policy. It is also clear that mobility does not affect the overall performance across these schedulers as significantly as the number of users in the system. All channel-&delay-aware schedulers still achieve a good goodput, between ~84% and ~92% of the system capacity, in the stationary and mobility cases. A goodput degradation is seen in the case of high mobility of 25 users because the channel feedback is not accurate to effectively exploit the user's channel condition. Greedy-Latency is observed to have the lowest packet drop percentage at the expense of a longer average delay that is still within the required latency. This scheduler improves overall the goodput and delay fairness indices against Greedy+PD. Finally, note that overall PF+PD remains the fairest scheduler even with the dropping is present. Introducing a simple packet dropping ensures that PF+PD can meet the maximum latency requirement.

ACKNOWLEDGMENT

The first author would like to thank the Ministry of Higher Education of Malaysia and Universiti Teknologi Malaysia to sponsor her research studies.

REFERENCES

[1] P. Pahalawatta, R. Berry, T. Pappas, and A. Katsaggelos, "Content-Aware Resource Allocation and Packet Scheduling for Video Transmission over Wireless Networks," Selected Areas in Communications, IEEE Journal on, vol. 25, pp. 749-759, 2007.

[2] G. A. Jefrey, G. Arunabha, and M. Rias, Fundamentals of WiMAX Understanding Broadband Wireless Networking: Prentice Hall, 2007.

[3] D. Wu, "QoS provisioning in wireless networks," Wireless Communications and Mobile Computing, vol. 5, pp. 957-969, 2005.

[4] T. Ali-Yahiya, A.-L. Beylot, and G. Pujolle, "An adaptive cross-layer design for multiservice scheduling in OFDMA based mobile WiMAX systems," Computer Communications, vol. 32, 2009.

[5] A. Lera, A. Molinaro, and S. Pizzi, "Channel-Aware Scheduling for QoS and Fairness Provisioning in IEEE 802.16/WiMAX Broadband Wireless Access Systems," Network, IEEE, vol. 21, pp. 34-41, 2007.

[6] Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems," in IEEE Std P802.16e/D12, ed, 2005.

[7] A. G. Kitti Wongthavarawat, "Packet scheduling for QoS support in IEEE 802.16 broadband wireless access systems," International Journal of Communication Systems, vol. 16, pp. 81-96, 2003.

[8] H. Safa, H. Artail, M. Karam, R. Soudan, and S. Khayat, "New Scheduling Architecture for IEEE 802.16 Wireless Metropolitan Area Network," in IEEE/ACS International Conference on Computer Systems and Applications, 2007. AICCSA '07., 2007, pp. 203-210.

[9] W.-H. Park, S. Cho, and S. Bahk, "Scheduler Design for Multiple Traffic Classes in OFDMA Networks," in Communications, 2006. ICC '06. IEEE International Conference on, 2006, pp. 790-795.

[10] Z. Sun, Y. Zhou, M. Peng, and W. Wang, "Dynamic Resource Allocation with Guaranteed Diverse QoS for WiMAX System," in Communications, Circuits and Systems Proceedings, 2006

[11] M. Nicolaou, A. Doufexi, S. Armour, and Y. Sun, "Scheduling Techniques for Improving Call Capacity for VoIP Traffic in MIMO-OFDMA Networks," in VTC 2009-Fall, 2009, 2009, pp. 1-5.

[12] R. Akmam, A. Doufexi, and D. Kaleshi, "Performance Evaluation of MIMO Downlink WiMAX for Different Schedulers," presented at the Wireless Advanced, London, UK, 2010.

[13] H. Wang and L. Dittmann, "Downlink resource management for QoS scheduling in IEEE 802.16 WiMAX networks," Comput. Commun., vol. 33, pp. 940-953, 2010.

[14] M. Andrew, K. Kumaran, K. Ramnan, A. Stolyar, R. Vijayakumar, and P. Whiting, "Scheduling in a Queuing System with Asynchronosly Varying Service Rates," 2004.

[15] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar, and P. Whiting, "Scheduling in a queueing system with asynchronously varying service rates," Probab. Eng. Inf. Sci, p. 18, 2004.

[16] S. N. Technologies. Qualnet [Online]. Available: http://www.scalable-networks.com/

[17] N. A. Ali, P. Dhrona, and H. Hassanein, "A performance study of uplink scheduling algorithms in point-to-multipoint WiMAX networks," Computer Communications. Adaptive Multicarrier Communications and Networks, vol. 32, pp. 511-521, 2009.

[18] R. Akmam, A. Doufexi, and D. Kaleshi, "Performance of MIMO Downlink WiMAX at Application Layer," presented at the Personal Indoor and Mobile Radio Conference, Turkey, 2010.

[19] C. I. Technology. (2008). Cisco IOS Quality of Service Solutions Configuration Guide. Available: www.cisco.com/warp/public/732/Tech/wfq.

[20] C. Valencia, "Scheduling Alternatives for Mobile WiMAX End-to-End Simulations and Analysis," Master Thesis, System and Computer Engineering, Carleton University, Ottawa, Canada, 2009.