KSII TRANSACTIONS ON INTERNET AND INFORMATION SYSTEMS VOL. 2, NO. 2, APRIL 2008

Bandwidth Management of WiMAX Systems and Performance Modeling

Yue Li, Jianhua He and Weixi Xing, *Non-members*
Institute of Advanced Telecommunications, Swansea University,
Wales, United Kingdom
[e-mail: {Y.Li, J.He, W.Xing}@swansea.ac.uk]
*Corresponding Author: Yue Li

Received February 26, 2008; revised April 10, 2008; accepted April 12, 2008;
published April 25, 2008

Abstract

WiMAX has been introduced as a competitive alternative for metropolitan broadband wireless access technologies. It is connection oriented and it can provide very high data rates, large service coverage, and flexible quality of services (QoS). Due to the large number of connections and flexible QoS supported by WiMAX, the uplink access in WiMAX networks is very challenging since the medium access control (MAC) protocol must efficiently manage the bandwidth and related channel allocations. In this paper, we propose and investigate a cost-effective WiMAX bandwidth management scheme, named the WiMAX partial sharing scheme (WPSS), in order to provide good QoS while achieving better bandwidth utilization and network throughput. The proposed bandwidth management scheme is compared with a simple but inefficient scheme, named the WiMAX complete sharing scheme (WCPS). A maximum entropy (ME) based analytical model (MEAM) is proposed for the performance evaluation of the two bandwidth management schemes. The reason for using MEAM for the performance evaluation is that MEAM can efficiently model a large-scale system in which the number of stations or connections is generally very high, while the traditional simulation and analytical (e.g., Markov models) approaches cannot perform well due to the high computation complexity. We model the bandwidth management scheme as a queuing network model (QNM) that consists of interacting multiclass queues for different service classes. Closed form expressions for the state and blocking probability distributions are derived for those schemes. Simulation results verify the MEAM numerical results and show that WPSS can significantly improve the network's performance compared to WCPS.

***Keywords*:** Bandwidth management scheme, WiMAX, queuing network model (QNM), generalized exponential (GE) distribution, first-come-first-served (FCFS), processor share (PS), maximum entropy (ME) principle, performance evaluation

Digital Object Identifier 10.3837/tiis.2008.02.001

1. Introduction

Modern mobile radio communications technologies have fundamentally changed our lives, and they are expected to provide more advanced services including high speed real-time multimedia services and the transmission of large volumes of data over wireless networks. However, due to the bandwidth constraints and the need for ubiquitous operation and mobility, current wireless systems can not efficiently support the expected services. Moreover, network infrastructures in rural areas and developing countries are relatively poor and providers are unwilling to install the necessary network equipment in regions that have a low profit potential. For those areas, broadband wireless technologies are very attractive [1]. Worldwide Interoperability for Microwave Access (WiMAX) [2] is one such broadband wireless technology, and it was invented particularly for metropolitan area networks. It is connection oriented, it can provide a very high date rate, it has large service coverage, and it offers flexible quality of service (QoS).

In the WiMAX system, service classes can be classified into Unsolicited Grant Services (UGS), Real-Time Polling Services (rtPS), Extended rtPS, No real-time Polling Services (nrtPS), and Best Effort (BE) [3]. How to manage the traffic from those service connections becomes one of the important issues in the WiMAX system. The bandwidth management scheme is therefore a critical component of WiMAX networks for managing the different kinds of traffic efficiently. However, due to the large number of connections and flexible QoS supported by WiMAX, the uplink access in WiMAX networks is very challenging since the medium access control (MAC) protocol must efficiently manage the bandwidth and related channel allocations. In this paper, we propose a novel bandwidth management scheme, namely, the WiMAX partial sharing scheme (WPSS), to obtain more efficient bandwidth management and better QoS. Under this scheme, any available bandwidth allocated for higher priority traffic can be efficiently used for lower priority traffic. On the other hand, the higher priority traffic can preempt bandwidth that is being used by lower priority traffic. Consequently, the scheme can improve the bandwidth utilization without violating the QoS of higher priority traffic and without delaying transmission.

Simulation modeling has been widely accepted as an efficient tool for studying detailed system behaviors. However, the simulation approach becomes costly, particularly as the system size increases considerably. Markov models can provide greater flexibility, but the associated numerical solutions may suffer from several drawbacks, such as the restrictive assumptions of the Poisson arrival process and/or state space explosion, which limit the applications of the Markov model approach for large scale wireless systems. The WiMAX system is especially novel, more complex, and more advanced. Because of the high frequency, high data rate, and large number of connections in WiMAX systems, the simulation tools and Markov model approaches are facing big challenges for evaluating the performances of WiMAX systems. Recently, we have applied the principle of maximum entropy (ME) [4][5][6] to evaluate the performances of a wireless GSM/GPRS cell with generalized exponential (GE) bursty traffic using a partial bandwidth sharing scheme. In this paper, we extended our previous work on the ME-based analytic model [4][5] to the performance modeling and evaluation of a WiMAX system in which multiclass bursty traffic flows are managed by WiMAX complete partitioning (WCPS) or the WiMAX partial sharing scheme (WPSS). In the analytical model, an open queuing network model (QNM) is proposed, which consists of five interacting multiclass generalized exponential (GE)-type queuing and delay systems with multiple servers and finite capacities. ME analytic solutions are derived, subject to appropriate GE-type queuing and delay theoretic mean value constraints and expressions. The model states and blocking probability distributions are determined by closed-form expressions. Typical numerical experiments are included to verify the credibility of the ME solutions compared to the simulation approach and to compare the WiMAX network performances under the WCPS and WPSS joint bandwidth management schemes.

The rest of the paper is structured as follows. Literature reviews and related WiMAX traffic classes are presented in Section 2. Two bandwidth management schemes are introduced in Section 3. Some preliminary remarks and the ME analysis of the bandwidth management systems are highlighted in Section 4. Typical experimental results are presented in Section 5. Concluding remarks follow in Section 6.

2. General Bandwidth Schemes and WiMAX Traffic Classes

Cost-effective algorithms for queuing and delay network models under various bandwidth management schemes are widely recognized as powerful and realistic tools for the performance evaluation and prediction of WiMAX performance with ever increasing volumes of multimedia traffic with different quality-of-service (QoS) guaranteed.

2.1 Literature Review

Earlier performance models for 2G wireless networks are based on the Global System for Mobile (GSM) telecommunications for the transmission of voice calls (e.g., [5]) and its extension based on the General Packet Radio Service (GPRS), which allows data communication with higher bit rates than those provided by a single GSM channel (e.g., [4][6][7][8]). Moreover, the performance models of the wireless third generation (3G) multi-service Universal Mobile Telecommunication Systems (UMTS) for multimedia traffic flows, such as voice calls, streaming media and data packets, have been of major interests for mobile network providers and have received widespread attention in the literature (e.g., [9][10][11]). Recent performance evaluation studies in the field are very often based on simulation modeling and the numerical solution of Markov models [5][6][7][8][9][10][11]. Moreover, Yi-Ting Mai presents his framework for improving the IP QoS and 802.16 QoS in [12]. It was confirmed that four types of traffic could be transmitted independently with QoS guarantees. For high-speed optical transmission, the reference model of 40 Gb/s-based photonic packet switches has been proposed for wide bandwidth management [13]. Those schemes can use the bandwidth management scheme to transmit high data rate traffic, but they are inefficient for improving the network performance. In this paper, WPSS is proposed not only to extend the earlier GSM/GPRS/UMTS bandwidth management schemes, but also to improve the WiMAX system efficiency and achieve better system performance.

2.2 WiMAX Service Classes

According to the 802.16e standard, there are five service classes defined, which are introduced below [3].

Unsolicited Grant Services (UGS): The UGS is designed to support real-time uplink service flows that generate and transport fixed-size data packets on a periodic basis, such as T1/E1 and voice over IP without silence suppression. The service offers fixed size grants on a real-time periodic basis, which eliminates the overhead and latency of SS (subscriber station) requests and assure that grants are available to meet the flow's real-time needs.

Real-Time Polling Services (rtPS): The rtPS is designed to support real-time uplink service flows that generate and transport variable size data packets on a periodic basis, such as moving pictures experts group (MPEG) video. The service offers real-time, periodic, unicast request opportunities that meet the flow's real-time needs and allow SSs to specify the size of their desired grant. This service requires more request overhead than UGS, but it supports variable grant sizes for optimum data transport efficiency.

Extended rtPS: Extended rtPS is a scheduling mechanism that builds on the efficiency of both UGS and rtPS. The extended rtPS provides unicast grants in an unsolicited manner as in UGS, thus saving the latency of a bandwidth request. However, whereas UGS allocations are fixed in size, ertPS allocations are dynamic.

No Real-time Polling Services (nrtPS): The nrtPS offers unicast polls on a regular basis, which assures that the uplink service flow receives request opportunities even during network congestion. The base station (BS) typically polls nrtPS CIDs on an interval on the order of one second or less.

Best Effort (BE): The intent of the BE type of service grant scheduling is to provide efficient service for best effort traffic in the uplink. In order for this service to work correctly, the Request/Transmission Policy setting must be set such that the SS is allowed to use contention request opportunities. This results in the SS using contention request opportunities as well as unicast request opportunities and unsolicited Data Grant Burst Types data transmission opportunities. All other bits of the Request/Transmission Policy are irrelevant to the fundamental operation of this scheduling service and should be set according to network policy.

Each SS to base station connection is assigned a service class as part of the creation of the connection. When packets are classified in the convergence sub-layer, the connection into which they are placed is chosen based on the type of QoS guarantees that are required by the application. In our system, the general traffic can include five types of classes, which are Voice IP (VoIP) without silence suppression, VoIP with silence suppression, Video Conferencing (VC), File Transmission (FTP), and HTTP. In this context, different WiMAX traffic has various priority and QoS requirements. With these conditions, the novel WiMAX scheduler will be proposed in following section.

3. Bandwidth Management Schemes for a WiMAX Network

With the development of wireless communication technology, the requirement of broadband wireless network has changed from 'last 100 meters' to 'last miles', and the WiMAX technology has been introduced as one of the world's candidate solutions. More recently, moderate mobility can be supported by WiMAX. In this section, we will discuss two bandwidth management schemes for WiMAX networks.

3.1 The WCPS & WPSS Bandwidth Management Schemes

Efficient joint bandwidth management schemes for WiMAX are critical for the support of mobile multimedia applications with required quality-of-service (QoS). Resources for those WiMAX traffic flows can be reserved statically or dynamically, and a combination of both is possible. Different cell capacity partitioning schemes for handling the transmission of UGS, rtPS, extend-rtPS, nrtps, and BE data packets can be defined. In such an environment, partitions of the available bandwidth may be created for related WiMAX traffic. There can be two main types of traffic handling schemes:

- WiMAX complete partitioning scheme (WCPS) divides the total cell capacity to simultaneously serve multimedia traffic. As a consequence, the transmission of UGS, rtPS, extend-rtPS, nrtps, and BE data packets can be studied separately.
- WiMAX partial sharing scheme (WPSS) is based on a partial sharing of radio channels where the conversational class of the UGS partition does not tolerate any delay, so it has the highest priority and it is independent of other traffic flows, followed by that of the rtPS partition and nrtPS partition, which are considered a high priority class and a low priority class respectively. Low priority traffic will reserve the channels from both the high priority and low priority partitions. The high priority class will take the higher priority role in the system. In this case, new incoming high priority data packets could preempt the reserved channels from the low priority traffic. Moreover, at any point in time, if the high priority class has some of its own channels idle, the low priority classes may use them, as appropriate, in order to enhance their own transmission capacity. New high priority class arrivals, however, will immediately preempt channels from their own channels and from any low priority classes that may use them, as appropriate (c.f., **Fig. 2**).

Two open QNMs of the WiMAX architecture under WCPS and WPSS related bandwidth management schemes can be seen in **Fig. 1** and **Fig. 2**, respectively. The compound Poisson process with geometrically distributed batch sizes is used to represent the arrival process of multiple class bursty traffic consisting of UGS, rtPS, extend-rtPS, nrtps, and BE data packets with equivalent GE-type interarrival times. Moreover, GE distributions are used to describe the UGS, rtPS, extend-rtPS, nrtps, and BE data packets channel transmission times. In this paper, we will focus only on uplink traffic.

The bandwidth management scheme WCPS (c.f., **Fig. 1**) operates fixed (non-interacting) partitions for UGS, rtPS, extend-rtPS, nrtps, and BE data packets. Thus, five interacting queuing and delay systems can be clearly analyzed independently. The joint bandwidth management protocol of WPSS is implemented under the following traffic handling operational conditions: At any given time, UGS will be independently transmitted through the fixed bandwidth. In the mean time, rtPS data packets may be acquired to increase the capacity of the nrtPS data packets, as appropriate. On the other hand, the extend-rtPS queuing system provides an opportunity to increase the capacity of the BE data packets delay system (c.f., **Fig. 2**). However, new arrivals with higher preemptive priority (e.g., rtPS data packets and extend-rtPS) will cause the immediate release of some or all of these channels to the rtPS/extend-rtPS queuing system. Subsequently, the transmission capacity of both nrtPS and BE data packets partitions will be progressively reduced or increased, as appropriate. Note that in the absence of any available capacity under WCPS and WPSS, the traffic flows of all the service classes (UGS, rtPS, extend-rtPS, nrtps, and BE data packets) will be lost on arrival.

Although WCPS cannot clearly achieve the best utilization for radio resources, it has the advantage of requiring a simpler management policy and implementation. Note that WCPS is the limiting case of WPSS under high loads.

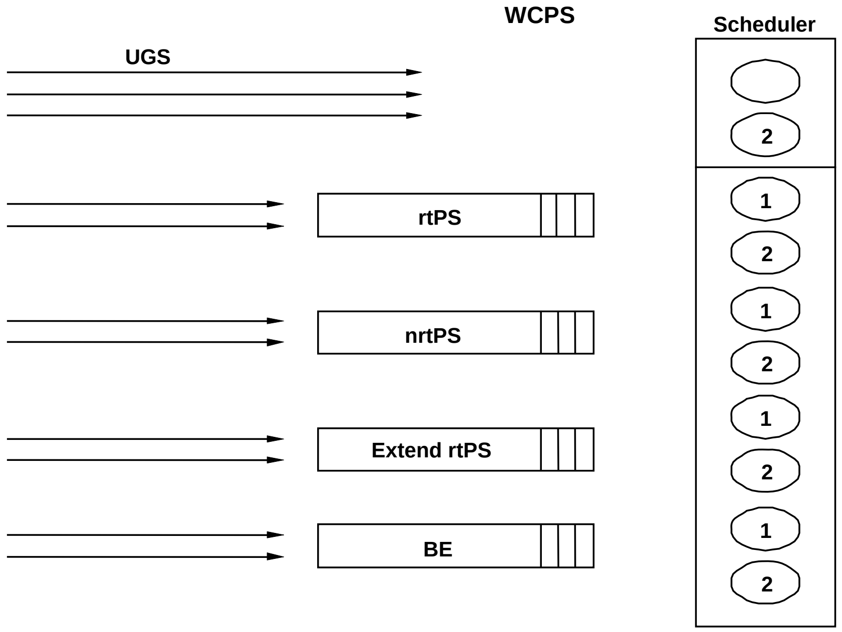

Fig. 1. An open QNM of a WiMAX system under WCPS

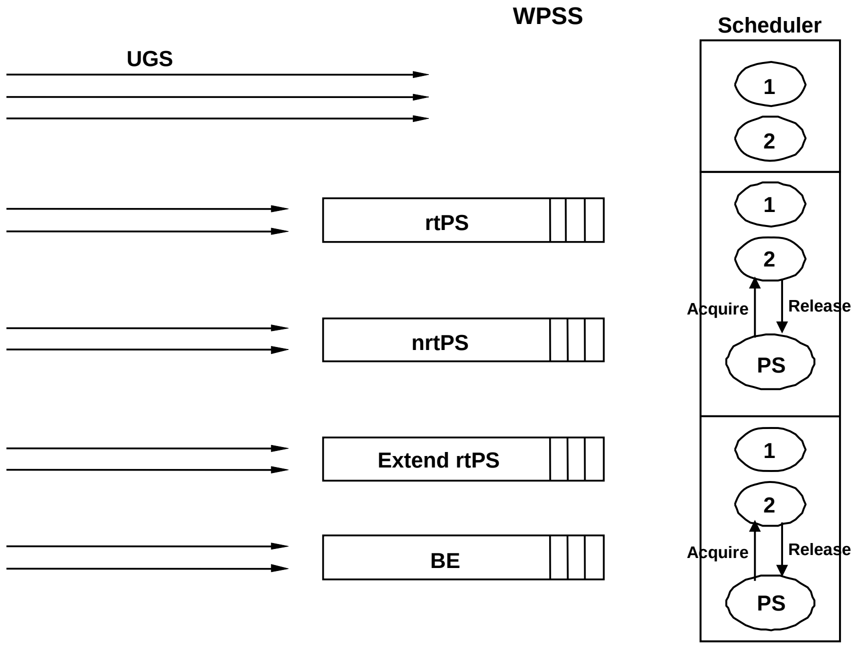

Fig. 2. An open QNM of a WiMAX system under WPSS

Moreover, in the area of queuing theory, the dynamic changing wireless channel can be generally modeled by a different server with various service rates. Therefore, we simply assume a clean wireless channel, over which packets will be correctly received unless collisions happen. General voice calls are abstracted as an M/M/c/c loss system with a Poisson distribution with c servers with particularly service rates.

From the our recent research [14][15][16][17], the open QNMs can be clearly decomposed into three interacting GE-type building block queuing and delay systems as follows: The UGS partition can be generally modeled by a classical pseudo-birth death GE/GE/c_1/c_1 loss system with GE-type classes of conversational (IP voice) calls and service rates with c_1 channels. Moreover, the rtPS partition can be represented by a finite capacity GE/GE/c_{21}/N_{21}/FCFS queuing system under FCFS. nrtPS is also described as a GE/GE/c_{22}/N_{22} delay system with a process sharing (PS) discipline. The extend-rtPS data packets scheme is developed as a GE/GE/c_{31}/N_{31}/FCFS queuing system. The BE data packets partition can be modeled by a finite capacity GE/GE/1/N_{32}/PS delay system with a multiple class of interactive and background data packets under a discriminatory PS rule. Note that under the PS rule, the available service capacity is shared evenly among all data packets belonging to the same class. However, in the presence of multiple data classes with different priority levels, the service capacity is shared according to PS with discrimination rates favoring higher priority classes. This is feasible because the data packets partition is physically capable of allocating all available bandwidth to one connection (subject to some battery restrictions). Also, when there are multiple connections, one time slot (channel) can be shared by eight different connections (c.f., [18]). Note that under the PS rule, N represents the maximum number of connections that can simultaneously share the available service capacity.

4. Analytical Model for the WiMAX Bandwidth Management Schemes

This section overviews the GE-type distribution and principle of maximum entropy (ME). Analytic solutions for WCPS, WPSS, and related ME analyses are included as well.

In this paper, the GE-type distribution is used to describe the WiMAX traffic flows. There are several reasons for selecting the GE-type distribution, which are explained below. The inter-arrival rate and squared coefficient of variation (SCV) of the inter-event times are mainly used to simulate the different traffic with a particular queuing model for WiMAX. For example, the UGS class can be modeled by the GE/GE/c/c loss system without any delay. The BE traffic could be used by the PS rule and the GE-type queuing model to model the multi-connections in one time slot or in several time slots. In this paper, we focus only on the GE-type distribution as an example. It should be noted that the GE-type distribution is one of the short range distributions. It can only approximate the different WiMAX traffic that is combined with a particular queue to characterize the traffic's character and assist in QoS provisioning. Moreover, facing the arrival of thousands of packets, the other methodologies may be incapable of obtaining an analytic solution due to the large scale and complex system design. Using GE-type distributions and related queuing models can help achieve performance measurements (utilization, mean queue length, and blocking probability etc) with a complex queuing network model. On the other hand, Maximum Entropy (ME) is a key solution to the problems of modeling an unlimited queuing network with a GE-type distribution, and the ME analytic solution obtained in our paper may apply to other traffic distributions (e.g., self-similar and long range distributions) with proper modifications. However, we place emphasis not only on the bandwidth management schemes, but also on the WiMAX analytic solutions against the simulation results. Our numerical results can give hints about the performances of the WiMAX bandwidth management schemes when they are used with other distributions.

4.1 GE-Type Distribution

The GE-type distribution is of the form [19][20]

$$F(t) = P(X \le t) = 1 - \tau e^{-\tau v t}, \tau = 2/(1 + C^2), t \ge 0 \tag{1}$$

where X is an inter-event time random variable and $\left\{1/v, C^2\right\}$ are the mean and squared coefficient of variation (SCV) of the inter-event times, respectively.

The GE distribution has a counting compound Poisson process (CPP) with geometrically distributed batch sizes with mean $1/\tau$. It may be meaningfully used to model the inter-arrival times of bursty multiple class mobile connections with different minimum capacity demands. Note that an IP packet length distribution is known to be non-exponential and should at least be described by the mean, $1/v$, and SCV, C^2. This is because IP packets are restricted by the underlying physical network, such as Ethernet and ATM, and thus, they have different packet lengths, typically 1500 bytes and 53 bytes, respectively.

The GE distribution may also be employed to model short range dependence (SRD) traffic with a small error. For example, an SRD process may be approximated by an ordinary GE distribution whose first two moments of the count distribution match the corresponding first two SRD moments. This approximation of a correlated arrival process by an uncorrelated GE traffic process may facilitate (under certain conditions) problem tractability with a tolerable accuracy and, thus, the understanding of the performance behavior of external SRD traffic in the interior of the network. It can be further argued that, for a given buffer size, the shape of the autocorrelation curve, from a certain point onwards, does not influence system behavior (c.f., [19]). Thus, in the context of system performance evaluation, an SRD model may be used to accurately approximate long range dependence (LRD) real traffic.

The nrtPS and BE data packets traffic has been shown to have a self-similar property or an LRD property, and those properties should be used in the simulations of the wireless systems in order to obtain a more realistic performance evaluation. However, it should be pointed out that if the self-similarity of the traffic must be modeled by analytic approaches, such as a Markov model or an ME, then the modeling task will be extremely difficult, if not impossible. Therefore only several specific traffic distributions are successfully studied as the input to the queuing systems. In this paper, we use the GE-type distribution to approximate the distributions of real data traffic, which has been successfully used as the traffic model in our previous research work. The GE-type distribution has a short range inter-arrival time character and a bulk size parameter, which can be used to model data packets traffic and real-time traffic. We have found that such an approximation does not considerably affect the accuracy of our analytic model on the performance evaluation of the queuing systems. Therefore, we continue to use the GE-type distribution to describe the traffic for the scheduling services in WiMAX systems and this enables us to propose an accurate and computationally tractable analytic model. To our knowledge, no such analytical model exists for the bandwidth management scheme for WiMAX systems. We believe that it is possible to extend our work with the GE-type distribution to other types of distributions. We will consider this research problem in our future research.

4.2 Maximum Entropy (ME) Formulation

The principle of ME (c.f., Jaynes [19]) provides a self-consistent method of inference for characterizing, under general conditions, an unknown but true probability distribution, subject to known (or, known to exist) mean value constraints. The ME solution can be expressed in terms of a normalizing constant and a product of Lagrangian coefficients corresponding to the constraints. In an information theoretic context, the ME solution is associated with the maximum disorder of the system states, so it is considered to be the least biased distribution estimate of all solutions satisfying the system's constraints. Major discrepancies between the ME distribution and an experimentally observed (c.f., [20]) or stochastically derived (c.f., [14]) distribution indicate that important physical or theoretical constraints have been overlooked. Conversely, experimental or theoretical agreement with the ME solution constitutes evidence that the constraints of the system have been properly identified.

In the field of systems modeling, expected values performance distributions, such as those relating to the marginal and joint state probabilities (i.e., number of jobs at an individual queue or a network) may be known to exist and thus, they can be used as constraints for the characterization of the form of the ME solution. Alternatively, these expectations may often be analytically established via classical queuing theory in terms of some moments of the inter-arrival time and service time distributions. An efficient analytic (exact or approximate) implementation of the ME solution clearly requires the a priori estimation of the Lagrangian coefficients via exact or asymptotic expressions involving basic system parameters and performance metrics, as appropriate. Hence, the principle of ME may be applied to characterize useful information theoretic approximations of performance distributions of queuing/delay systems and related networks.

The ME solution for the joint state probability distribution of an arbitrary open network with queuing and delay stations, subject to marginal mean value constraints, can be interpreted as a product-form approximation. Thus, entropy maximization implies a decomposition of a complex network into individual stations, each of which can be analyzed separately with revised inter-arrival and service times. Moreover, the marginal ME state probability of a single station, in conjunction with suitable formulae for the estimation of flow moments in the network, can play the role of a cost-effective analytic building block towards the computation of the performance metrics of the entire network. Further information on ME formalism and queuing system can be found in Kouvatsos [15].

This section highlights the ME methodology, as applied to the analysis of the GE/GE/1/N/PS and GE/GE/c/N/FCFS/CBS queuing systems. Note that the GE/GE/c/c loss system is a special case of the GE/GE/c/N/FCFS/CBS queuing system with $N = c$.

Moreover, further details on the mathematical proofs associated with key analytic GE-type results can be found in [14][15].

For each class $i\,(i = 1, 2, \ldots, R, R > 1)$, let $\{1/\lambda_i, C_{ai}^2\}$ and $\{1/\mu_i,\ C_{si}^2\}$ be the mean and SCV of the interarrival and service time distributions, respectively. Note that μ_i ($i = 1, 2, \ldots, R$) under the PS rule is defined subject to discriminatory weights. Moreover, for either a GE/GE/c/N/FCFS/CBS queuing system or a GE/GE/1/N/PS delay system, let at any given time

- $\mathbf{S} = (c,\ c_1,\ c_2, \ldots, c_R)$, $\sum_{R}^{i=1} c_i \le N$, where c_i, ($i = 1, 2, \ldots, R$) is the class of the ith job in either system with c being the class of the job in service under the FCFS rule;

- $\mathbf{Q}$ be the set of all feasible states of **S**.
- **P(S)** , $\mathbf{S} \in \mathbf{Q}$ be the stationary state probability.
- n_i be the number of entities of class $i\,(i = 1, 2, ..., R)$;
- π_i be the blocking probability that an arrival of class i will find the system at full capacity.

For each state **S**, $\mathbf{S} \in \mathbf{Q}$, and class i, $i = 1, 2, ..., R$, the following auxiliary functions are defined:

$n_i(\mathbf{S}) =$ the number of class i jobs present in either the GE/GE/1/N/PS or the GE/GE/c/N/FCFS/CBS system.

$$s_i(\mathbf{S}) = \begin{cases} 1, \text{if the job in service is of class i;} \\ 0, \text{otherwise,} \qquad \text{GE/GE/1/N/PS} \end{cases}$$

$$s_{ik}(\mathbf{S}) = \begin{cases} 1, \text{if } n_i \geq k \text{ and } k \leq c - \sum_{j=1}^{i-1} n_j; \\ \\ 0, \text{otherwise,} \quad GE/GE/c/N/FCFS/CBS. \end{cases}$$

$$f_i(\mathbf{S}) = \begin{cases} 1, \text{if } \sum_{i=1}^{R} n_i(\mathbf{S}) = N, \text{and } s_i(\mathbf{S}) = 1; \\ \\ 0, \text{otherwise,} \qquad GE/GE/1/N/PS \end{cases}$$

$$f_{ik}(\mathbf{S}) = \begin{cases} 1, \text{if } \sum_{i=1}^{R} n_i(\mathbf{S}) = N, \text{and } s_{ik}(\mathbf{S}) = 1; \\ 0, \text{otherwise,} \quad GE/GE/c/N/FCFS/CBS. \end{cases}$$

Note that the constraints of the server state are $s_i(S)$ or $s_{ik}(S)$, which is related to the utilization of the bandwidth. On the other hand, the constraints of the queue capacity are designed as $f_i(S)$ or $f_{ik}(S)$.

Suppose that the following mean value constraints about the state probability P(**S**) are known to exist:

(i) Normalization

$$\sum_{\mathbf{S} \in Q} P(\mathbf{S}) = 1. \tag{2}$$

(ii) Probabilities $\{U_i, i = 1, 2, ..., R\}$, $\{U_{ik}, i = 1, 2, ..., R; k = 1, 2, ..., c\}$

$$\begin{cases} \sum_{\mathbf{S}\in Q} s_i(\mathbf{S})P(\mathbf{S}) = U_i, \quad 0 < U_i < 1, \\ \qquad i = 1, 2, ..., R, \qquad \text{GE/GE/1/N/PS;} \\ \sum_{\mathbf{S}\in Q} s_{ik}(\mathbf{S})P(\mathbf{S}) = U_{ik}, \quad 0 < U_{ik} < 1, \\ \qquad i = 1, 2, ..., R; k = 1, 2, ..., c, \qquad \text{GE/GE/c/N/FCFS/CBS.} \end{cases} \tag{3}$$

(iii) Average number in the system { $L_i, i = 1, 2, ..., R$ },

$$\sum_{\mathbf{S}\in\mathbf{Q}} n_i(\mathbf{S})P(\mathbf{S}) = L_i, \quad i = 1, 2, ..., R. \tag{4}$$

with $U_i < L_i < N$ for GE/GE/1/N/PS and $U_{i1} < L_i < N$ for GE/GE/c/N/ $FCFS/CBS$.

(iv) Full buffer state probabilities { $\phi_i, \phi_{ik}, i = 1, 2, ..., R; k = 1, 2, ..., c$ }

$$\begin{cases} \sum_{\mathbf{S}\in\mathbf{Q}} f_i(\mathbf{S})P(\mathbf{S}) = \phi_i, \quad 0 < \phi_i < 1, \qquad \text{GE/GE/1/N/PS;} \\ \sum_{\mathbf{S}\in\mathbf{Q}} f_{ik}(\mathbf{S})P(\mathbf{S}) = \phi_{ik}, \quad 0 < \phi_{ik} < 1, \quad \text{GE/GE/c/N/FCFS/CBS,} \end{cases} \tag{5}$$

satisfying the class flow balance equations, namely

$$\lambda_i(1 - \pi_i) = \mu_i U_i, i = 1, ..., R. \tag{6}$$

The form of the ME joint state probability distribution { $P(\mathbf{S}), \mathbf{S} \in \mathbf{Q}$ } can be characterized by maximizing the entropy functional $H(\mathbf{P}) = -\sum_{\mathbf{S}\in\mathbf{Q}} P(\mathbf{S}) \log P(\mathbf{S})$, subject to prior information expressed by mean value constraints (2)-(5). By employing Lagrange's method of undetermined multipliers, the following solutions are obtained:

$$P(\mathbf{S}) = \begin{cases} \frac{1}{Z} \prod_{i=1}^{R} g_i^{s_i(\mathbf{S})} x_i^{n_i(\mathbf{S})} y_i^{f_i(\mathbf{S})}, \forall \mathbf{S} \in \mathbf{Q}, \qquad \text{GE/GE/1/N/PS;} \\ \frac{1}{Z} \prod_{i=1}^{R} (\prod_{k=1}^{c} g_{ik}^{s_{ik}(\mathbf{S})}) x_i^{n_i(\mathbf{S})} y_{ik}^{f_{ik}(\mathbf{S})}, \forall \mathbf{S} \in \mathbf{Q}, \quad \text{GE/GE/c/N/FCFS/CBS,} \end{cases} \tag{7}$$

where $Z = 1/P(0)$ is the normalizing constant, $\{g_i, x_i, y_i, i = 1, 2, ..., R\}$ and $\{g_{ik}, x_i, y_{ik}, i = 1, 2, ..., R; k = 1, 2, ..., c\}$ are the Lagrangian coefficients corresponding to constraints (3)-(5) per class, respectively.

4.3 The Aggregate ME Probability Distribution

By using equations (7), the aggregate state probabilities $\{P(n), n = 0,1,\ldots,N\}$ are given by

$$P(n) = \begin{cases} \dfrac{1}{Z}\left(\sum_{i=1}^{R} g_i x_i y_i^{f_{ik}(\mathbf{n})}\right) X^m, n \in [1,N], GE/GE/1/N/PS \\ \dfrac{1}{Z}\prod_{i=1}^{R}\left\{\prod_{k=1}^{c} g_{ik}^{s_{ik}(\mathbf{n})} x_i y_{ik}^{f_{ik}(\mathbf{n})}\right\} X^m, n \in [1,N], GE/GE/c/N/FCFS \end{cases} \tag{8}$$

where $Z = 1/P(0)$, $X = \sum_{R}^{i=1} x_i$ and $\sum_{i=1}^{R} n_i = n$, $m = \begin{cases} n-1, & \text{if } n \le N-1; \\ N, & \text{if } m = N. \end{cases}$

4.4 The Lagrangian Coefficients

The Lagrangian coefficients $\{g_i, x_i\, i = 1,2,\ldots,R\}$ and $\{g_{ik}, x_i, i = 1,2,\ldots,R; k = 1,2,\ldots,c\}$, can be approximately determined via closed form asymptotic expressions relating to the ME solution of the corresponding GE-type infinite capacity queues at equilibrium [15]. Moreover, the Lagrangian coefficients $\{y_i, y_{ik}, \forall i,k\}$ can be determined via the flow balance equation (6) (c.f., [16]). To this end, appropriate formulae can be established, namely for ($i = 1,2,\ldots,R$) and ($k = 1,2,\ldots,c$)

$$\begin{cases} g_i = \dfrac{\sigma_i \gamma_i \rho_i}{\sigma_i \rho_i + \gamma_i - \sigma_i \gamma_i}, & \text{GE/GE/1/N/PS;} \\ g_{ik} = \dfrac{\sigma_i c \rho_i + (k-1)\gamma_i (1-\sigma_i)}{k * (\gamma_i (1-\sigma_i) + \sigma_i)}, & \text{GE/GE/c/N/FCFS/CBS.} \end{cases} \tag{9}$$

$$x_i = \begin{cases} \dfrac{\gamma_i + \sigma_i \rho_i - \gamma_i \sigma_i}{\gamma_i + \sigma_i \rho_i - \gamma_i \sigma_i \rho_i}, & \text{GE/GE/1/N/PS;} \\ \dfrac{\sigma_i \rho_i + \gamma_i (1-\sigma_i)}{\sigma_i \rho_i (1-\gamma_i) + \gamma_i}, & \text{GE/GE/c/N/FCFS/CBS.} \end{cases} \tag{10}$$

$$y_i \text{ or } y_{ik} = \frac{1-\rho_i}{1-x_i} \frac{\sigma_i}{\gamma_i (1-\sigma_i) + \sigma_i} \tag{11}$$

where $\sigma_i = \frac{2}{Ca_i^2 + 1}$, $\gamma_i = \frac{2}{Cs_i^2 + 1}$ and $\rho_i = \begin{cases} \lambda_i / \mu_i, & \text{GE/GE/1/N/PS;} \\ \lambda_i / c\mu_i, & \text{GE/GE/c/N/FCFS/CBS.} \end{cases}$

4.5 The Blocking Probability

A universal expression for the marginal blocking probabilities { $\pi_i, i = 1,2,\ldots,R$ } of a stable multiple class GE/GE/1/N/PS delay system and GE/GE/c/N/FCFS/CBS queuing systems can be approximated by focusing on a tagged job within an arriving bulk and making use of GE-type probabilistic arguments.

To this end, let $\sigma_i = \frac{2}{Ca_i^2+1}$ and $\gamma_i = \frac{2}{Cs_i^2+1}$ be the GE-type interarrival and service time non-zero stage selection probabilities associated with the GE/GE/1/N/PS and GE/GE/c/N/FCFS/CBS systems. Given that the system is in state $\mathbf{n} = (n_1, n_2, ..., n_R)$ with $n = \sum_{i=1}^{R} n_i$, the number of available buffer spaces is equal to $N - n$. By focusing on a tagged job within an arriving bulk of class i ($i = 1, 2, ..., R$), the following blocking probability can be clearly determined:

(i) P(a class i tagged job is blocked and its bulk finds the queue in state 0 = (0, 0, …, 0))

$$= \begin{cases} \delta_i(0)(1-\sigma_i)^N P(0), & GE/GE/1/N/PS; \\ \delta_i^c(0)(1-\sigma_i)^N P(0), & GE/GE/c/N/FCFS/CBS. \end{cases} \tag{12}$$

where $\delta_i(0) = \dfrac{\gamma_i}{\gamma_i(1-\sigma_i)+\sigma_i}$.

(ii) P(a class i tagged job is blocked and its bulk finds a queue in state $\mathbf{n} = (n_1, n_2, ..., n_R), \sum_{i=1}^{R} n_i \le N$

$$= \begin{cases} (1-\sigma_i)^N P(\mathbf{n}), & GE/GE/1/N/PS; \\ \begin{cases} \delta_i^{c-n}(0)(1-\sigma_i)^{N-n} P(\mathbf{n}), 0 < n < c \\ (1-\sigma_i)^{N-n} P(n), c \le n \le N \end{cases} & GE/GE/c/N/FCFS/CBS. \end{cases} \tag{13}$$

where $\delta_i(0) = \dfrac{\gamma_i}{\gamma_i(1-\sigma_i)+\sigma_i}$.

Combining Eqs. (12) - (13), after some manipulation, the blocking probabilities $\{\pi_i, i = 1, 2, ..., R\}$ can be expressed by

$$\pi_i = \begin{cases} \sum_{n=0}^{N} \delta_i(n)(1-\sigma_i)^{N-n} P(n), & GE/GE/1/N/PS; \\ \sum_{n=0}^{N} \delta_i^L(0)(1-\sigma_i)^{N-n} P(n), & GE/GE/c/N/FCFS/CBS, \end{cases} \tag{14}$$

where $\delta_i(0) = \frac{r_i}{r_i(1-\sigma_i)+\sigma_i}$, $\delta_i(n) = 1(\forall n > 0)$, and

$$L = \begin{cases} c & \text{if } n = 0, \\ max(0, c-n) & \text{if } 0 < n < c \\ 0 & \text{if } c \le n \le N. \end{cases}$$

4.6. Weighted Performance Measures under WPSS

All performance metrics of interest for the QNM of **Fig. 2** under WPSS may be determined by making use of weighted averages taking probabilistically into account the dynamic increase/decrease of the low priority transmission capacity. More specifically, low priority channels will receive capacity under the WPSS variable transmission scheme depending on the availability of free high priority channels. Thus, the average performance measure could be defined:

- $P(n)$ probability of n jobs in the system
- S Statistic performance measures
- μ_{lp} service rate of the low priority class
- μ_{hp} service rate of the high priority class
- l number of channels engaged from the high priority class
- c number of high priority channels

Low priority packets will get some available channels from the high priority class, but it will give them back to the high priority class when the new incoming high priority packets arrive, so the statistic (S) of the low priority will be:

$$U_{lp} = \sum_{i=1}^{N} P(i);$$

$$L_{lp} = \sum_{i=0}^{N} P(i);$$

$$S_{lp} = \sum_{i=0}^{c} S_{lp}(\mu_{lp} \leftarrow \mu_{lp} + \mu_{hp}(c-l)) * P_{hp}(n=l)$$

The proposed analytic GE/GE/1/N/PS delay model differs from and, in some respects, extends overall the MMPP/M/1 delay model suggested by Kouvatsos et al. [20]. Although the latter model incorporates a Markov modulated Poisson arrival process (MMPP), it is only applicable to a single class of data packets, it assumes exponential transmission times and, being an infinite capacity delay model, it does not capture the adverse blocking effect on system performance. Moreover, the GE-type delay model can be solved via closed-form expressions as opposed to those requiring computationally demanding matrix geometric methods.

5. Numerical Results

This section presents typical numerical results, with the purposes to (i) illustrate the credibility of the ME methodology against the simulation approach, (ii) compare the bandwidth management capabilities of WCPS and WPSS, and (iii) to assess the impact of GE-type bursty traffic on the network performances. Our numerical experiments are run with the following reference parameters, configured as UGS (5Mbps), rtPS (14.4Mbps), extend-rtPS (12Mbps), nrtPS (12.5Mbps), and BE data packets (6.5Mbps) (c.f., [21][22]). The experiment input is { $\mu = 5.0$, $c = 8$, $Ca^2 = 5$, $Cs^2 = 5$} for UGS, {$\mu = 14.4$, $Ca_2 = 2$, $Cs^2 = 5$, $c = 2$, $N = 5$} for the rtPS class, {$\mu = 12.5$, $Ca_2 = 5$, $Cs^2 = 5$, $N = 10$} for the nrtPS class, $\{\mu = 12, c = 2, Ca^2 = 5, Cs^2 = 5\}$ for the extend-rtPS class, and $\{\lambda = 5.0, \mu = 6.25, Cs^2 = 5, N = 15\}$ for the BE class. Note that the Java programming language is used to design the simulation for the QNMs of **Fig. 1** and **Fig. 2**, and to obtain the simulation results at the 95 % confidence intervals.

Typical numerical experiments show that, over a wide range of parameter configurations, the ME results are very comparable in accuracy to those obtained via simulation. Moreover, it can be seen that the interarrival time SCV has an adverse effect on the performance metrics of different service classes.

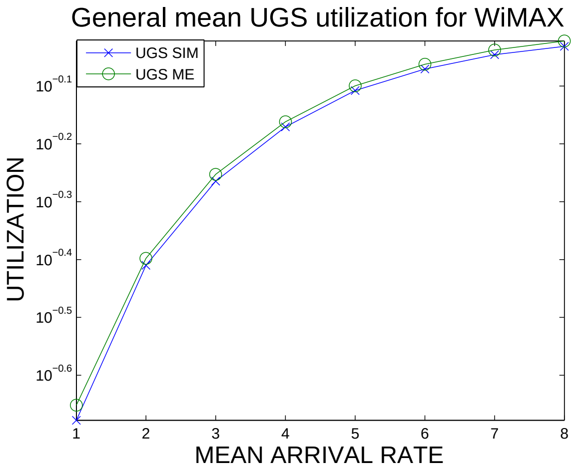

Fig. 3. Effect of traffic variability on the utilization of UGS for a multiple class GE/GE/c/c loss system

Fig. 3 illustrates the channel utilization of the UGS class via the ME analytical solution, which shows a perfect match between the simulation and the analytic results. The mean queue length is shown in **Fig. 4**, which describes the delay performance of the rtPS traffic flows. As a lower priority class, nrtPS will share the channel capacity with the rtPS class.

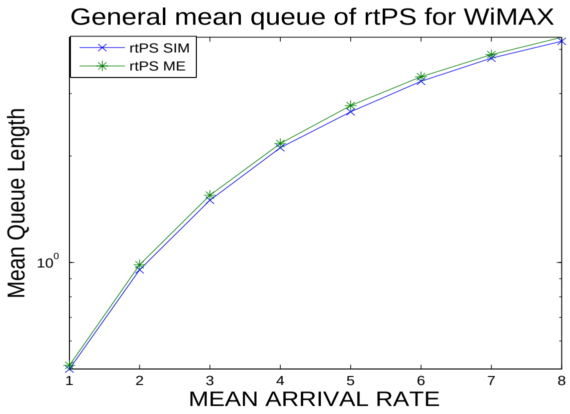

Fig. 4. Effect of traffic variability on the aggregate mean number of rtPS for a GE/GE/c/N/FCFS/CBS queuing system

The effect of the two bandwidth management schemes, WCPS and WPSS, can be seen in **Fig. 5**. The utilization of bandwidth is much higher with WPSS than with WCPS. The general utilization of the extend-rtPS class is shown in **Fig. 6**. **Fig. 7** shows that the analytic mean number of BE data packets decreases rapidly with increasing external interarrival-time SCVs, which demonstrates that WPSS is an efficient bandwidth management scheme. **Fig. 5** and **Fig. 7** clearly demonstrate that WPSS can achieve much better performance, higher utilization, and a lower mean queue length. Note that more numerical experiments with a wide range of parameter configurations can be seen in [14].

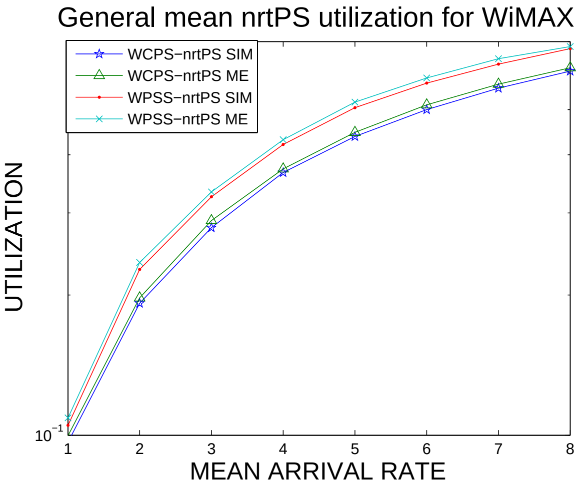

Fig. 5. Effect of traffic variability on the utilization of nrtPS for a GE/GE/1/N/PS delay system under WCPS and WPSS

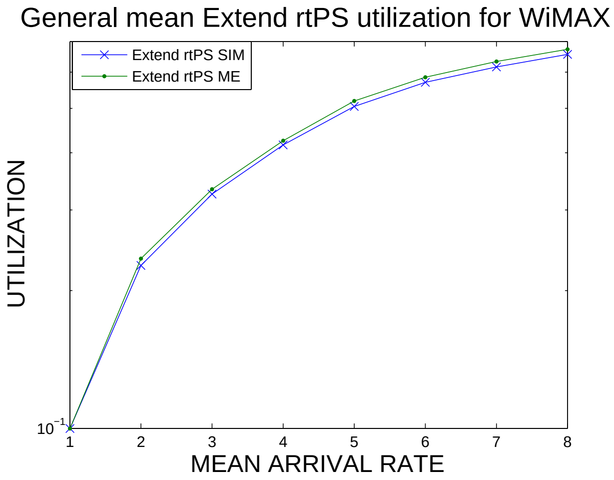

Fig. 6. Effect of traffic variability on the general utilization of extend rtPS for a GE/GE/c/N/FCFS/CBS queuing system

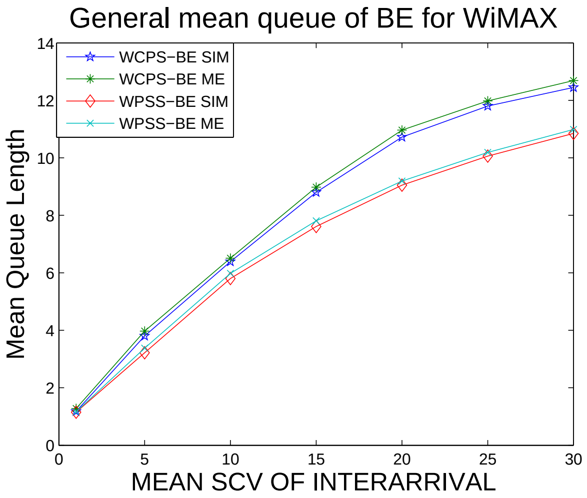

Fig. 7. Effect of traffic variability on the mean number of BE data packets for a multiple class GE/GE/1/N/PS delay system under WCPS and WPSS

6. Conclusion

In this paper, we investigated two bandwidth management schemes for WiMAX systems to understand how the bandwidth management schemes will behave and how to effectively configure and optimize the schemes, which are critical for efficient bandwidth utilization and QoS provisioning. Because WiMAX systems have high data rates and are connection oriented, there can be a large number of data packets and connections for the systems to handle, so understanding the performances of the bandwidth management schemes is inefficient and difficult when using the traditional simulation approach. Due to the large number of connections and corresponding system states, the Markov model approach is also not efficient for modeling the WiMAX systems. In this context, we proposed an efficient Maximum Entropy based analytical model to evaluate the performances of the bandwidth management schemes. The analytical model can be used to avoid the huge state spaces of the Markov models as well as the long simulation times for large-scale WiMAX systems. In the analytical model, a triple interacting GE-type queuing and delay system with multiple servers and finite capacities is studied, subject to appropriate GE-type queuing and delay theoretic mean value constraints. Closed-form expressions for the analytical model states and blocking probability distributions are obtained. Typical numerical results verify the credibility of the ME solution compared to the simulation results at 95 % confidence intervals and confirmed the performance superiority of WPSS (compared to WCPS) for WiMAX systems. Finally, the numerical results verify the adverse effect of bursty traffic on the network performances of WiMAX systems.

Acknowledgement

This work has been partially funded by the European Union through the Welsh Assembly Government. The authors sincerely thank the anonymous reviewers for their valuable and constructive comments, and the great help from Professor Minho Jo.

References

[1] S. J. Vaughan-Nichols, "Achieving Wireless Broadband with WiMax," *Computer*, Vol. 37, Issue 6, pp. 10-13, June 2004.

[2] D. Pareek, "WiMax: Taking Wireless to the MAX," Auerbach Publishers Inc. June, 2006.

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

[4] D. D. Kouvatsos, Y Li and I. Awan, "Performance Modeling of a Wireless GSM/GPRS Cell under Partial Sharing Scheme," *NETWORKING 2004*, pp. 1252-1262, May 2004.

[5] K.Al-Begain, G. Bolch and M. Telek, "Scalable Schemes for Call Admission and Handover Handling in Cellular Networks with Multiple Services," *Journal on Wireless Personal Communications*, Volume 15, No. 2, Kluwer Academic Publishers, pp. 125-144, 2000.

[6] R. Litjens and R. Boucherie, "Radio Resource Sharing in GSM/GPRS Network," *ITC Specialist Seminar on Mobile Systems and Mobility*, Lillehammer, Norway, Mar. 22-24, pp. 261-274, 2000.

[7] Y-R. Haung, Y-B. Lin and J-M. Ho, "Performance Analysis for Voice/Data Integration on a Finite-Buffer Mobile System," *IEEE Transaction on Vehicular Technology*, Vol. 49, No. 2, Mar. 2000.

[8] C.H. Foh, B. Meini, B. Wydrowski and M.Zuerman, "Modeling and Performance Evaluation of GPRS," *Proc. of IEEE VTC*, 2001, Rhodes, Greece, pp. 2108-2112, May 2001.

[9] H. Kaaranen, A. Ahtiainen, L. Laitinen and S.N. Laitinen, "UMTS Networks, Architecture, Mobility and services," John Wiley and Sons, ISBN: 0471-48654, 2001.

[10] Y. R. Haung, "Performance Analysis for UMTS Networks with Queued Radio Access Bearers," *IEEE Transactions on Vehicular Technology*, Vol. 51, No. 6, pp 1330-1337, 2002.

[11] Y. Ohfa, K. Kawahara, T. Ikeuaga and Y. Ore, "Performance Evaluation of Channel Switching Scheme for Packet Data Transmission in Radio Network Controller," *Networking 2000*, LNCS 2345, pp. 648-659, 2002.

[12] Yi-Ting Mai, Chun-Chuan Yang and Yu-Hsun Lin, "Cross-layer QoS framework in the IEEE 502.16 network," *ICACT 2007*, pp. 2090 - 2095 Feb. 12-14, 2007.

[13] H. Harai and M. Murata, "High-speed buffer management for 40 gb/s-based photonic packet switches," *IEEE/ACM Transactions on Networking*, Volume 14, Issue 1, pp. 191-204, Feb. 2006.

[14] Y. Li, D.D. Kouvatsos and I. Awan, "Some Recent Analytic Results for multiclass GE-type Queuing and Delay system," Research Report 08-04, Netwroks and Performance Engineering Reaseach Group, University of Bradford, August, 2004.

[15] D. D. Kouvatsos and I.U. Awan, "MEM for arbitrary closed queuing networks with RS-blocking and multiple job classes," *Annals of Operations Research*, Special Issue on Queuing Networks with Blocking, pp. 231-269, 1998.

[16] W. Xing, Y. Li and D. Kouvatsos, "An Efficient WiMAX System Design and Structure Protocol for the IRIS Project," *The forth international conference on performance modeling and evaluation of Heterogeneous Networks* (*HET-NET'S 2006*), Bradford, UK, 2006.

[17] W. Xing, Y. Li and X. H. Hguyen, "WiMAX Fast-moving Access Network," *The 2008 IEEE Wireless Communications and Networking Conference* (*IEEE WCNC*), Las Vegas, USA, 2008.

[18] V.K.N. Lau and S.V. Maric, "Mobility of Queued Call Requests of a New Call-Queuing Technique for Cellular Systems," *IEEE Transactions on Vehicular Technology*, Vol. 47, No. 2 pp. 480-488, 1998.

[19] E.T.Jaynes, "Prior Probabilities," *IEEE Trans. Syst. Sci. Cybern.*, Vol. 4, pp 227-241, 1968.

[20] D.D. Kouvatsos, "Entropy Maximisation and Queuing Network Models," *Annals of Operation Research*, Vol. 48, pp. 63-126, 1994.

[21] C. Cicconetti, L. Lenzini, E. Mingozzi and C. Eklund, "Quality of service support in IEEE 802.16 networks," *IEEE/ACM Transactions on Networking*, Volume 20, Issue 2, pp. 50-55, Mar.-Apr. 2006.

[22] C. Cicconetti, A. Erta, L. Lenzini and E. Mingozzi, "Performance Evaluation of the IEEE 802.16 MAC for QoS Support," *IEEE Transactions on Mobile Computing*, Volume 6, Issue 1, pp. 26-38, Jan. 2007.

Yue Li obtained his PhD from Bradford University in 2005. Dr. Li's research focuses on devising analytic solutions and cost-effective algorithms for the performance modeling and dimensioning of the 4G wireless cell architecture. Dr. Li has had fifteen papers published in the field of Performance Modeling and he is proficient in Java, C++, and C++ Builder, which are essential in creating new applications for networks. He is a member of the Networks and Performance Engineering Research Group and the IEEE Computer and Communication Group. He has interests in developing novel wireless 4G models.

Jianhua He is lecturer in the Institute of Advanced Telecommunications, Swansea University, UK. He received both his BEng and MEng in electronics engineering from the Huazhong University of Science and Technology (HUST), China, and his PhD in communications engineering from the Nanyang Technology University, Singapore. He joined HUST as an Associate Professor in 2001. From 2004 to 2006, he was with the Department of Electronic & Electrical Engineering, University of Bristol, UK. He was with the Department of Computing & Electronic Systems, University of Essex, UK from 2006 to 2007. His research interests include medium access control and networking for heterogeneous wireless networks (WiFi, WiMAX, WPAN, WCDMA, sensor networks, mesh networks, and mobile ad hoc networks), dynamic radio resource management and spectrum access, cross-layer performance evaluation and optimization for third generation (3G) mobile communication systems and beyond. He has published 60 journal and conference papers. He serves on the editorial board of the Wiley International Journal of Wireless Communication and Mobile Computing (WCMC), KSII Transactions on Internet and Information Systems, and serves as a guest editor for the International Journal of Autonomous and Adaptive Communications Systems. He is a chair or TPC member of many international conferences. He is a Member of IEEE.

Weixi Xing completed his PhD at Cambridge University, UK. In 1988, following the completion of his initial BA and MA degrees at the Beijing University of Technology in China, Weixi worked for eleven years in China's ICT industry and in 1997, became Senior Manager of research and development at the Shenzhen Asian Olympic Technology Co. Ltd, where he organized the development of social emergency receiving and disposal systems. This involved combining fixed-line communication, wireless communication, Internet and Intranet, GIS (Geographical Information System), and a large screen display system to form a multifunction social emergency service center. The system was subsequently awarded the National Class New Product Certificate in China. Dr. Xing is overseeing the IRIS project at Swansea, which aims to transform the current telecommunications broadband market, seeking to find low cost techniques of providing broadband delivery to businesses and homes with a high quality of service.