QoS-Guaranteed Bandwidth Shifting and Redistribution in Mobile Cloud Environment

Sudip Misra, *Senior Member, IEEE*, Snigdha Das, Manas Khatua, *Student Member, IEEE*, and Mohammad S. Obaidat, *Fellow, IEEE*

Abstract—Mobile cloud computing (MCC) improves the computational capabilities of resource-constrained mobile devices. On the other hand, the mobile users demand a certain level of quality-of-service (QoS) provisioning while they use services from the cloud, even if the interfacing gateway changes due to the mobility of the users. In this paper, we identify, formulate, and address the problem of QoS-guaranteed bandwidth shifting and redistribution among the interfacing gateways for maximizing their utility. Due to node mobility, bandwidth shifting is required for providing QoS-guarantee to the mobile nodes. However, shifting alone is not always sufficient for maintaining QoS-guarantee because of varying spectral efficiency across the associated channels, coupled with the corresponding protocol overhead involved with the computation of utility. We formulate bandwidth redistribution as a utility maximization problem, and solve it using a modified descending bid auction. In the proposed scheme, named as *AQUM*, each gateway aggregates the demands of all the connecting mobile nodes and makes a bid for the required amount of bandwidth. We investigate the existence of Nash equilibrium (NE) in the proposed solution. Theoretically, we deduce the maximum and minimum selling prices of bandwidth, and prove the convergence of AQUM. Simulation results establish the correctness of the proposed algorithm.

Index Terms—Mobile cloud computing, auction theory, Nash equilibrium, bandwidth shifting, bandwidth redistribution

1 Introduction

Mobile devices are increasingly becoming an essential part of our everyday life due to the rapid reduction in cost of hardware, improved portability, and increasing computational capability. The built-in data exchange feature in the modern mobile devices allows the users to run powerful web applications such as mobile banking, online gaming, video and image processing, online shopping, health management, and finance management. All these applications deal with real-time data streaming. Although the mobile devices have improved hardware and software, their limited energy source is a persistent problem. Consequently, executing computationally intensive applications on mobile devices remains a standing issue in mobile networks.

MCC [11], [14] is an integration of cloud computing [2] into the mobile environment. In MCC, the data processing and storing applications are moved from the mobile devices to the servers in a cloud, and, thus, improves the efficiency and lifetime of the mobile devices. In case of high computational resource and application requirements such as video streaming, audio, and data services, the mobile users request services from the cloud servers through an interfacing gateway. The gateway communicates with the cloud service provider (CSP) for allocating shared resources which are required for resolving the mobile user's request. Thereafter, connection is set up between the mobile user and the cloud server through the interfacing gateway, and, then, the mobile user is capacitated to use the resources of the cloud servers. At this juncture, we mention that the words "node", "device", and "user" are used interchangeably in the rest of the manuscript.

- S. Misra, S. Das, and M. Khatua are with the School of Information Technology, Indian Institute of Technology, Kharagpur, West Bengal 721302, India. E-mail: {smisra, snigdhadas, manas.khatua}@sit.iitkgp.ernet.in.
- M.S. Obaidat is with the Department of Computer Science and Software Engineering, Monmouth University, NJ 07764. E-mail: obaidat@monmouth.edu.

Manuscript received 15 Sept. 2013; revised 19 Nov. 2013; accepted 22 Nov. 2013. Date of publication 10 Dec. 2013; date of current version 30 July 2014.
Recommended for acceptance by I. Bojanova, R.C.H. Hua, O. Rana, and M. Parashar.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TCC.2013.19

1.1 Motivation

The cloud servers provide resources on demand. To fulfill on-demand request of real-time applications and computations, bandwidth is an essential component that controls the rate of transmission. It may happen that the cloud server provides real-time services as per the request, but the mobile nodes are not able to receive the service due to the lack of bandwidth. Moreover, the mobile users demand a certain level of QoS provisioning in terms of delay, jitter, response time, and reliability, while they use service from the cloud. When a mobile device changes its location, the corresponding gateway for maintaining the connectivity with the cloud also changes. Therefore, the aggregated bandwidth requirement for the gateway also changes, which creates the necessity of *bandwidth shifting*, provided the previous allocation was optimal. Further, bandwidth shifting alone is insufficient for maintaining QoS while maximizing the revenue of each gateway. It happens due to the variation in the spectral efficiency across channels, and the communication protocols across devices. We consider that the gateway earns revenue from the users for providing requested bandwidth and QoS-guarantee, while paying for getting that amount of bandwidth from the cloud servers. We further assume that the gateways are responsible for ensuring such QoS requirements. Depending upon different attributes such as channel spectral efficiency, and protocol overheads,

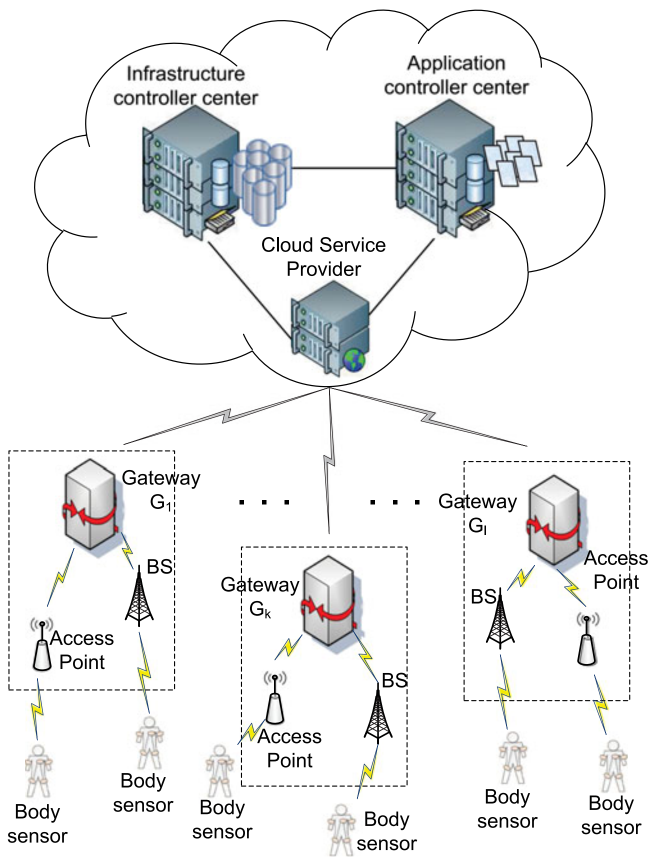

Fig. 1. Mobile cloud computing architecture.

each gateway utilizes different percentages of allocated bandwidth from the CSP. Thus, bandwidth utilization of the previous interfacing gateway may differ from that of the current one. This difference violates the QoS requirement in the newly connected gateway, while the previous gateway can provide more QoS than its present requirement. Therefore, proper *redistribution of bandwidth* is essential in real-time applications for fulfilling the QoS requirements for computational services.

1.2 Contributions

In this paper, we address the problems of bandwidth shifting and redistribution resulting from varying demand from gateways. It is pertinent to clarify at this juncture that the bandwidth redistribution problem differs from the traditional bandwidth allocation problem in that, a while the former concerns allocating proportional bandwidth to all the gateways (and, in turn, to the users), even if only a few gateways change their bandwidth demand, the latter concerns allocating bandwidth to that gateways who have changed the bandwidth demand. A schematic view of mobile cloud architecture is shown in Fig. 1. We do not consider the bandwidth allocation process between gateways and mobile nodes. We assume that the CSP is authorized for bandwidth shifting and allocation, and these functions are performed for the gateways only. In this work, we consider QoS-guarantee in terms of *service delay*. Other aspects of QoS may be considered for extending this work in the future.

Definition 1 (Service delay). *Service delay is the total time required for providing a service from CSP to a mobile node.*

We formulate this as utility maximization problem for gateways. For solving the above optimization problem, we use the concepts of auction theory *[31] from applied economics. Each gateway submits a bid to the CSP, based on the requested bandwidth and QoS demand from the mobile nodes connected to it. The CSP allocates the required bandwidth through a payment system. We formulate the utility of a gateway with the help of* revenue function *and* cost function. *The gateway defines the revenue function based on the revenue per unit allocation of effective bandwidth, and the revenue per unit service delay for the QoS requirement. The revenue of a gateway completely depends on the allocated bandwidth and QoS-guarantee between the mobile nodes and the gateway. In other words, it is postulated that higher revenue leads to increased QoS protection. It may be clarified that our focus in this work is not on the revenue maximization aspects of CSP. The cost function explains the pricing strategy between the gateway and the CSP. We prove the existence of* Nash Equilibrium *(*NE*) in the proposed scheme. Further, we theoretically deduce the extrema of selling price of bandwidth, and prove that the algorithm converges in finite number of iterations. We summarize the main* contributions *in this paper as follows:*

- *We theoretically prove the requirement of bandwidth shifting followed by bandwidth redistribution in a typical mobile cloud environment.*
- *We propose an* auction theory-based QoS-guaranteed utility maximization *(*AQUM*) algorithm that redistributes the total available bandwidth optimally.*
- *We analyze the existence of Nash equilibrium and the convergence criteria of the redistribution algorithm.*

1.3 Paper Organization

The remainder of the paper is organized as follows. In Section 2, we briefly present few relevant related works reported in the literature. In Section 3, we discuss some preliminary concepts required for understanding the proposed solution. We discuss the network model in Section 4. In Section 5, we theoretically prove the necessity of bandwidth shifting, and redistribution. We present the utility maximization problem formulation followed by an auction theory-based solution approach in Section 6. We present the experimental results followed by discussions in Section 7. Finally, we conclude the paper in Section 8, while discussing about how this work can be extended in the future.

2 Related Work

Even though MCC [11], [14] provides many advantages, Dinh et al. [11] highlighted some drawbacks in regards issues such as service availability, low bandwidth, network management, QoS-guarantees, and pricing problem in MCC. The background of these issues offer the motivational platform of our present work on QoS-guaranteed optimal bandwidth redistribution in MCC. Zhang et al. [37] designed a combinatorial auction scheme for heterogeneous resource allocation in mobile cloud system. Several works exist on optimal bandwidth allocation in various other networks. Jin and Kwok [19] recommended a limited bandwidth sharing solution for centralized mobile users using coalition game theory. This model does not provide any information regarding the amount of bandwidth sharing among the users. To overcome this limitation, Jung et al. [21] extended the scheme by incorporating the distribution policy, which evaluates the amount of bandwidth usage among users using Markov Decision Process (MDP). Krishna et al. [23] proposed a recommender system on cloud using learning automata.

There exist few allocation schemes that help to maintain QoS requirements of network nodes. For example, allocation schemes were proposed for ensuring maintaining equal expected access delay [29], fair allocation [3], guaranteed bandwidth [12], delay guarantee [20], and service differentiation [33]. For restricting the hand-off failure due to insufficient resource, a QoS-guaranteed measurement preservation scheme is proposed in [28]. For achieving end-to-end fair bandwidth allocation, Tang et al. [34] proposed a max-min fair maximum throughput bandwidth allocation (MMBA) scheme followed by lexicographical max-min fair bandwidth allocation (LMMBA) scheme for wireless mesh networks integrated with cognitive radio. Fei et al. [13] proposed a QoS-guaranteed fair up-link dynamic bandwidth allocation algorithm for allocating bandwidth from base stations to relay stations in IEEE 802.16j-based vehicular networks. A hierarchical QoS-aware dynamic bandwidth allocation algorithm [18] was proposed for assigning bandwidth depending upon the user requirements and weights of priority queue. Lin et al. [27] addressed the problem of varying QoS requirement for different data incentive applications in cloud computing system.

Concurrently, dynamic spectrum allocation in cognitive radio networks (CRNs) is also an interesting resource allocation issue. Zhu et al. [38] addressed the dynamic bandwidth allocation problem by building strategies using stochastic differential game for both the secondary users (SU) and the service providers (SP). Huang et al. [17] proposed a spectrum sharing mechanism using an auction-based approach keeping the 'interference temperature' under threshold. Chen et al. [7] addressed the spectrum sharing problem for multi-licensed primary users (PUs) using auction theory. In their proposed framework, a licensed PU shares the unused spectrum to the unlicensed secondary user based on the interference temperature threshold of the PU of CRNs. In [9], three auction-based mechanisms were proposed for distributively allocating spectrum in CRN. They compared their own algorithms based on three characteristics—convergence, social welfare, and cheat-proof. A novel auction-based approach for sharing of dynamic spectrum in CRNs [35] was proposed, in which a PU is assigned the spectrum to the SUs based on the requirement of the SUs without violating its own performance.

Chaikijwatana and Tachibana [6] investigated the problem of social surplus for efficient bandwidth allocation in wireless networks using generalized VCG auction mechanism with network coding. The authors considered that the total required bandwidth is always greater than the total available bandwidth. In [24], Kun et al. considered only the individual user performances for bandwidth allocation.

TABLE 1
Summary of Notations

Notation	Description
G_i	Gateway i
I	Total number of gateways
N_i	Set of nodes connected to a gateway G_i
U_i	Utility of gateway G_i
E_i	Spectral efficiency of a channel associated with gateway G_i
T_{ik}	Transmission delay required for accessing a service by a mobile node N_{ik} connected to gateway G_i
Z_i	Ratio of the effective bandwidth to the total available bandwidth for a gateway G_i
d_i	Service delay associated with a gateway G_i
B_i	Allocated bandwidth to a gateway G_i
B_{tot}	Total available bandwidth in CSP
α_i	Protocol overhead corresponding to a gateway G_i
b_i	Bid value given by a gateway G_i
s_i	shifting factor of gateway G_i
r_i	Revenue per unit service delay received by gateway G_i
q_i	Revenue per unit transmission rate received by gateway G_i
p	Cost price per unit bandwidth allocation charged by CSP
d_{θ_i}	Threshold value of service delay corresponding to G_i
ϕ_i	Minimum requirement of bandwidth for a gateway G_i to maintain its own operations and QoS
δ	A parameter contains positive number
β	Reserved bid for CSP
Δp	Price decrement factor
BER_i^{trgt}	Target bit error rate for a gateway G_i

Their scheme allocates same bandwidth to all users with same service request. Chai et al. [5] relaxed this limitation by considering the joint system performance of multiple access in a heterogeneous network. A multi-hop iterative auction-based bandwidth allocation mechanism for flow contention problem was proposed by Kao et al. [22]. Many researchers have also used utility theory as a strategy for bandwidth allocation. For example, the utility-based resource allocation problem is reported in [4], [8], [25] and [26]. However, the fundamental differing characteristic of the work presented in this paper from the existing ones is that we consider node mobility.

Many authors have also explored the bandwidth allocation problem in cloud computing scenario. In [1], a service level agreement-aware dynamic bandwidth allocator (DBA) is proposed in which the bandwidth is allocated among the virtual machines (VMs) for different application requirements associated with each VM. Papagianni et al. [30] proposed a unified resource allocation framework for networked clouds. Das et al. [10] considered the mapping of cloud server and mobile nodes in MCC environment. Similarly, many other works on bandwidth allocation exist in the literature. However, none of these addresses the problem of providing QoS-guaranteed bandwidth allocation in MCC environment, while the utility of each gateway is optimal. In this paper, we propose QoS-guaranteed bandwidth shifting and redistribution schemes using auction theory in MCC environment.

3 Preliminaries

We use two basic concepts—spectral efficiency of a channel, and auction theory, in problem formulation and solution approach design, respectively. In this Section, we describe them in brief for easier understanding of the solution approach. At the outset, we list in Table 1 the notations used in this paper.

3.1 Channel Spectral Efficiency

According to the Shannon capacity theory [15], the maximum rate of channel capacity (c) in an additive white Gaussian noise (AWGN) channel is expressed as

$$c = B\ log_2\Big(1 + \frac{P}{N_0 B}\Big), \tag{1}$$

where B, P and N_0 define the channel bandwidth, transmit power, and power spectral density of the noise, respectively. Therefore, the largest possible spectral efficiency of a channel for reliable communication is obtained from Shannon's spectral efficiency (E_{sh}) as follows:

$$E_{sh} = \frac{c}{B} = log_2\Big(1 + \frac{P}{N_0 B}\Big) = log_2(1 + \gamma), \tag{2}$$

where, γ is the signal to noise power ratio (SNR) at the receiver.

In MCC, each gateway may use adaptive modulation scheme for adjusting their transmission rate depending on channel quality. Therefore, the spectral efficiency (E) relative to the theoretical maximum efficiency is expressed by Goldsmith and Chua [16] as

$$E = log_2(1 + \psi\gamma), \tag{3}$$

where,

$$\psi \leq \frac{1.5}{\ln(0.2/BER^{trgt})}. \tag{4}$$

Depending on the requirements of a specific application, the target BER (BER_i^{trgt}) of gateway i may vary.

3.2 Auction Theory

In applied economics, auction theory is well known for modeling buying and selling of commodities and services. Similarly, auction theory is also useful for exchanging commodities in the network applications [36]. Apart from other auctions such as Sealed-Bid, Open-Cry, First-Price, and Second-Price [36], conventional auction [36] is more popular in the context of exchanging network commodities due to its simplicity. Conventional auction is mainly classified into two categories based on the bidding schemes—ascending or descending bid auction. For the descending bid auction, the commodity seller sets the maximum selling price (p_{max}) of an asset. Each buyer i calculates the utility $U(i)$ based on the valuation $R(i)$ and the cost $C(i)$ of the product. In the vector form, the computation of utilities is represented as follows:

$$\mathbf{U} = \mathbf{R} - \mathbf{C}. \tag{5}$$

In each iteration of the auction process, the price per unit allocation decreases by some positive value. The auction process terminates when either the buyer accepts to pay the seller's price or the price becomes zero. In the perspective of network allocation, the auction process terminates when the seller fulfills the selling criteria and confirms the selling price to the buyers. Each buyer pays the seller according to the allocation of the assets.

4 Mobile Cloud Network Model

Consider a simple mobile cloud environment in which all operations follow the discrete time model with normalized time-slots $t \in \{0, 1, 2, \ldots\}$. There are one CSP and I single-channel gateways $\mathbf{G} = \{G_1, G_2, \ldots, G_I\}$ connected with the CSP through wireless channel. Let us assume that spectral efficiency of each channel is different and represented by the vector $\mathbf{E(t)} = \{E_1(t), E_2(t), \ldots, E_I(t)\}$.

We further consider that each gateway has K number of mobile nodes connected with it at time t via any mobile network. Hence, $\mathbf{N}_i(t) = \{N_{i1}(t), N_{i2}(t), \ldots, N_{iK}(t)\}$, and, thus, $\mathbf{N} = \bigcup_{i=1}^{I} \mathbf{N}_i(t)$. In this work, we consider QoS-guarantee in respect of service delay. Let us consider that the total available bandwidth of the CSP equals B_{tot}. If a mobile node requests any service from the cloud server, the service is provided through a gateway. We consider the adequate bandwidth requirement of the gateways for successful execution of the requested services by the mobile nodes. Let $\mathbf{B}(t) = \{B_1(t), B_2(t), \ldots, B_I(t)\}$ denote the allocated bandwidth vector for the gateways $\mathbf{G}$ at time t.

4.1 Service Delay Calculation

We define T_{ik} as the transmission delay required for accessing a service by the mobile node N_{ik}, if the total available bandwidth B_{tot} is completely allocated to the gateway G_i. In other words, T_{ik} is the ideal transmission delay. Hence, the total transmission delay (T_i) for the gateway G_i equals $\sum_{k=1}^{|N_i|} T_{ik}$, where $|.|$ indicates the cardinality of a set.

Let us consider that, at time t, the CSP allocates B_i amount of bandwidth to the gateway G_i. Therefore, the transmission rate of gateway G_i is computed as $(1-\alpha_i)E_i B_i$, where α_i is the protocol overhead corresponding to the gateway G_i. If Z_i represents the ratio of the effective bandwidth of a gateway to the total available bandwidth in CSP, then, we have $Z_i = (E_i \alpha_i B_i)/(B_{tot})$. As we consider the total service delay (d_i) with respect to each gateway, the delay vector is defined as $\mathbf{d} = \{d_1, d_2, \ldots, d_I\}$. For the above network model, d_i for the gateway G_i is computed as follows:

$$d_i = \frac{\sum_{k=1}^{|N_i|} T_{ik}}{Z_i} = \frac{B_{tot} \sum_{k=1}^{|N_i|} T_{ik}}{E_i(1-\alpha_i)B_i}. \tag{6}$$

5 Bandwidth Shifting

We consider a mobile cloud network. It may be stressed that the nodes are mobile in such environment. In this section, we theoretically prove that node mobility triggers the necessity of bandwidth shifting, if the cloud server does not have any unused reserved bandwidth for future use. Subsequently, we prove that bandwidth shifting alone is not always sufficient for providing QoS-guarantee.

5.1 Necessity of Bandwidth Shifting

When a mobile node changes its location, the corresponding gateway for maintaining connectivity with the cloud changes. Therefore, the aggregated bandwidth requirement for the gateway also changes. For maintaining QoS in terms of service delay, the present gateway checks the total transmission delay for all connecting nodes, and the allocated bandwidth. In Theorem (1), we prove that the modified service delay increases with node mobility.

Theorem 1. *Let us consider that a mobile node changes its location in a MCC environment, and thus, the interfacing gateway from G_i to G_j. Then, the service delay in G_j prior to the change in node location is less than that following the change, i.e., $d_j < \hat{d}_j$, where $\hat{d}_j$ defines the modified service delay of G_j after the node changes location.*

Proof. We consider the case in which at least one mobile node, say N_{ik}, changes its location as well as the gateway from G_i to G_j. Using Equation (6), we calculate the service delay occurring at gateway G_j prior to the change in location of a node, as follows:

$$d_j = \frac{B_{tot} \sum_{k=1}^{|N_j|} T_{jk}}{E_j(1-\alpha_j)B_j}. \tag{7}$$

Similarly, we compute the service delay occurring at gateway G_j following the change in node location, as follows:

$$\hat{d}_j = \frac{B_{tot}\left(\sum_{k=1}^{|N_j|} T_{jk} + T_{ik}\right)}{E_j(1-\alpha_j)B_j}. \tag{8}$$

As the value of T_{ik} is positive, it can be inferred that,

$$d_j < \hat{d}_j. \tag{9}$$

This concludes the proof. □

Corollary 1. *Let us consider that a mobile node changes its location in a MCC environment, and thus, the interfacing gateway from G_i to G_j. Then, the service delay in G_i before the node location change is greater than that in G_i after the change, i.e., $d_i > \hat{d}_i$, where $\hat{d}_i$ defines the service delay of G_i after the node changes location.*

From Theorem (1), it may be inferred that the modified service delay for gateway G_j equals $\hat{d}_j$. The cases are imminent if the gateway wants to maintain QoS-guarantee in terms of service delay. If $\hat{d}_j$ is less than or equal to d_{θ_j}, then bandwidth shifting is not required, where d_{θ_j} is the threshold value of the service delay corresponding to the gateway G_j for providing QoS-guarantee. Otherwise, we conclude that bandwidth shifting from the previously connected gateway to the new one is essential.

5.2 Problem beyond Bandwidth Shifting

Let us consider that a mobile node, say N_{ik}, changes its location, and consequently, the interfacing gateway from G_i to G_j also changes. We further consider that the corresponding bandwidth of each gateway before and after the node's location change are B_i, $\hat{B}_i$, and B_j, and $\hat{B}_j$, respectively. Similarly, we denote the total transmission delay as T_i, $\hat{T}_i$, and T_j, and $\hat{T}_j$, respectively.

Assumption 1. *Let the mobile node N_{ik} changes its location, and, thus, the interfacing gateway from G_i to G_j, while it should satisfy the following criteria.*

(a) $B_i = \hat{B}_j$. (b) $T_i = \hat{T}_j$.
(c) $\alpha_i = \alpha_j$. (d) $E_i = E_j$.

Theorem 2. *Let the Assumptions 1(a), (b), and (c) hold for a cloud network. The service delay at the previously connected gateway before the node location change is unequal with the service delay at the currently connected gateway after the node location change, because of the variation in spectral efficiency of channels.*

Proof. We consider the case where only one node changes its interfacing gateway from G_i to G_j. From Equation (6), we compute the service delay occurring at gateway G_i before the node location change, as follows:

$$d_i = \frac{B_{tot} T_i}{E_i(1-\alpha_i)B_i}. \tag{10}$$

Similarly, we compute the service delay occurring at gateway G_j after the node location change, as follows:

$$\hat{d}_j = \frac{B_{tot} \hat{T}_j}{E_j(1-\alpha_j)\hat{B}_j}. \tag{11}$$

After taking the ratio between $\hat{d}_j$ and d_i, we get,

$$\frac{\hat{d}_j}{d_i} = \frac{E_i}{E_j}. \tag{12}$$

According to the spectral efficiency of a channel, as computed in Equation (3), the values of E_i and E_j are different, as they are correspond to two different channels. This concludes that the values of $\hat{d}_j$ and d_i are unequal. □

Proposition 1. *Let the Assumptions 1(a), (b), and (d) hold for a cloud network. The service delay in the previously connected gateway before the node location change is unequal with the service delay in the currently connected gateway after the node location change, because of differing protocol overheads.*

From Theorem 2, we notice that the QoS-guarantee provided by the gateways fails in few conditions, as the service delay changes with node mobility. This necessitates the requirement for bandwidth redistribution. Specifically, if a node changes its location, and, thus, the gateway from G_i to G_j, then bandwidth redistribution is necessitated, if $\hat{d}_j > d_{\theta_j}$. Similarly, for the different protocol overheads, we need bandwidth redistribution in some cases too.

6 Bandwidth Redistribution

In this Section, we first describe the formulation of the optimization problem governing QoS-guaranteed bandwidth redistribution. Thereafter, we present an auction-based solution scheme.

6.1 Utility Maximization Problem Formulation

We design an utility function for computing the overall benefit of each interfacing gateway. The utility function of the gateway depends on the service it provides to the mobile nodes and the bandwidth it buys for providing the services. Each gateway pays certain price for getting the required bandwidth from the CSP. On the other hand, each gateway charges certain amount of revenue from the mobile nodes for providing the services to them. Additionally, the gateway demands an extra charge from the mobile nodes for assuring the QoS in terms of service delay. Taking into account the above three factors contributing to the gateway utility, the utility function for all the gateways can be expressed as follows:

$$\mathbf{U}(b_i, \mathbf{b}_{-i}, p) = \mathbf{R}^Q(b_i, \mathbf{b}_{-i}) + \mathbf{R}^B(b_i, \mathbf{b}_{-i}) - \mathbf{C}(b_i, p), \tag{13}$$

where $\mathbf{b} = \{b_1, b_2, \ldots, b_{i-1}, b_i, b_{i+1}, \ldots, b_I\}$ denotes the requested bid vector by the interfacing gateways. In other words, the submitted bid vector $\mathbf{b}$ also represents the requested bandwidth vector in which b_i defines the requested bandwidth of the interfacing gateway G_i. $\mathbf{b}_{-i} = \{b_1, b_2, \ldots, b_{i-1}, b_{i+1}, \ldots, b_I\}$ defines the bid vector eliminating the bid of the gateway G_i, and p represents the confirmed price per unit bandwidth allocation given by the CSP. The vectors $\mathbf{R}^Q(b_i, \mathbf{b}_{-i})$ and $\mathbf{R}^B(b_i, \mathbf{b}_{-i})$ define the gateway revenue function based on the QoS in terms of service delay, and bandwidth allocation, respectively. $\mathbf{C}(b_i, p)$ defines the bandwidth allocation cost function. We model the three utility components $\mathbf{R}^Q(b_i, \mathbf{b}_{-i})$, $\mathbf{R}^B(b_i, \mathbf{b}_{-i})$, and $\mathbf{C}(b_i, p)$ as follows.

1. *Service delay specific revenue function modeling*. The gateway with small service delay increases the overall revenue. Hence, we formulate the service delay specific gateway revenue function as follows:

$$\mathbf{R}^Q(b_i, \mathbf{b}_{-i}) = \mathbf{s} - \mathbf{r}\mathbf{d}, \tag{14}$$

where, the vector $\mathbf{r} = \{r_1, r_2, \ldots, r_I\}$ defines the revenue per unit service delay occurring at the corresponding interfacing gateways, and $\mathbf{s} = \{s_1, s_2, \ldots, s_I\}$ defines the shifting factor. The shifting factor changes the negative impact of the revenue calculation associated with service delay.

2. *Bandwidth allocation specific revenue function modeling*. High bandwidth allocated to the connected gateway results in an increase in the overall revenue. Specifically, we say that the increase in effective bandwidth of a gateway results in an increment of its revenue. So, we formulate the bandwidth allocation specific revenue function as follows:

$$\mathbf{R}^B(b_i, \mathbf{b}_{-i}) = \mathbf{q}\mathbf{E}(\mathbf{1} - \boldsymbol{\alpha})\mathbf{B}, \tag{15}$$

where, the vector $\mathbf{q} = \{q_1, q_2, \ldots, q_I\}$, $\mathbf{1} = \{1, 1, \ldots, 1\}$, and $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \ldots, \alpha_I\}$ denote the revenue per unit transmission rate, and protocol overhead, respectively.

3. *Cost function modeling*. With the increase in bandwidth allocation, the overall cost of a gateway increases. So, we define the cost function of the gateway as the multiplication of the submitted bid and the price per unit bandwidth allocation, as follows:

$$\mathbf{C}(b_i, p) = pb_i. \tag{16}$$

Hence, the utility function of a gateway G_i is defined as

$$U_i = s_i - \frac{r_i B_{tot} \sum_{k=1}^{|N_i|} T_{ik}}{E_i(1-\alpha_i)B_i} + q_i E_i B_i (1-\alpha_i) - pb_i. \tag{17}$$

We consider service delay as a utility maximization constraint for performing bandwidth redistribution. To provide QoS-guarantee in terms of service delay, the value of the service delay of the gateway G_i should be less than or equal to the threshold value. Hence, the following inequality holds,

$$d_i \leq d_{\theta_i}. \tag{18}$$

Finally, we propose a utility maximization scheme using the revenue and cost functions of the gateway. The utility maximization problem is expressed as follows:

$$\begin{aligned} &maximize\ \mathbf{U} \\ &subject\ to\ d_i \leq d_{\theta_i}, \forall i \in (1, I). \end{aligned} \tag{19}$$

Lemma 1. *The utility, shown in Equation (17), is a strictly increasing function of BER in the range* $BER < 0.2$.

Proof. Taking the first derivative of the utility function, as shown in Equation (17), with respect to BER, we get,

$$\frac{\partial U_i}{\partial (BER_i)} = \frac{\partial U_i}{\partial E_i} \times \frac{\partial E_i}{\partial (BER_i)}, \tag{20}$$

where,

$$\frac{\partial U_i}{\partial E_i} = \frac{r_i B_{tot} \sum_{k=1}^{|N_i|} T_{ik}}{E_i^2(1-\alpha_i)B_i} + q_i(1-\alpha_i)B_i \tag{21}$$

$$\frac{\partial E_i}{\partial (BER_i)} = \frac{1.5\gamma}{BER_i \ \ln 2 \ \ln(\frac{0.2}{BER_i})(\ln(\frac{0.2}{BER_i}) + 1.5\gamma)}. \tag{22}$$

All the parameters in Equations (21) and (22) are positive, and the value of α_i ranges in between $[0, 1)$. Further, the value of $\ln(\frac{0.2}{BER_i})$ is always greater than 0, if BER_i is less than 0.2. Hence, we have, $\frac{\partial U_i}{\partial (BER_i)} > 0$. This concludes that the utility function is a strictly increasing function of BER in the range $BER < 0.2$. □

6.2 AQUM: Auction-Based QoS-Guaranteed Utility Maximization

An auction is a mechanism or a set of business rules for exchanging commodities based on the bidding price. Therefore, we use the auction theory-based approach for solving the QoS-guaranteed bandwidth redistribution problem in MCC, following the methodology similar to the one described in [35]. In the auction process, each gateway participates as a bidder, and the CSP acts as an auctioneer-cum-seller. We use *descending price auction* for determining the optimum bandwidth allocation—this maximizes the utility vector. In descending price auction, the seller sets the initial ceiling price for each unit allocation, and decreases the price over time until either the price becomes zero or any buyer accepts the price for buying the commodity. In this problem, we consider a tradeoff between the unit price and the bandwidth request, as, in general, the requested amount of bandwidth reduces with the increase in price per unit allocation. For incorporating the tradeoff condition, we modified the termination condition of the descending price auction. In our modified descending price auction process, the price decreases over time until the total bid reaches the total available bandwidth. In the interim, if the price p becomes less then p_{min}, the ceiling price is reset again for continuing the auction process. We describe the basic steps of the modified descending price auction for the present problem.

1. Initialization: Each gateway G_i knows its Shannon spectral efficiency E_i, protocol overhead α_i, and revenue per unit service delay r_i. We assume that r_i is determined based on the QoS-guarantee between the gateway and the connecting mobile nodes. In this work, we disregard how the value of r_i is determined—an issue which will be investigated as future work.

Initially, the CSP broadcasts its reserve bid β, which is a positive constant value required for its own use. The CSP also broadcasts the price p per unit allocation to all the gateways, and sets the initial value of p as p_{max}.

2. Bid: Each gateway G_i submits a bid b_i ($0 < b_i \leq B_{tot}$), which represents the minimum required bandwidth at the initial stage required for the QoS-guarantee constraint. The submitted bid by the gateway G_i should satisfy the service delay constraint, which considers the requests of all mobile nodes and the minimum bandwidth requirement for performing its own operation. Hence, the bid amount is computed as

$$b_i = \sum_{k=1}^{|N_i|} b_{ik} + \phi_i, \tag{23}$$

where ϕ_i represents the minimum requirement of bandwidth to maintain its own operations and QoS in the network. In every iteration, the bid is increased by some positive value δ, such that,

$$b_i(t+1) = b_i(t) + \delta. \tag{24}$$

3. Allocation: In each iteration, CSP aggregates all the bid values such that $\sum_{i=1}^{I} b_i$, and adds the aggregated value to its own reserve bid β. Finally, the CSP compares the computed value with the maximum availability of bandwidth B_{tot}. If $(\sum_{i=1}^{I} b_i(t) + \beta) \geq B_{tot}$, then the CSP concludes the auction process, and allocates B_i following Equation (25) to the gateway G_i. Otherwise, the CSP sets $p(t+1) = p(t) - \Delta p$. If the value of $p(t+1)$ decreases below the minimum price p_{min}, the CSP resets the price to the maximum value p_{max}. Finally, the CSP broadcasts $p(t+1)$ to all the gateways for the next iteration. The auction process continues until $(\sum_{i=1}^{I} b_i(t) + \beta)$ is greater than or equal to B_{tot}. Let X be the maximum number of iterations required to get optimum bandwidth allocation. As p decreases in discrete steps, sometimes we have $(\sum_{i=1}^{I} b_i(t) + \beta) > B_{tot}$. For maintaining the total available bandwidth equal to the total allocated bandwidth, we allocate B_i amount of bandwidth instead of b_i, in which B_i is computed using the bid vector $\mathbf{b}$, as follows:

$$B_i = \frac{b_i(t)}{\sum_{i=1}^{I} b_i(t)} (B_{tot} - \beta). \tag{25}$$

4. Payment: The gateway G_i pays the cost value C_i to the CSP for the allocation of bandwidth b_i units by adopting the pre-pay mechanism [35]. Then, C_i is computed as follows:

$$C_i = p b_i. \tag{26}$$

The pseudocode of auction-based QoS-guaranteed utility maximization and bandwidth redistribution scheme is presented in Algorithm 1. We deduce the pricing limit of the auction scheme in Theorem 3.

Algorithm 1 AQUM: Auction-based Bandwidth Redistribution

Inputs: p_{max}, β
Output: B

CSP broadcasts $p(t)$ to all gateways
Gateway calculates **b(t)** and **U(t)**
for $i = 1$ to I **do**
    if $(U_i(t) > U_i(t-1))$ **then**
        Gateway G_i submits bid $b_i(t)$
    else
        Gateway G_i submits bid $b_i(t-1)$
    end if
end for
if $\left(\sum_{i=1}^{I} b_i(t) + \beta \geq B_{tot}\right)$ **then**
    CSP calculates **B** and allocates to the gateways
    CSP confirms the final price $p(t)$ to the gateways
else
    CSP revises the price $p(t+1) = p(t) - \Delta p$
    if $(p(t+1) < p_{min})$ **then**
        CSP reset the price $p(t+1) = p_{max}$
    end if
    Goto Step 1 for next iteration
end if

Theorem 3. *The pricing limit for the maximum and minimum bids are represented as*

$$p_{min} = \frac{q_i E_i (1-\alpha_i)(I-1)(B_{tot}-\beta)}{I^2 B_{tot}} + \frac{r_i (I-1) \sum_{k=1}^{|N_i|} T_{ik}}{E_i(1-\alpha_i)(B_{tot}-\beta)}, \text{for } b_i = B_{tot} \ \forall i \in (1, I),$$

$$p_{max} = \frac{q_i E_i (1-\alpha_i)(I-1)(B_{tot}-\beta)}{I^2 \phi} + \frac{r_i B_{tot} (I-1) \sum_{k=1}^{|N_i|} T_{ik}}{E_i(1-\alpha_i)(B_{tot}-\beta)\phi}, \quad \text{for} \quad b_i = \phi \ \forall i \in (1, I).$$

Proof. Each gateway G_i submits a bid value based on the aggregated bandwidth requirement of mobile nodes which equals $\sum_{k=1}^{|N_i|} b_{ik}$, and its own requirement which equals ϕ_i. Without any external demand, a gateway submits a positive bid ϕ_i as its minimum requirement. Finally, the incremental bid indicates that the maximum possible value of b_i is B_{tot}. Let us consider that $\phi_i = \phi$, $\forall i$. Therefore, we conclude that, $\phi \leq b_i \leq B_{tot}$.

Taking the first derivative of the utility function shown in Equation (17), with respect to b_i, we get,

$$\frac{\partial U_i}{\partial b_i} = \left(q_i E_i (1-\alpha_i)(B_{tot}-\beta) \sum_{\forall j \neq i} b_j \right) \Big/ \left(\sum_{j=1}^{I} b_j \right)^2 + \frac{r_i B_{tot} \sum_{k=1}^{|N_i|} T_{ik} \sum_{\forall j \neq i} b_j}{E_i (1-\alpha_i)(B_{tot}-\beta) b_i^2} - p. \tag{27}$$

We know that the optimum value exists at $\frac{\partial U_i}{\partial b_i} = 0$. Therefore, we get,

$$p = \frac{q_i E_i (1-\alpha_i)(B_{tot}-\beta)\sum_{\forall j \neq i} b_j}{(\sum_{j=1}^{I} b_j)^2} + \frac{r_i B_{tot} \sum_{k=1}^{|N_i|} T_{ik} \sum_{\forall j\neq i} b_j}{E_i(1-\alpha_i)(B_{tot}-\beta) b_i^2}. \quad (28)$$

In Lemma (2), we prove that p is a decreasing function with respect to both b_i and $\sum_{\forall j \neq i} b_j$. Finally, p_{max} and p_{min} are expressed as follows:

$$p_{min} = \frac{q_i E_i (1-\alpha_i)(I-1)(B_{tot}-\beta)}{I^2 B_{tot}} + \frac{r_i (I-1)\sum_{k=1}^{|N_i|} T_{ik}}{E_i(1-\alpha_i)(B_{tot}-\beta)}, \text{for} b_i = B_{tot}\ \forall i \in (1, I),$$

$$p_{max} = \frac{q_i E_i (1-\alpha_i)(I-1)(B_{tot}-\beta)}{I^2 \phi} + \frac{r_i B_{tot}(I-1)\sum_{k=1}^{|N_i|} T_{ik}}{E_i(1-\alpha_i)(B_{tot}-\beta)\phi}, \text{for} b_i = \phi\ \forall i \in (1, I).$$

This concludes the proof. □

Lemma 2. *p is a decreasing function with respect to both b_i and $\sum_{\forall j\neq i} b_j$ if the following inequality holds:*

$$b_i > \frac{\sum_{\forall j\neq i} b_j}{\left[\frac{E_i(1-\alpha_i)(B_{tot}-\beta)\sqrt{q_i}}{\sqrt{r_i B_{tot}\sum_{k=1}^{|N_i|} T_{ik}}} - 1\right]}.$$

Proof. Taking the first derivative of the price function, as derived in the Theorem 3 (shown in Equation (28)), with respect to b_i, we get,

$$\frac{\partial p}{\partial b_i} = -\frac{2 r_i B_{tot}\sum_{k=1}^{|N_i|} T_{ik}\sum_{\forall j\neq i} b_j}{E_i(1-\alpha_i)(B_{tot}-\beta)(b_i)^3} - \frac{2 q_i E_i (1-\alpha_i)(B_{tot}-\beta)\sum_{\forall j\neq i} b_j}{\left(\sum_{j=1}^{I} b_j\right)^3}. \quad (29)$$

Similarly, taking the first derivative of the same price function with respect to $\sum_{\forall j\neq i} b_j$, we get,

$$\frac{\partial p}{\partial b_j} = \frac{r_i B_{tot}\sum_{k=1}^{|N_i|} T_{ik}}{E_i(1-\alpha_i)(B_{tot}-\beta)(b_i)^2} - \frac{q_i E_i (1-\alpha_i)(B_{tot}-\beta)\left(\sum_{\forall j\neq i} b_j - b_i\right)}{\left(\sum_{j=1}^{I} b_j\right)^3}. \quad (30)$$

From both the Equations (29) and (30), we conclude that p is a decreasing function with respect to both b_i and $\sum_{\forall j\neq i} b_j$ provided the following inequalities hold:

$$\frac{\partial p}{\partial b_i} < 0, \quad (31)$$

and

$$\frac{\partial p}{\partial b_j} < 0. \quad (32)$$

Evidently, as all the parameters in Equation (29) contain positive values, we say that the inequality in Equation (31) is always true. For proving the other inequality in Equation (32), let us consider that the inequality is true. Therefore,

$$\begin{aligned} &\frac{\partial p}{\partial b_j} < 0 \\ &\Rightarrow \left[q_i E_i (1-\alpha_i)(B_{tot}-\beta)\left(\sum_{\forall j\neq i} b_j - b_i\right)\right] \Big/ \left(\sum_{j=1}^{I} b_j\right)^3 \\ &> \frac{r_i B_{tot}\sum_{k=1}^{|N_i|} T_{ik}}{E_i(1-\alpha_i)(B_{tot}-\beta)(b_i)^2} \\ &\Rightarrow \left(\frac{b_i}{\sum_{j=1}^{I} b_j}\right)^2 - 2\left(\frac{b_i}{\sum_{j=1}^{I} b_j}\right)^3 > \frac{r_i B_{tot}\sum_{k=1}^{|N_i|} T_{ik}}{q_i E_i^2 (1-\alpha_i)^2 (B_{tot}-\beta)^2}. \end{aligned} \quad (33)$$

As, $2(\frac{b_i}{\sum_{j=1}^{I} b_j})^3 > 0$, $\forall b_i$, $i \in I$, we approximate the Equation (33) as,

$$\begin{aligned} \left(\frac{b_i}{\sum_{j=1}^{I} b_j}\right)^2 &> \frac{r_i B_{tot}\sum_{k=1}^{|N_i|} T_{ik}}{q_i E_i^2 (1-\alpha_i)^2 (B_{tot}-\beta)^2} \\ \Rightarrow b_i &> \frac{\sum_{\forall j\neq i} b_j}{\left[\frac{E_i(1-\alpha_i)(B_{tot}-\beta)\sqrt{q_i}}{\sqrt{r_i B_{tot}\sum_{k=1}^{|N_i|} T_{ik}}} - 1\right]}. \end{aligned}$$

Hence, we conclude that the inequality in Equation (32) is satisfied if and only if,

$$b_i > \frac{\sum_{\forall j\neq i} b_j}{\left[\frac{E_i(1-\alpha_i)(B_{tot}-\beta)\sqrt{q_i}}{\sqrt{r_i B_{tot}\sum_{k=1}^{|N_i|} T_{ik}}} - 1\right]}. \quad (34)$$

This concludes the proof. □

6.3 Convergence and Nash Equilibrium

In this Section, we provide the convergence and Nash equilibrium analysis of the proposed algorithm, AQUM.

Lemma 3. *The utility function in Equation (17) is continuous over the interval $0 < b_i \leq B_{tot}$.*

Proof. The basic criterion of a continuous function in a range of values is as follows:

Let $f(x)$ be a real valued function defined on a subset X of the real numbers $\mathbb{R}$, that is $f : X \to \mathbb{R}$. Then, $f(x)$ is said to be continuous at a point $\lambda_0 \in X$, if for any $\epsilon > 0$ there exists a $\mu > 0$ such that for all $x \in X$ with $|x - \lambda_0 < \mu|$, the inequality $|f(x) - f(\lambda_0)| < \epsilon$ is valid.

Likewise, we have a utility function U_i, a real-valued function defined on a subset Y of real numbers $\mathbb{R}$. Specifically, $Y \in (\phi, B_{tot})$, where $\phi > 0$. We assume that there exists a $\mu > 0$ such that for all $b_i \in Y$, and $\lambda_0 \in Y$, the inequality $|b_i - \lambda_0| < \mu$ is valid. Therefore, we get,

$$|U_i(b_i, \mathbf{b}_{-i}, p) - U_i(\lambda_0, \mathbf{b}_{-i}, p)|$$
$$= \left| \frac{q_i E_i(1-\alpha_i)(B_{tot}-\beta)(b_i-\lambda_0)\sum_{\forall j \neq i} b_j}{\sum_{j=1}^{I} b_j(\sum_{\forall j \neq i} b_j + \lambda_0)} + \frac{r_i B_{tot} \sum_{k=1}^{|N_i|} T_{ik}(b_i-\lambda_0)\sum_{\forall j \neq i} b_j}{(B_{tot}-\beta)b_i\lambda_0 E_i(1-\alpha_i)} - p(b_i - \lambda_0) \right|,$$
$$\Rightarrow |U_i(b_i, \mathbf{b}_{-i}, p) - U_i(\lambda_0, \mathbf{b}_{-i}, p)| < \frac{q_i E_i(1-\alpha_i)(B_{tot}-\beta)(b_i-\lambda_0)\sum_{\forall j \neq i} b_j}{\sum_{j=1}^{I} b_j(\sum_{\forall j \neq i} b_j + \lambda_0)} + \frac{r_i B_{tot} \sum_{k=1}^{|N_i|} T_{ik}(b_i-\lambda_0)\sum_{\forall j \neq i} b_j}{(B_{tot}-\beta)b_i\lambda_0 E_i(1-\alpha_i)}$$
$$\Rightarrow |U_i(b_i, \mathbf{b}_{-i}, p) - U_i(\lambda_0, \mathbf{b}_{-i}, p)| < \epsilon,$$

where,

$$\epsilon = \frac{q_i E_i(1-\alpha_i)(B_{tot}-\beta)(b_i-\lambda_0)\sum_{\forall j \neq i} b_j}{\sum_{j=1}^{I} b_j(\sum_{\forall j \neq i} b_j + \lambda_0)} + \frac{r_i B_{tot} \sum_{k=1}^{|N_i|} T_{ik}(b_i-\lambda_0)\sum_{\forall j \neq i} b_j}{(B_{tot}-\beta)b_i\lambda_0 E_i(1-\alpha_i)}.$$

As $b_i \in (\phi, B_{tot})$, $\forall i \in (1, I)$, all the terms in Equation (36) are positive. Hence, $\epsilon > 0$. Therefore, we conclude that U_i is a continuous function over the bid range (ϕ, B_{tot}), where $\phi > 0$. □

Theorem 4. *There exists a Nash Equilibrium for every gateway's individual non-zero bid.*

Proof. Rosen [32] proved the existence of equilibrium point for every concave n-person game. In this problem, taking the second derivative of the utility function shown in Equation (17) with respect to b_i, we get,

$$\frac{\partial^2 U_i}{\partial b_i^2} = -\frac{2 r_i B_{tot} \sum_{k=1}^{|N_i|} T_{ik} \sum_{\forall j \neq i} b_j}{E_i(1-\alpha_i)(B_{tot}-\beta)(b_i)^3} - \frac{2 q_i E_i(1-\alpha_i)(B_{tot}-\beta)\sum_{\forall j \neq i} b_j}{(\sum_{j=1}^{I} b_j)^3}. \quad (37)$$

Substituting the condition shown in Equations (34) and (37), we get,

$$\frac{\partial^2 U_i}{\partial b_i^2} < 0. \quad (38)$$

This implies that the utility function of each gateway is a concave function with respect to its own bid. It is shown in Lemma (3) that the utility function of each gateway is continuous in the range (ϕ, B_{tot}). This implies that there exists a NE, provided the condition shown in the Equation (34) is satisfied. □

Theorem (5) below helps in proving the convergence property of the proposed algorithm, AQUM.

Theorem 5. *The AQUM algorithm concludes in a finite number of iterations.*

Proof. According to the bidding policy, the bidding value is increased by some positive value δ in every iteration. Therefore, $b_i(t+1) = b_i(t) + \delta$. Hence, $b_i(t+1) > b_i(t)$. Again, for a sufficiently large value of t, we have $B_{tot} \leq b_i(t+1)$. Let X define the number of iterations. Then, there exists a finite value of X such that $\sum_{i=1}^{I} b_i(X) \geq B_{tot}$, as $B_{tot} \leq b_i(t+1)$. This implies that the algorithm concludes in a finite number of iterations. □

Lemma 4. *Consider the worst case scenario where all the gateways in a cloud network have no additional request, except their own operational requirement ϕ, which is equal for all the gateways. In this scenario, the maximum number of times a gateway increases the bid value is defined as,*

$$H_{max} = \frac{B_{tot} - I\phi - \beta}{I\delta}. \quad (39)$$

Proof. In each iteration, the bidding value increases by a positive constant δ. Even though there are some iterations possible in the AQUM algorithm, when the bidding value remains same, those iterations are not considered in the calculation of H_{max}. Therefore, at the Hth iteration, the bidding value equals $(\phi + \delta H)$. According to the proposed bidding strategy, the auction concludes, when the sum of the total bid and the reserved bid reaches at least B_{tot}. That is,

$$\sum_{i=1}^{I}(\phi + \delta H) + \beta = B_{tot}$$
$$\Rightarrow H_{max} = \frac{B_{tot} - I\phi - \beta}{I\delta}.$$

This concludes the proof. □

7 Numerical Results

In this Section, we present numerical simulation results of the proposed AQUM algorithm for the MCC environment. Initially, we present an example scenario followed by parameter settings. We show the necessity of bandwidth redistribution, and establish results of convergence and Nash equilibrium. Each experiment was executed 30 times, and the ensemble average values are plotted with 95 percent confidence interval.

7.1 Parameter Settings

Let us consider an MCC environment, as shown in Fig. 1, with one CSP and five gateways $(G_1, \ldots, G_5)$. Each gateway G_i has two connected mobile nodes (N_{i1}, N_{i2}). We consider the total available bandwidth $B_{tot} = 30$ Mbps, $p_{max} = 50$, revenue per unit transmission rate $q_i = 60, \forall i \in I$, and revenue per unit service delay $r_i = 60, \forall i \in I$. In all the experiments, we consider fixed SNR value of 10 dB. The target BER and protocol overhead α of all gateways are considered to be varying. Initially, the ideal transmission time and bandwidth demand for two nodes of each gateway are 0.1

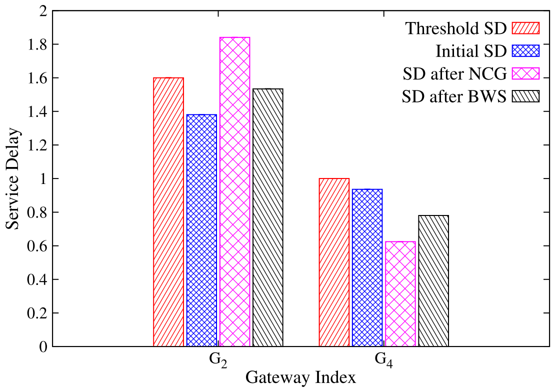

Fig. 2. Service delay comparison. Prior to and after a node changes gateway from G_4 to G_2, and subsequently, prior to and after corresponding bandwidth shifting.

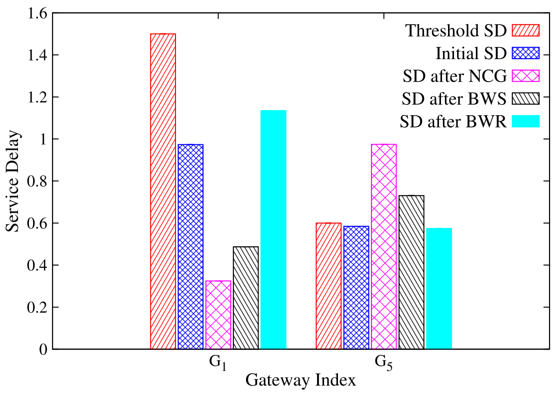

Fig. 3. Service delay comparison. Prior to and after a node changes gateway from G_1 to G_5, and subsequently, after bandwidth shifting followed by redistribution.

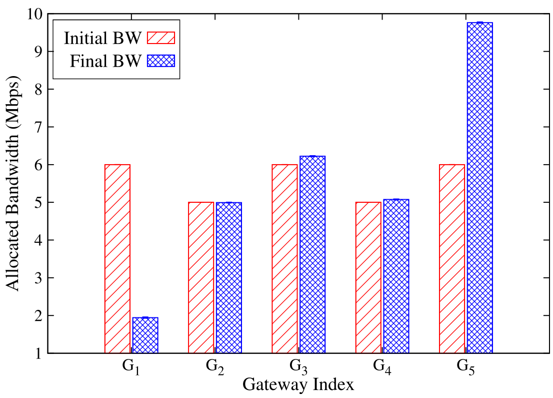

Fig. 4. Final bandwidth allocation.

sec, 0.2 sec, and 1 Mbps, 2 Mbps, respectively. The notations SD, NCG, BWS, and BWR adopted in the legends of the figures define the service delay, node changes gateway, bandwidth shifting, and bandwidth redistribution, respectively. Value of the shifting parameter s depends on delay threshold and revenue per unit service delay.

7.2 Bandwidth Shifting

Suppose the target BER vector for the gateways is $\{10^{-4}$, 10^{-5}, 10^{-4}, 10^{-3}, $10^{-2}\}$, and the protocol overhead vector is $\{0.02,$, 0.02, 0.02, 0.008, $0.008\}$. In this experiment, let, the node N_{41} changes gateway from G_4 to G_2. Fig. 2 shows the necessity for bandwidth shifting, as the service delay of gateway G_2 crosses its threshold value after sharing its own bandwidth with the newly connected node N_{41}. Furthermore, it is evident that bandwidth shifting is sufficient for assuring service delay guarantee in both G_2 and G_4 gateways, while the others remain undisturbed.

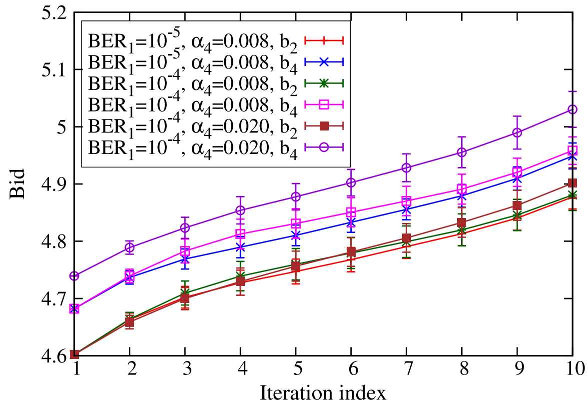

Fig. 5. Bid increment.

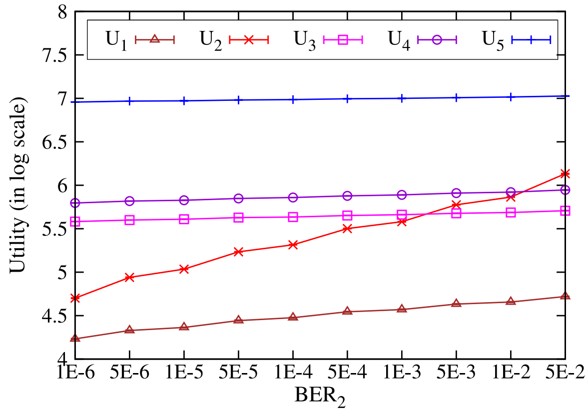

Fig. 6. Utility of each gateway in logarithmic scale.

In the same setup of parameters, instead of node N_{41}, if the node N_{12} changes its gateway from G_1 to G_5, bandwidth shifting alone is not sufficient for providing service delay guarantee, as is evident from Fig. 3. In this experiment, bandwidth redistribution ensures service delay guarantee.

7.3 Bandwidth Redistribution

Considering the same experimental settings, as described in Section 7.2, we executed the AQUM algorithm. The redistributed bandwidth for each gateway is shown in Fig. 4. It is observed that the bandwidth allocation for gateways G_3 and G_4 gets changed due to change in the position of node N_{12}. The redistribution mechanism also supports the service delay requirement of each gateway, as depicted in Fig. 3.

Fig. 5, shows the bidding statistics of gateways G_4 and G_2 for three sets of values of BER and α – $\{BER_1 = 10^{-4}$, $\alpha_4 = 0.008\}$, $\{BER_1 = 10^{-4}$, $\alpha_4 = 0.02\}$, and $\{BER_1 = 10^{-5}$, $\alpha_4 = 0.008\}$. We consider a random distribution for bid increment in the interval $(0.01, 0.1)$. As a gateway either increases or maintains the old bid in the AQUM algorithm, on an average, it increases until the algorithm converges.

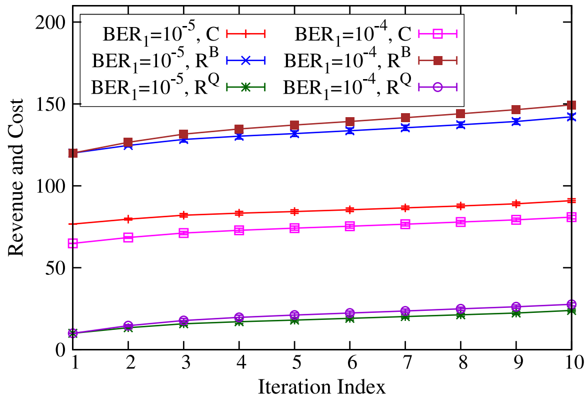

Fig. 7. Revenue and cost under different parameter settings.

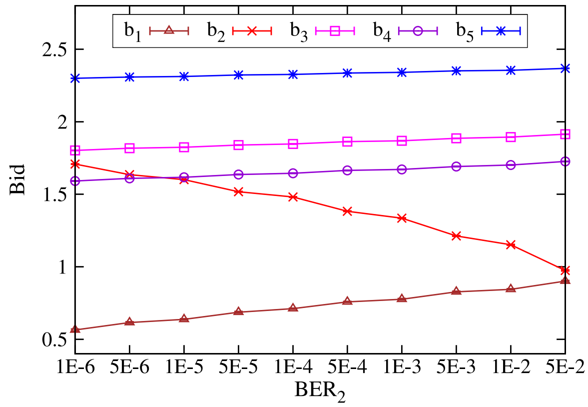

Fig. 9. NE of bid under different channel quality.

Fig. 6 shows the comparison of the computed utility of each gateway. It can be inferred that the utility increases with increase in target BER for any gateway, which again establishes the correctness of Lemma 1. We also measured the revenue earned by a gateway due to bandwidth allocation and QoS provisioning, and the cost it pays for getting bandwidth from the CSP, which is shown in Fig. 7. The figure depicts that the revenue increases, while the cost decreases, with channel quality as intended. Consequently, the utility of a gateway also increases with channel quality.

7.4 Convergence

Fig. 8 shows the convergence of the proposed algorithm, AQUM. In this experiment, we computed the total bid of all the gateways for each iteration cycle. Suppose the default target BER vector for the gateways is $\{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}\}$, and the protocol overhead vector is $\{0.02,, 0.02, 0.02, 0.008, 0.008\}$. We considered two sets of modified values of BER and α, which are $\{BER_1 = 10^{-4},\ \alpha_4 = 0.008\}$, and $\{BER_1 = 10^{-4}, \alpha_4 = 0.02\}$. The BER and α of the other gateways remain fixed. Following illustrates that the AQUM algorithm converges at iteration index 17, 19, and 20, respectively, for the default and the modified sets of BER and α values mentioned above.

7.5 Nash Equilibrium

Fig. 9 depicts that when the channel quality corresponding to gateway G_2 improves, the gateway G_2 decreases its bid. On the contrary, the other gateways either increase their bid or maintain the old value. This indicates the impact of competitiveness among the gateways. It is observed from the figure that the bid trajectories of the gateways intersect at different points, thereby helping in locating the existence of NE.

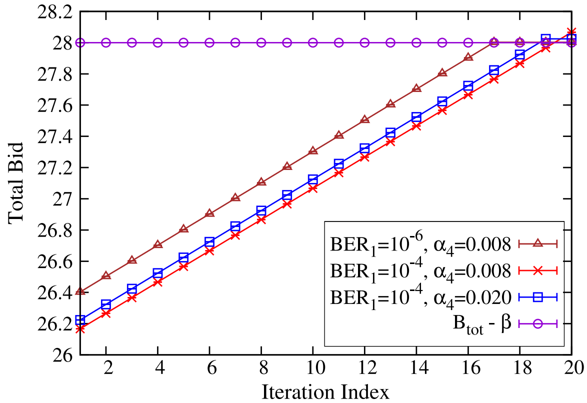

Fig. 8. Convergence of AQUM algorithm.

8 Conclusion

In this paper, we have identified and addressed the problem of bandwidth shifting and redistribution in an MCC environment. The bandwidth redistribution problem differs from traditional bandwidth allocation problem in that while the former concerns allocating proportional bandwidth to all the gateways (and, in turn, to the users), even if only a few gateways change their bandwidth demand, the latter concerns allocating bandwidth to that gateways who have changed the bandwidth demand. We have proposed an auction-based QoS-guaranteed utility maximization algorithm for maximizing the revenue of each gateway, while it maintains QoS of mobile nodes by purchasing bandwidth from the service provider. We have investigated the existence of NE and proved that the algorithm converges within a finite number of iteration.

Even though the proposed algorithm, AQUM, maximizes the utility for performing the auction process, each gateway needs to know the bid value of others, which, in real environment, is not always feasible. A distributed algorithm, therefore, is necessary. Further, we have considered a random bidding increment within a range. Analysis on the selection of different bidding strategies and their implications on utility needs to be discussed in future.

References

[1] A. Amamou, M. Bourguiba, K. Haddadou, and G. Pujolle, "A Dynamic Bandwidth Allocator for Virtual Machines in a Cloud Environment," *Proc. IEEE Consumer Comm. and Networking Conf.*, pp. 99-104, Jan. 2012.

[2] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "Above the Clouds: A View of Cloud Computing," *Comm. ACM*, vol. 53, pp. 50-58, 2010.

[3] C.M. Assi, Y. Ye, and S. Dixit, "Dynamic Bandwidth Allocation for Quality-Of-Service over Ethernet PONs," *IEEE J. Selected Areas in Comm.*, vol. 21, no. 9, pp. 1467-1477, Nov. 2003.

[4] L. Badia, S. Merlin, and M. Zorzi, "Resource Management in IEEE 802.11 Multiple Access Networks with Price-Based Service Provisioning," *IEEE Trans. Wireless Comm.*, vol. 7, no. 11, pp. 4331-4340, Nov. 2008.

[5] R. Chai, X. Wang, Q. Chen, and T. Svensson, "Utility-Based Bandwidth Allocation Algorithm for Heterogeneous Wireless Networks," *Science China Information Sciences*, vol. 56, no. 2, pp. 95-107, 2013.

[6] P. Chaikijwatana and T. Tachibana, "VCG Auction-Based Bandwidth Allocation with Network Coding in Wireless Networks," *Proc. 10th WSEAS Int'l Conf. Applied Computer and Applied Computational Science*, pp. 104-109, 2011.

[7] L. Chen, S. Iellamo, M. Coupechoux, and P. Godlewski, "An Auction Framework for Spectrum Allocation with Interference Constraint in Cognitive Radio Networks," *Proc. IEEE INFOCOM*, pp. 1-9, 2010.

[8] L. Chen, B. Wang, X. Chen, X. Zhang, and D. Yang, "Utility-Based Resource Allocation for Mixed Traffic in Wireless Networks," *Proc. IEEE INFOCOM*, pp. 91-96, 2011.

[9] Y. Chen, Y. Wu, B. Wang, and K.J.R. Liu, "Spectrum Auction Games for Multimedia Streaming over Cognitive Radio Networks," *IEEE Trans. Comm.*, vol. 58, no. 8, pp. 2381-2390, Aug. 2010.

[10] S. Das, S. Misra, M. Khatua, and J.J.P.C. Rodrigues, "Mapping of Sensor Nodes with Servers in a Mobile Health-Cloud Environment," *Proc. IEEE 15th Int'l Conf. E-Health Networking, Applications and Services*, Oct. 2013.

[11] H.T. Dinh, C. Lee, D. Niyato, and P. Wang, "A Survey of Mobile Cloud Computing: Architecture, Applications, and Approaches," *Wireless Comm. and Mobile Computing*, vol. 13, no. 18, pp. 1587-1611, 2013.

[12] J. Elias, F. Martignon, and A. Capone, "An Efficient Dynamic Bandwidth Allocation Algorithm for Quality of Service Networks," *Proc. First IFIP TC6 Int'l Conf. Autonomic Networking*, pp. 132-145, 2006.

[13] R. Fei, K. Yang, S. Ou, and X. Cheng, "A QoS-Aware Dynamic Bandwidth Allocation Algorithm for Base Stations in IEEE 802.16j-Based Vehicular Networks," *Proc. IEEE Conf. Wireless Comm. and Networking*, pp. 1-6, 2010.

[14] N. Fernando, S.W. Loke, and W. Rahayu, "Mobile Cloud Computing: A Survey," *Future Generation Computer Systems*, vol. 29, pp. 84-106, 2013.

[15] A. Goldsmith, *Wireless Communication*. Cambridge Univ. Press, 2005.

[16] A.J. Goldsmith and S.-G. Chua, "Variable-Rate Variable-Power MQAM for Fading Channels," *IEEE Trans. Comm.*, vol. 45, no. 10, pp. 1218-1230, Oct. 1997.

[17] J. Huang, R.A. Berry, and M.L. Honig, "Auction-Based Spectrum Sharing," *J. Mobile Networks and Applications*, vol. 11, no. 3, pp. 405-418, 2006.

[18] L. Jiang, M. Fu, and Z. Le, "Hierarchical QoS-Aware Dynamic Bandwidth Allocation Algorithm for Wireless Optical Broadband Access Network," *Proc. Int'l Conf. Electronics, Comm. and Control*, pp. 4329-4332, 2011.

[19] X. Jin and Y. Kwok, "Cloud Assisted P2P Media Streaming for Bandwidth Constrained Mobile Subscribers," *Proc. IEEE 16th Int'l Conf. Parallel and Distributed Systems*, pp. 800-805, 2010.

[20] J. Joutsensalo, T. Hmlinen, J. Siltanen, and K. Luostarinen, "Delay Guarantee and Bandwidth Allocation for Network Services," *Proc. IEEE Conf. Next Generation Internet Networks*, pp. 129-134, 2005.

[21] E. Jung, Y. Wang, I. Prilepov, F. Maker, X. Liu, and V. Akella, "User-Profile-Driven Collaborative Bandwidth Sharing on Mobile Phones," *Proc. Frist ACM Workshop Mobile Cloud Computing and Services: Social Networks*, Jun. 2010.

[22] B.R. Kao, N.N. Tu, I.S. Hwang, and K.R. Lai, "Auction-Based Bandwidth Allocation in Multi-Hop Wireless Ad Hoc Networks," *Wireless Personal Comm.*, pp. 1-17, 2013.

[23] P.V. Krishna, S. Misra, D. Joshi, and M.S. Obaidat, "Learning Automata Based Sentiment Analysis for Recommender System on Cloud," *Proc. Int'l Conf. Computer, Information and Telecomm. Systems*, May. 2013.

[24] Z. Kun, D. Niyato, and P. Wang, "Optimal Bandwidth Allocation with Dynamic Service Selection in Heterogeneous Wireless Networks," *Proc. IEEE Global Telecomm. Conf.*, 2010.

[25] W.-H. Kuo and W. Liao, "Utility-Based Resource Allocation in Wireless Networks," *IEEE Trans. Wireless Comm.*, vol. 6, no. 10, pp. 3600-3606, Oct. 2007.

[26] W.-H. Kuo and W. Liao, "Utility-Based Radio Resource Allocation for QoS Traffic in Wireless Networks," *IEEE Trans. Wireless Comm.*, vol. 7, no. 7, pp. 2714-2722, July 2008.

[27] J.-W. Lin, C.-H. Chen, and J.M. Chang, "QoS-Aware Data Replication for Data Intensive Applications in Cloud Computing Systems," *IEEE Trans. Cloud Computing*, vol. 1, no. 1, pp. 101-115, Jan.-June 2013.

[28] X. Luo, I. Thng, B. Li, S. Jiang, and L. Yin, "A Dynamic Measurement-Based Bandwidth Allocation Scheme with QoS Guarantee for Mobile Wireless Networks," *Proc. IEEE Conf. Wireless Comm. and Networking*, vol. 3, pp. 947-951, 2000.

[29] K. Maly, C.M. Overstreet, X. Qiut, and D. Tan, "Dynamic Bandwidth Allocation in a Network," *Proc. ACM SIGCOMM Computer Comm.*, vol. 18, pp. 13-24, 1988.

[30] C. Papagianni, A. Leivadeas, S. Papavassiliou, V. Maglaris, C.C. Pastor, and A. Monje, "On the Optimal Allocation of Virtual Resources in Cloud Computing Networks," *IEEE Trans. Computers*, vol. 62, no. 6, pp. 1060-1071, June 2013.

[31] S. Parsons, J.A. Rodriguez Aguilar, and M. Klein, "Auctions and Bidding: A Guide for Computer Scientists," *ACM Computing Surveys*, vol. 43, no. 2, pp. 10:1-10:59, 2011.

[32] J.B. Rosen, "Existence and Uniqueness of Equilibrium Points for Concave N-Person Games," *Econometrica*, vol. 33, pp. 520-534, July 1965.

[33] S. Shelford, G.C. Shoja, and E.G. Manning, "Optimal Bandwidth Allocation for Dynamically Priced Network Services," *Proc. IEEE Third Int'l Conf. Broadband Comm., Networks and Systems*, pp. 1-8, 2006.

[34] J. Tang, R. Hincapi, G. Xue, W. Zhang, and R. Bustamante, "Fair Bandwidth Allocation in Wireless Mesh Networks with Cognitive Radios," *IEEE Trans. Vehicular Technology*, vol. 59, no. 2, pp. 1487-1496, Mar. 2010.

[35] X. Wang, Z. Li, P. Xu, Y. Xu, X. Gao, and H. Chen, "Spectrum Sharing in Cognitive Radio Networks—An Auction-Based Approach," *IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 40, no. 3, pp. 587-596, June 2010.

[36] Y. Zhang, C. Lee, D. Niyato, and P. Wang, "Auction Approaches for Resource Allocation in Wireless Systems: A Survey," *IEEE Comm. Surveys and Tutorials*, vol. 15, no. 3, pp. 1020-1041, Third Quarter, 2013.

[37] Y. Zhang, D. Niyato, and P. Wang, "An Auction Mechanism for Resource Allocation in Mobile Cloud Computing Systems," *Proc. Eighth Int'l Conf. Wireless Algorithms, Systems, and Applications*, 2013.

[38] K. Zhu, D. Niyato, and P. Wang, "Dynamic Bandwidth Allocation under Uncertainty in Cognitive Radio Networks," *Proc. IEEE Global Telecomm. Conf.*, pp. 1-5, 2011.

Sudip Misra is an Associate Professor in the School of Information Technology at the Indian Institute of Technology Kharagpur. He received his Ph.D. degree in Computer Science from Carleton University, in Ottawa, Canada. His current research interests include algorithm design for emerging communication networks. Dr. Misra is the author of over 190 scholarly research papers. He has won *eight research paper awards* in different conferences. He has been awarded the *IEEE ComSoc Asia Pacific Outstanding Young Researcher Award* at IEEE GLOBECOM 2012, Anaheim, California, USA. He was also the recipient of several academic awards and fellowships such as the *Young Scientist Award* (National Academy of Sciences, India), Young Systems Scientist Award (Systems Society of India), Young Engineers Award (Institution of Engineers, India), *(Canadian) Governor General's Academic Gold Medal* at Carleton University, the *University Outstanding Graduate Student Award* in the Doctoral level at Carleton University and the *National Academy of Sciences, India—Swarna Jayanti Puraskar* (Golden Jubilee Award). He was also awarded the Canadian Government's prestigious *NSERC Post Doctoral Fellowship* and the *Humboldt Research Fellowship* in Germany. Dr. Misra is the *Editor-in-Chief* of the International Journal of Communication Networks and Distributed Systems (IJCNDS), Inderscience, Switzerland. He has also been serving as the *Associate Editor* of the Telecommunication Systems Journal (Springer SBM), Security and Communication Networks Journal (Wiley), International Journal of Communication Systems (Wiley), and the EURASIP Journal of Wireless Communications and Networking. He is also an Editor/Editorial Board Member/Editorial Review Board Member of the IET Networks and IET Wireless Sensor Systems. Dr. Misra has 8 books published by Springer, Wiley, and World Scientific. He was invited to chair several international conference/workshop programs and sessions. Dr. Misra was also invited to deliver *keynote/invited lectures* in over 25 international conferences in USA, Canada, Europe, Asia and Africa.

Snigdha Das received the BTech degree in information technology from the West Bengal University of Technology, India, in 2010. She is currently working toward the MS degree at the School of Information Technology, Indian Institute of Technology Kharagpur, India, and is working as a junior project officer in Development of Feasibility Assessment Model for Adaption of Underground Coal gasification Technology in North-Eastern Region of India funded by DeitY, Government of India. She was an associate software engineer in CSC for two years. Her current research interests include mobile computing and cloud computing.

Manas Khatua received the BTech degree in computer science and engineering from the University of Kalyani, India, in 2003, and the MTech degree in information technology from Bengal Engineering and Science University, Shibpur, India, in 2007. He is currently working toward the PhD degree at the School of Information Technology, Indian Institute of Technology, Kharagpur, India, and is working as a senior research fellow in an Underwater Sensor Network project funded by DeitY, Government of India. He has teaching experience of more than two years in an engineering college and served as an assistant system engineer in TCS for two and half years. His research interests include Wi-Fi, wireless sensor networks, and cloud computing. He is a student member of the IEEE.

Mohammad S. Obaidat (S'85-M'86-SM'91-F'05) received the MS and PhD degrees in computer engineering from Ohio State University. He is an internationally well-known academic, researcher, and scientist. He is currently a full professor of computer science with Monmouth University, New Jersey, where he served as the chair of the Department of Computer Science and the director of the Graduate Program. He also served as the advisor to the President of Philadelphia University, and as the president of the Society for Modeling and Simulation International, SCS. He has received extensive research funding and has published more than 30 books and 600 refereed technical articles. He was the recipient of many awards including many Best Paper Awards in IEEE International conferences such as GLOBCOM 2009, DCNET, AICCSA, and CITS. He also received many awards such as SCS prestigious McLeod Founders Award, SCS Presidential Award, IEEE ComSoc-GLOBECOM 2010 Outstanding Leadership Award, Fulbright Distinguished Award, Nokia Research Fellowship Award, and SCS distinguished service award. He is the editor-in-chief of three scholarly journals and is also an editor and an advisory editor of numerous international journals and transactions, including IEEE Journals and Transactions. He has chaired numerous international conferences and given numerous keynote speeches. He is a fellow of SCS. More information is available at http://bluehawk.monmouth.edu/mobaidat.