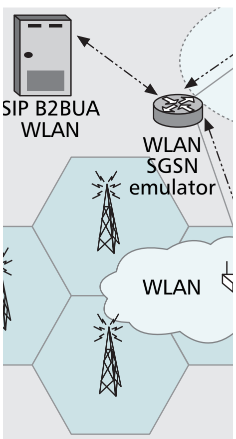
ACCEPTED FROM OPEN CALL

DESIGNING VOIP SESSION MANAGEMENT OVER INTERWORKED WLAN-3G NETWORKS

KUMUDU S. MUNASINGHE, M. RUBAIYAT KIBRIA, AND ABBAS JAMALIPOUR,
UNIVERSITY OF SYDNEY



The authors present a mobility-aware novel interworking network design: inter-connecting Universal Mobile Telecommunications System (UMTS)) and Wireless Local Area Network (WLAN) over an IP based common platform. The design facilitates VoIP session management.

ABSTRACT

In an all-IP internetworked heterogeneous environment, ongoing VoIP sessions from roaming users will be subject to frequent vertical handoffs across network boundaries. Ensuring uninterrupted service continuity for these handoff calls requires successful session management among the participating access networks. As such, a mobility-aware novel interworking network design (interconnecting UMTS and WLAN over an IP-based common platform) [1] is presented in this article that facilitates VoIP session management, including session establishment and seamless session handoff across different networks. For comparison purposes, VoIP session management is evaluated in terms of session establishment, handoff delays, transient packet loss, end-to-end traffic delays, and jitter value for different voice codecs, which demonstrate satisfactory and feasible results. In the event (e.g., network congestion, buffer overflow) that session continuity cannot be guaranteed (also known as outage) across network boundaries, this article proposes an algorithm that compensates the user by reducing the unit service charge of future sessions (governed by the outage period) through a noncooperative game-theory-based pricing mechanism.

INTRODUCTION

Voice over IP (VoIP) has proven to be a spectacular success, and one of the fastest growing segments in the Internet and telecommunications industry. It has been a revolutionary alternative to the traditional telephony system due to its optimal resource utilization and cost efficiency. Although VoIP is well known for its capabilities of supporting voice sessions over wireless technologies such as high-data-rate wireless local area networks (WLANs) and wide coverage third-generation (3G) cellular networks, its behavior has not been sufficiently addressed in the current literature for interworked WLAN-3G systems.

Since interworking affects the VoIP and other 3G cellular service access by a roaming user in a hotspot area via the WLAN and vice versa [2], this article investigates these behaviors in terms of VoIP session establishment delays, vertical handoff delays (with service continuity), transient packet losses, and end-to-end delays. A detailed description of the interworking platform is available in [1]. Unlike other existing interworking proposals [3], the proposed architecture is capable of achieving session continuity during a vertical handoff between WLAN and Universal Mobile Telecommunications System (UMTS) networks, thereby guaranteeing user satisfaction. The novelty of this approach lies in its adoption of the Third Generation Partnership Project's (3GPP) IP multimedia subsystem (IMS) architecture [4] for supporting real-time session negotiation and management with additional controls as inspired by [5, 6]. The IMS, as introduced in UMTS Release 5 within its core network, comprises the required characteristics for controlling real-time multimedia sessions and plays an essential role in the provisioning of IP multimedia services.

The key contribution of this article therefore lies in presenting an in-depth investigation of the behavior and management of a VoIP session for roaming users across network boundaries, and proposing an appropriate design framework (not previously explored) that incorporates the Session Initiation Protocol (SIP) REFER method. An OPNET-based simulation model is developed to analyze the VoIP session establishment and vertical handoff delays from WLAN to UMTS and from UMTS to WLAN, transient packet loss during each handoff scenario, end-to-end VoIP traffic delays, and jitter value for popular voice codecs. Simulation results show that the proper choice of voice codec and the semantics of the vertical handoff mechanism significantly improve the VoIP session management performance. However, such provisioning is incumbent on the availability of service connectivity. In instances when the network experiences dynamic link conditions such as congestion, link failure, and buffer overflow, such an assumption

This work is supported by ARC Linkage Project, in collaboration with SingTel Optus. The review of this manuscript was coordinated by Associate Prof. Ekram Hossain.

 1536-1284/08/$25.00 © 2008 IEEE

fails to offer a comprehensive design architecture. Since VoIP constitutes the minimum service guarantee (among multiple service types, e.g., background traffic, streaming multimedia) during an outage period (t), failure to offer session continuity violates the service level agreement (SLA), especially for differentiated subscription classes such as gold, silver, and economic (as expected in future all-IP networks [7]). To atone for the SLA violation, a noncooperative game-theory-based pricing mechanism is incorporated in the design framework to derive the equilibrium user compensation (cost reduction defined by the Nash equilibrium) across different service types. Each time the session continuity in the interworking architecture is disrupted, the pricing algorithm is invoked to compensate the user for the service loss.

The remainder of this article is organized as follows. We present an overview of the proposed interworked architecture. This is followed by a comprehensive description of the OPNET-based networking simulator and analysis of the performance results. The game-theory-based compensation mechanism is presented next, followed by some concluding remarks.

VoIP over Interworked WLAN-UMTS

This section briefly introduces the underlying WLAN-UMTS interworking mechanism and the implementation of VoIP with specific attention to its vertical handoff mechanism.

Architectural Overview

The WLAN-UMTS interworking platform considered in this article is illustrated in Fig. 1. One of the primary design strategies of this model is that the WLAN is coupled at the general packet radio service (GPRS) gateway support node (GGSN) of the UMTS core network. In fact the WLAN is connected to the GGSN via a serving gateway support node (SGSN) emulator, which connects the WLAN as a separate radio network controller (RNC) to the same UMTS network. Another notable design consideration is that the IMS is used for centralized session management, which enables the UMTS core network to centrally manage both terminal as well as session mobility across heterogeneous networks. Interested readers may refer to [1] for more specific and detailed information on the architectural design.

As depicted in Fig. 1, the data flow that originates from the cellular interface of the source mobile host (MH) gets routed via the SGSN and GGSN to its destination. Regardless of the adopted routing approach, the end-to-end data flow bypasses the IMS network, which is specifically designed for session control signaling. Note that signaling messages such as call setup, call termination, and other session management related functionalities are carried out by the Internet Engineering Task Force's (IETF's) SIP [8]. Unlike the end-to-end data flow, signaling messages from the MH are routed via the call session control functions (CSCFs) of the IMS. When the MH requires to establish a session, this request is always sent to the serving CSCF (S-CSCF) of the home network via the proxy CSCF (P-CSCF) of the visiting network. During the exchange of IMS-SIP signaling (SIP signaling within the IMS structure), both the SGSN and GGSN act as routers by merely forwarding the IMS-SIP messages.

VoIP Session Establishment

Prior to establishing a 3GPP-IMS controlled data session (VoIP in this case), the MH is required to perform a service registration function to update the IMS on its current location. The MH acts as a SIP client and sends a SIP registration message to the relevant S-CSCF via the visiting P-CSCF and home I-CSCF. Following the service registration and activation of the Packet Data Protocol (PDP) context, the MH becomes ready to establish a VoIP session. The message sequence for negotiating a VoIP session managed by the 3GPP IMS is illustrated in Fig. 2. Note that the message sequence flow in Fig. 2 corresponds to the data and signaling flow depicted in Fig. 1.

As can be seen, the session establishment procedure is initiated by a SIP INVITE message sent from the UMTS interface of the source MH (step 1). This initial message is forwarded from the P-CSCF to the S-CSCF of the serving (home or visiting) network to the destination. Next, the destination responds with a 183 Session Progress containing a Session Description Protocol (SDP) answer (if the INVITE contains a request to follow the precondition call flow model) with information of media streams and codecs that the destination is able to accept for this session (step 2). The acknowledgment for the reception of this provisional response by a precondition acknowledgment (PRACK) request follows (step 3). If the destination does not receive a PRACK response within a determined time, it retransmits the provisional response. When the PRACK request successfully reaches the destination, a 200 OK response is generated by the destination with an SDP answer (step 4). Next, an UPDATE request is sent by the source containing another SDP offer, in which the source indicates that the resources are reserved at its local segment (step 5). Upon receiving this UPDATE request, the destination generates a 200 OK response (step 6). Once this is done, the MH can start the VoIP flow, and the session is in progress (via the UMTS interface).

VoIP Session Handoff

As the network selection algorithm triggers the WLAN interface, the need for a vertical handoff mechanism arises for handing off an ongoing VoIP session to the WLAN interface (or vice versa). Furthermore, in order to ensure seamless continuity of services, both terminal/IP and session mobility must be addressed within this vertical handoff framework. Therefore, the SIP REFER method is chosen to exclusively handle vertical handoff, which is capable of providing both terminal and session mobility [9]. The basic steps involved in the SIP REFER-based session handoff process are illustrated in Fig. 2.

Here, the UMTS interface locally notifies the WLAN interface with a SIP REFER request (step 8) via the IMS. The REFER request contains a Refer-To header line containing the destination SIP uniform resource identifier (URI)

In the WLAN-UMTS interworking platform considered in this article, one of the primary design strategies is that the WLAN is coupled at the Gateway GPRS Support Node (GGSN) of the UMTS core network.

and a Replaces header line identifying the existing session to be replaced by the new session. Following this, the WLAN interface sends the corresponding node (CN) an INVITE message with the Replaces header received from the previous REFER request (step 10). Also, the new IP address and port numbers are included in the SDP body of this INVITE message. The receipt of the Replaces header indicates that the initial session is to be replaced by the incoming INVITE request and hence should be terminated. Once the WLAN interface successfully establishes a session with the CN, it sends a NOTIFY request to the UMTS interface updating the final status of the REFER transaction (step 12). This NOTIFY message contains the session information of the newly established session, allowing the UMTS interface to subsequently retrieve the session (if so desired). Once the data flow is established between the WLAN and the CN, the UMTS interface tears down its session with the CN (steps 15 and 16). Note that in the event that the provided information in the Replace header does not match any existing session, the triggered INVITE does not replace the initial session and is processed normally. Thus, any failed session handoff attempt cannot destroy the initial session.

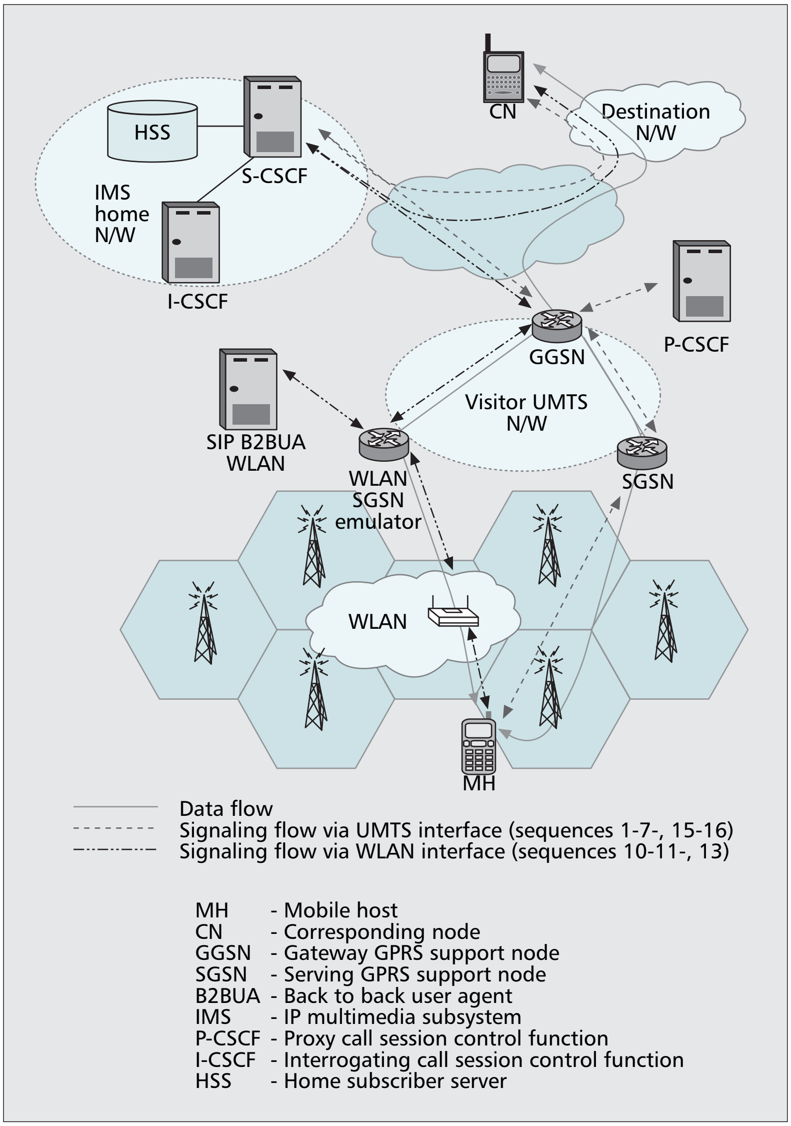


Figure 1. *The WLAN-3G cellular interworking architecture.*

Network Modeling

Simulation Platform

In order to investigate the performance of VoIP session management in the interworking framework, a network simulation model is developed using OPNET Modeler 11.5. Since OPNET's standard SIP components do not adhere to the specifications of the 3GPP's IMS model, substantial modifications are required. As such, modifications are made to SIP proxy servers: user agent server (UAS) to function as different CSCFs, user agent client (UAC) processes to communicate with modified UASs, IMS-SIP based messaging and flow between the CSCFs, roaming facility between multiple domains, and facility for introducing process delay controls (i.e., for messages sent between CSCFs and the home subscriber server [HSS] queries). Thus, a fully functional 3GPP IMS model for OPNET is constructed and integrated with OPNET's existing UMTS Special Module, capable of using IMS-based SIP signaling for session management. This newly developed 3GPP IMS model is an enhanced version of the basic IMS signaling model, which is currently available under the contributed models library of the OPNET University Program [10].

Figure 3 illustrates the constructed simulation scenario. In the proposed model a simple WLAN is connected via an SGSN emulator to the GGSN of the visiting UMTS network. The PCSCF (WLAN) can be seen as a SIP back-to-back user agent (B2BUA), which is capable of interworking with the IMS-SIP and forwarding SIP requests. The S-CSCF is the only IMS node implemented in the home UMTS network. This is because the I-CSCF is mainly used for the SIP registration process, and it is assumed that both UMTS and WLAN interfaces of the MH have already been registered. The CN, which is a SIP UAC, is connected to a destination IP network via a public IP backbone. It is important to note the assumptions made when obtaining these results. Both the UMTS and WLAN belong to different IP subnets, and IP addressing and routing are statically assigned. Since there are no multiple networks available (except for one UMTS and one WLAN), the need for a network selection algorithm has been eliminated. Handoff decisions are individually based on the signal strength of the WLAN, to/from which the MH either roams in or roams out.

As illustrated by Figs. 1 and 2, a SIP REFER-based (or pure SIP-based) signaling and handoff scenario is simulated. A series of VoIP connections are established for simulating VoIP session establishment and vertical handoff delays. Five different types of voice codecs are used for voice traffic generatio: G.711 (data rate of 64 kb/s), G.726 (data rate of 32 kb/s), GSM (data rate of 13 kb/s), G.729 (data rate of 8 kb/s), and G.723.1 (data rate of 5.3 kb/s). Of the above five codecs, G.723.1, GSM, and G.729 are currently the most widely used in GPRS and UMTS systems.

Besides these, the generated voice packets are considered to have a fixed IP header of 40 bytes, which includes a 12-byte Real-Time Transport Protocol (RTP) header, an 8-byte User Datagram Protocol (UDP) header, and a 20-byte IP header. Beyond this, depending on the transmission medium, an additional overhead of 34 bytes are added if the IEEE 802.11 medium access control (MAC) layer is used, and a minimum overhead of 6 bytes are added if the UMTS terrestrial radio access network (UTRAN) is used. Furthermore, no header compression option is considered at the UTRAN Packet Data Convergence Protocol (PDCP) layer, no silence suppression is used, and no playout buffer is used to compensate for jitter.

Simulation Results

The average session establishment and vertical handoff delays for the above two scenarios can be stated as follows. In the case of the SIP REFER method, the average IMS-based VoIP session establishment delays over the UMTS and WLAN interfaces are 165 ms and 152 ms, respectively, whereas the average vertical handoff delays for WLAN to UMTS and UMTS to WLAN are 302 ms and 282 ms, respectively.

For comparison purposes, we also simulate the network under a second scenario where the SIP-ReINVITE method is incorporated as the chosen mobility management mechanism. Since SIP Re-INVITE only provides session handoff and is incapable of providing IP/terminal mobility, a Mobile IP version 4 (MIPv4) home agent (HA) needs to be introduced [11]. Results indicate that the average session establishment delay over the UMTS and WLAN interfaces for a joint MIP-SIP mechanism are 179 ms and 166 ms, respectively. Also, the average vertical handoff delays for WLAN to UMTS and UMTS to WLAN are 180 ms and 94 ms, respectively. It is important to note that despite its relatively low session establishment and vertical handoff delays, substantial changes in the form of adding a MIP HA at the UMTS core network are required. This involves a total overhaul of 3GPP's recommended mobility management specifications, which makes its implementation practically infeasible for the considered scenario in Fig. 1. Interested readers are referred to [12] for detailed descriptions of specific design considerations and scenarios for using a combined SIP-MIP approach.

The above results indicate close correlation between the session establishment delays for both scenarios (i.e., SIP REFER and SIP Re-INVITE). However, there is a substantial difference between the session handoff delays of these two cases. For example, the SIP REFER mechanism experiences around 55 percent higher handoff delay than the MIP-SIP mechanism. This high vertical handoff delay of SIP REFER method can be attributed to the relatively large 3GPP SIP message sizes and heavy application-layer-based IMS latencies (e.g., HSS lookup, SIP REFER-based message exchange, and routing all SIP signaling via the home network). In contrast, SIP Re-INVITE only uses high overhead IMS signaling for session handoff and uses MIPv4 for terminal mobility management, which has a relatively smaller overhead. Such a reduction in signaling related latencies eventually result in a relatively low handoff delay.

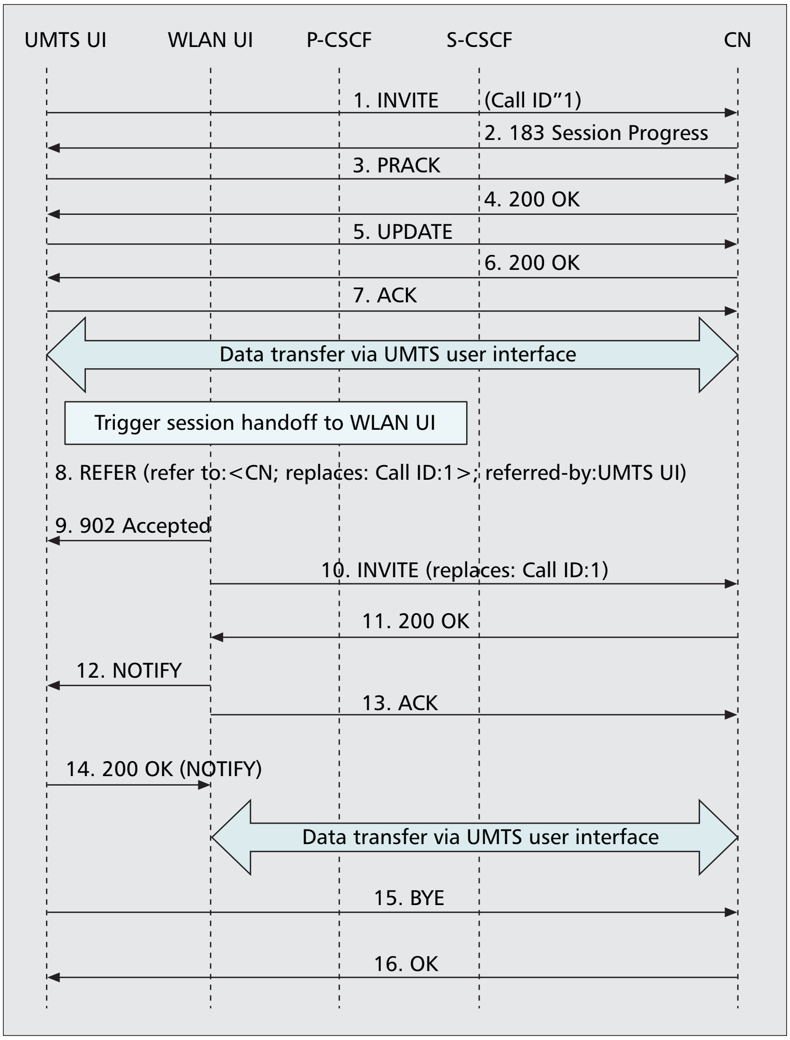
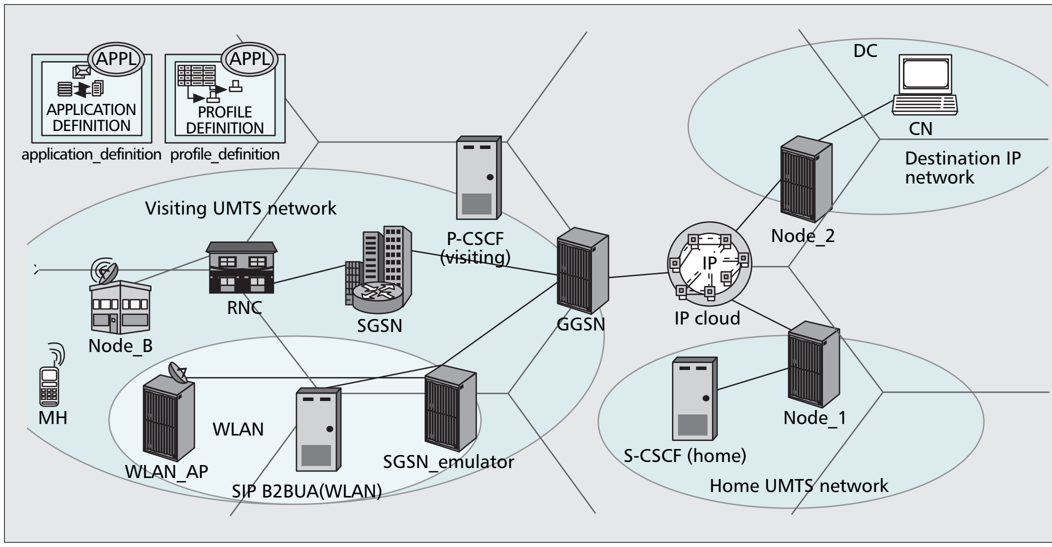


Figure 2. *IMS-SIP-based VoIP session flow.*

Figure 4a and Fig. 4b illustrate the transient packet loss scenarios for WLAN-to-UMTS and UMTS-to-WLAN handoff over different codecs. The packet loss is observed by simulating a break-before-make type handoff as described in detail in [1]. As previously stated, since the SIP REFER method incurs relatively higher vertical handoff delay, it also incurs relatively higher transient packet loss. This is very clearly illustrated in Figs. 4a and 4b. Furthermore, the complicated structure of the UTRAN tends to increase the session setup and vertical handoff delays at the UMTS interface in contrast to the rather simple WLAN, which eventually contributes to relatively higher transient packet loss. It is also expected that with other traffic (e.g., background traffic) present in the network, the processing capability of a node is likely to diminish, resulting in an exponential increase of the vertical handoff delay and packet loss.

Figure 5a, on the other hand, illustrates the end-to-end delay metric investigation for WLAN and UMTS interfaces. This is dependent on the end-to-end VoIP signaling and data paths, the codec, and the payload size of the packets. In

It is evident from these simulation results that the voice capacity over an interworked system is a function of the system parameters, transmission rate, voice packet payload length (depending on the codec used), existing traffic load, and the sampling period.



■ **Figure 3.** *The OPNET simulation model.*

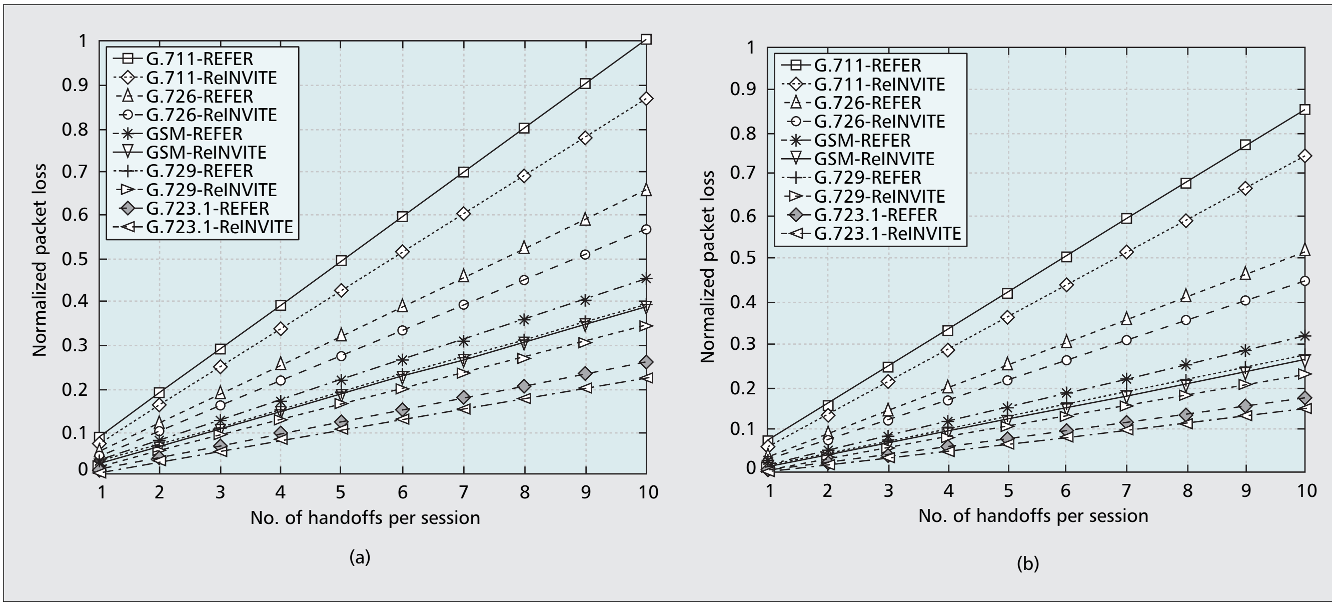
fact, the delays from the endpoints to the codecs at both ends, encoder (processing) delay, algorithmic delay, packetization delay, serialization delay, queuing and buffering delay, and fixed portion of the network delay yield the end-to-end delay for the connection (i.e., WLAN or UMTS). Note that neither SIP REFER nor SIP Re-INVITE has an impact on this metric. Furthermore, longer packetization intervals (sampling periods) result in creating relatively larger voice payloads. These relatively large payloads with proportionately smaller IP header overheads lead to better bandwidth utilization. However, as the period of time required for constructing a single packet increases, the amount of time it takes for this packet to reach the other end and be decoded increases as well. Thus, it can be said that longer packetization intervals may lead to higher latencies. This phenomenon is demonstrated in the case of codec G.723.1 (sampling period 30 ms) in Fig. 5a. Additionally, the fact that G.723.1 is a codec that requires very high processor powers to provide high-quality audio compression has also contributed to its increase in end-to-end delay. Jitter, the variation of interarrival delay, is another factor that affects the end-to-end delay, especially during a vertical handoff (when the link capacity changes, etc.). Figure 5b shows the selection of different VoIP codecs in terms of jitter values.

It is evident from these simulation results that the voice capacity over an interworked system is a function of the system parameters, transmission rate, voice packet payload length (depending on the codec used), existing traffic load, and sampling period. The G.723.1 and G.729 codecs appear to have the worst end-to-end delays. However, this can easily be overlooked since they are significantly below the International Telecommunication Union's (ITU's) recommendation for one-way end-to-end delay threshold of 150 ms. Furthermore, G.723.1 and G.729 also have the lowest jitter values. Conversely, G.711 and G.726 result in relatively high transient packet loss and packet delay variation during vertical handoff. Since G.723.1 shows the lowest transient packet loss and jitter, it can be selected as the most appropriate scheme for the considered framework. However, these performance measures (for both call setup and vertical handoff) are subject to continuous service availability. To provide a comprehensive design architecture, the interworking framework therefore needs to address the outage period t (SLA violation) that ensues from unpredictable network conditions such as link failure, congestion, and buffer overflow. To this effect, a game-theory-based pricing algorithm is presented in the next section that complements the design framework in offering fair service provisioning.

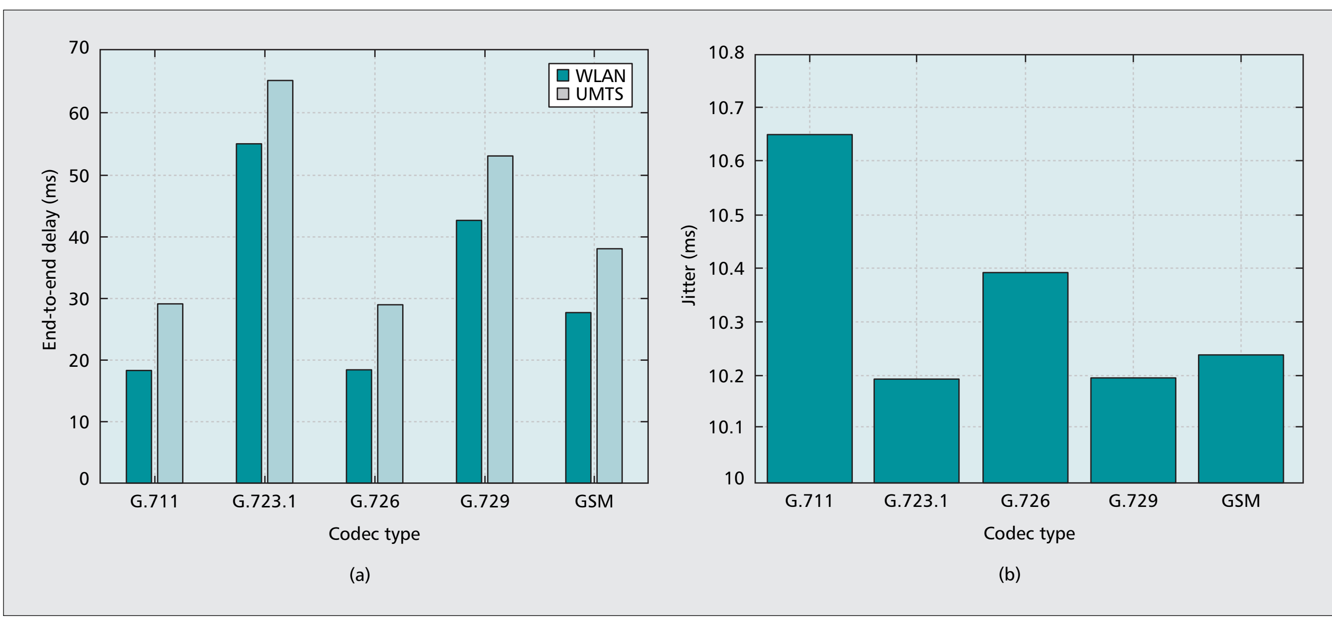
NONCOOPERATIVE GAME-THEORY-BASED OUTAGE COMPENSATION

Within the context of VoIP (basic service) outage in the WLAN-3G interworking framework, SLA violation of "always good connectivity" encompasses the differentiated subscription classes (e.g., gold, silver and economic [7], based on monthly premiums) that guarantee certain levels of perceived quality of service (QoS) from the provider or operator. Game theory [13], in particular noncooperative game theory (preferred over other games for its deterministic nature), offers a viable solution to atone for this outage loss. The atonement is in the form of service cost reduction where a lossy utility function (perceived loss) is utilized to derive the highest amount of cost reduction the network can accommodate, especially in the case of vertical handoff call dropping. Obviously, it forms an integral part of managing seamless VoIP session transfer.

Defined as "a formal modeling approach that studies the situation where multiple players or entities make decisions in an attempt to maximize their returns," noncooperative game theory therefore consists of a set of players, a set of strategies, and a collection of payoffs. In literature it is commonly represented in normal or matrix form (when players act simultaneously) and extensive form (when players act sequential-



■ **Figure 4.** *Transient packet loss during vertical handoff: a) UMTS-to-WLAN handoff; b) WLAN-to-UMTS handoff.*



■ **Figure 5.** *End-to-end delay and jitter comparison: a) average end-to-end delay during vertical handoff; b) average jitter for codec variation during vertical handoff.*

ly), and classified as symmetric (strategies independent of players) and asymmetric (mostly players have nonidentical strategies); zero-sum (one benefits at the expense of others) and non-zero-sum (sum of benefits may not be zero); with perfect (previous strategies known to all players) and imperfect information (previous strategies not a priori), and having pure (deterministic) and mixed strategies (random with probability distribution). Although the term *non-cooperative* refers to the absence of cooperation between the players, it is better identified as a *procedural* game where different actions (or strategies) available to the players are defined a priori. A classic example of non-cooperative game theory is the Prisoner's Dilemma (illustrated in Table 1) where two suspects (A and B) caught in a home robbery are given choices (confess or remain silent) with associated consequences or jail times (two years if they both confess, six months if they both remain silent, and ten years if either of them confesses). Translated into a strategic game theory,[1] their decision can then be formally represented as

$$D_{PD} = [M, \{U_i\}, \{P_i\}],$$

where M is the number of players (in this case two), and U_i and P_i denote the utility and payoff (strategic outcome, subject to the other player's rational decision) function (jail terms), respectively. As shown in Table 1, the payoff function (2,2) is referred to as the Nash equilibrium, which includes each player's best strategy given

[1] *Pure, finite, symmetric, non zero-sum and with imperfect information.*

Implementation wise, every time the terminal experiences service outage for an ongoing session, the compensation algorithm dynamically calculates the maximum cost reduction (Nash equilibrium points in an iterative form).

A

B		Silent	Confess
	Silent	1/2, 1/2	10,0
	Confess	0,10	2,2

Table 1. *Strategic form of Prisoner's Dilemma.*

the other player chooses the equilibrium strategy as well. In other words, when both players are rational and strive to achieve the best result, Nash equilibrium points correspond to distribution outcomes where neither player can benefit by changing their strategies.

Since it is easier to manipulate and hence formulate the game theory as a two-player problem (similar to the Prisoner's Dilemma), the available traffic types (e.g., VoIP, streaming multimedia, background traffic) in the interworked architecture are classified as flat-rate-based (i.e., service charge = cents per session or call) and volume-based (i.e., service charge = cents per megabyte of traffic) (henceforth referred to as service types). Note that this classification is in line with the assumption of end-to-end IP-based services (including voice service) which can be quantified by the session or volume. Assuming Δ_t and Δ_v to denote the flat rate cost and volume-based cost, while the step size cost reduction (interval defined by the network provider or operator) pertaining to the service types are represented by $\Delta_{rt} = \acute{\Delta}, 2\acute{\Delta}, 3\acute{\Delta}, \ldots, \acute{\Delta}_{max} : \acute{n}$ elements; $\acute{\Delta}_{max} \leq \Delta_t$ and $\Delta_{rv} = \bar{\Delta}, 2\bar{\Delta}, 3\bar{\Delta}, \ldots, \bar{\Delta}_{max} : \bar{n}$ elements; $\bar{\Delta}_{max} \leq \Delta_v$ ($\acute{n}$ and $\bar{n}$ may not be equal and is limited by Δ_t and Δ_v), the proposed noncooperative game theory can be described as below as a normal (or strategic), asymmetric, with imperfect information, and zero-sum algorithm.

Players: Player 1 is the cost reduction per unit time (Δ_{rt}), and player 2 is the cost reduction per unit volume (Δ_{rv}).

Strategy: For every outage period, the players reduce the cost of service in step sizes (over the range defined by the network provider or operator) in relation to the outage period t.

Payoff: Payoff is the overall loss/gain (in cents) function the network incurs based on the independent selection by the individual players. The selection will be a Nash equilibrium if and only if none of the players can gain any benefit by changing its strategy, subject to the other player's current choice. To this effect, the Nash equilibrium point corresponds to the highest cost reduction achievable by the players without cooperation.

A categorical description of the compensation algorithm is given in the following subsections.

Lossy Utility Function

In the proposed algorithm, service cost reduction amounts to revenue loss for the network provider or operator. Since this is a perceived utility loss for the network, in this article it is referred to as the *lossy utility*. It is represented by two utility functions reflecting resource utilization and revenue los,s respectively. The former utilizes $w_i.log(1 + kx_i \cdot \Delta_{rt} \cdot t)$ or $w_i.log(1 + kx_i \cdot \Delta_{rv} \cdot t)$ (originating from the concept of fair resource allocation [14]) as the resource utility function, where i denotes the ith player, x_i is the resource utilization ($x_i \leq X_i$ (required resource)), w_i is the weight factor assigned to each player, and k_i is the scaling factor. Since higher resource consumption prior to the outage period translates to larger cost reduction by the network, the term $w_i.log(1 + kx_i \cdot \Delta_{rt} \cdot t)$ or $w_i.log(1 + kx_i \cdot \Delta_{rv} \cdot t)$ reflects this reduction level for resources utilized (x_i). Note that the allocated resource depends on the subscription classes because a higher premium entails a user enjoying more resources. The sigmoid function [7] (often utilized to approximate user satisfaction or dissatisfaction with respect to the perceived QoS), on the other hand, represents the latter (revenue loss utility function); that is,

$$1 - \frac{1}{1 + \exp^{-\alpha(\Delta_y \cdot t - \Delta_{rt} \cdot t - \beta)}} \quad \text{for } \Delta_{rt} \cdot t \leq \Delta_y \cdot t$$

or

$$1 - \frac{1}{1 + \exp^{-\alpha(\Delta_y \cdot t - \Delta_{rv} \cdot t - \beta)}} \quad \text{for } \Delta_{rv} \cdot t \leq \Delta_y \cdot t,$$

where y denotes the random service parameter (revenue loss here), α is the sigmoid function steepness, β is the center of the curve (or the acceptable region of operation), and $x_{\Delta rt}$ and $x_{\Delta rv}$ denote the allocated resources for the service types, subject to $\Delta_{rt} \cdot t \leq \Delta_y \cdot t$ or $\Delta_{rv} \cdot t \leq \Delta_y \cdot t$. By incorporating these utility functions for cost reduction purposes, the lossy utility of flat rate and volume-based cost reduction can be commonly represented as

$$\text{Lossy Utility} = \text{Resource Utility} - \text{Revenue Utility.} \tag{1}$$

Note that the first term in Eq. 1 can be regarded as the utility gain of the lossy function since lower allocated resources (above the minimum requirement) will result in lower cost reduction. On the other hand, α in the second term stipulates the level of subscription: the higher the value of α, the higher the premium (or subscription type) and the perceived QoS. For a VoIP (13 kb/s requirement) outage of t = 2 min, Δ_t = 35 cents/min, and Δ_v = 50 cents/min, $\beta = 0$, w_1, $w_2 = 0.2$, k_1, k_2 = 0:5, and α = {0.1, 0.2, 0.3}, Fig. 6 depicts the lossy utility function of the flat rate service profile. Because of the more stringent SLA regarding the perceived QoS, the penalty for higher premium users (i.e., gold users) is greater than that for the other subscription classes. The decreasing section of the utility function therefore corresponds to the revenue utility exceeding the resource utility gain.

Payoff Function

Here payoff function reflects the loss (total utility) incurred by the network provider or operator in offering cost reduction to the user. Note that the total utility [15] includes the player's individual lossy utility function and the new utility function governing the step size contributions of the two players; that is,

$$\frac{\Delta_{rt}}{\Delta_{rt} + \Delta_{rv}}$$

and

$$\frac{\Delta_{rv}}{\Delta_{rt} + \Delta_{rv}}$$

for flat rate and volume-based, respectively. With such payoff functions, the proposed game theory can be represented in strategic form as in Table 2 where the rows and columns denote Δ_{rt} and Δ_{rv}, respectively, while each element in the matrix gives their corresponding joint payoffs $P(.)$ (scalar values denoting total compensation in cents).

Nash Equilibrium Selection

To find out the solution (i.e., Nash equilibrium: may or may not be a dominant strategy) of the proposed two-player noncooperative game, we utilize the best response function. In game theory, the best response function corresponds to the strategy that produces the highest payoff (reduction in this case) for a player in response to the other player's strategy. Once the best responses of all the players are identified, the Nash equilibrium is derived as the set of best responses (i.e., a set of scalar values) from all the players so that none can benefit by changing their strategies. Mathematically, if B_{rt} and B_{rv} are the best response of the two service types given by $B_{rt} = P(\tilde{\Delta}_{rt}, \Delta_{rv})$ and $B_{rv} = P(\Delta_{rt}, \hat{\Delta}_{rv})$, the solution will be a Nash equilibrium (B^*_{rt}, B^*_{rv}) if and only if $B^*_{rt} = B_{rt}$ and B^*_{rv} subject to the constraint that $(\Delta_{rt} + \Delta_{rv}) \cdot t \leq \Delta_y \cdot t$. Since the strategy profile (Δ_{rt} or Δ_{rv}) is closed (pure strategy), the players are assumed to be rational and attempt to maximize their payoff (offer maximum cost reduction), and have common knowledge about the other player meeting these requirements, the proposed noncooperative game possesses Nash equilibrium solutions.

Implementation-wise, every time the terminal experiences service outage for an ongoing session, the compensation algorithm dynamically calculates the maximum cost reduction (Nash equilibrium points in an iterative form). In graphic representation this can be depicted by the intersection/crossover point of the best response functions for Δ_{rt} and Δ_{rv} respectively. Note that for all forms of service outage, the compensation takes effect immediately after service re-initialization. Figure 7 shows the selection of Nash equilibrium solutions (denoted $Nash_e$, $Nash_s$, $Nash_g$) over different subscriptions classes for a flat rate service outage of $t = 2$ min. The service types are modeled after VoIP (flat rate: 13 kb/s) and streaming multimedia (volume-based: 1 Mb/s), respectively, where the utilized resources are governed by $x \leq X$, X being the required bandwidth. To emulate a real-life scenario, Fig. 7 considers different unit service charges where the charges decrease with respect to higher premium subscriptions, such as economic: Δ_t = 35 cents/min and Δ_v = 50 cents/min;

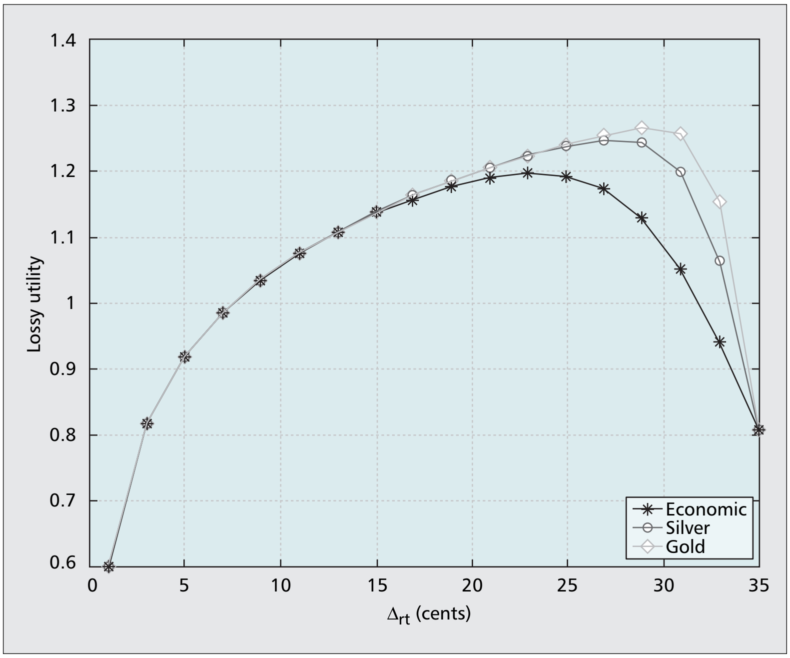


Figure 6. *Lossy utility of Δ_{rt} for VoIP service outage.*

silver: Δ_t = 33 cents/min and Δ_v = 45 cents/min; and gold: Δ_t = 30 cents/min and Δ_v = 40 cents/min. It is evident from the figure that the equilibrium points (Δ_{rt}, Δ_{rv}) for the three subscription classes are given as economic {13, 13}, silver {(16, 9)(13, 13)}, and gold {(19, 5)(16, 9)(10, 13)(7, 17)}. Since there are multiple Nash equilibrium solutions, the smallest point in the solution set is selected to minimize the network loss for individual subscription classes, that is,

$$\arg \min_{\forall R}(B^*_{rt}, B^*_{rv}).$$

The fact that gold subscription offers the highest amount of cost reduction for a flat rate service outage is due to the stringent SLA about the perceived QoS and the corresponding higher penalty value. Not only is the cost reduction higher, but the higher premium also entitles a user to greater cost reduction options, and is in conformance with current commercial practice (for post-paid users). A similar principle is applicable to performance comparisons between silver and economic subscriptions as well. Although the algorithm has been proposed for session discontinuity of either a flat rate or volume-based service profile, it can easily be extended to joint service outage.

Conclusions

It can be concluded that VoIP session management over an interworked WLAN-3G system with the 3GPP's IMS acting as a coupling arbitrator is a promising but very challenging task. In order to investigate the performance measures of VoIP session management over such a system, a novel simulation model (including enhanced IMS signaling) was developed. Simulation results reveal that the application-layer-based additional IMS processing latencies are

	$\bar{\Delta}$	$\cdots$	$k\bar{\Delta}$	$\cdots$	$\bar{\Delta}_{max}$
$\acute{\Delta}$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$2\acute{\Delta}$	$\cdots$	$\cdots$	$P(2\acute{\Delta}, k\bar{\Delta})$	$\cdots$	$\cdots$
$\cdots$	$P(l\acute{\Delta}, \bar{\Delta})$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$\acute{\Delta}_{max}$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$

Table 2. *Strategic form of the compensation algorithm.*

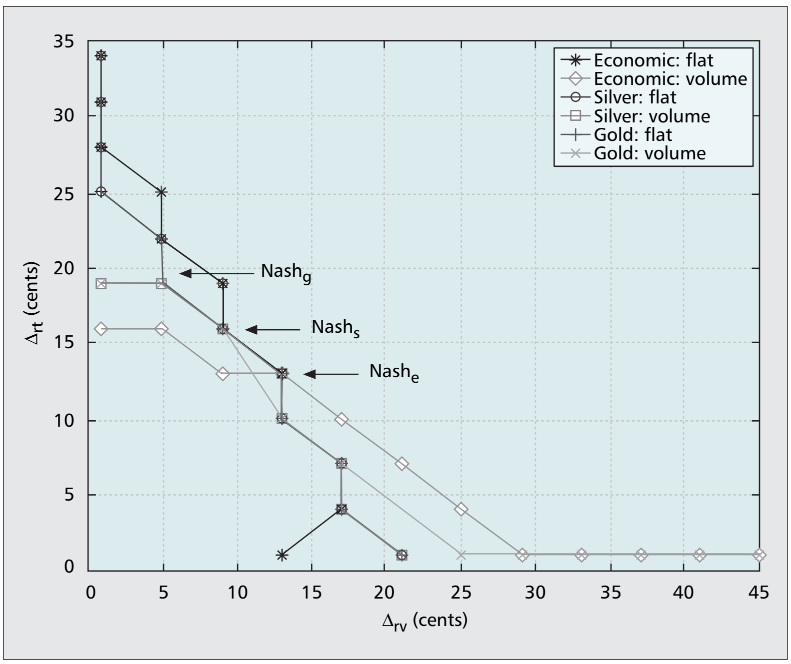


Figure 7. *Comparison of Nash equilibrium solutions for different subscription classes.*

capable of having a substantial impact on these performance measures, except end-to-end delay and jitter values. Instead, both of these parameters for the considered voice codecs indicate acceptable levels for real-time VoIP communications. To redress the issue of service discontinuity from dynamic network conditions (primarily affecting vertical handoff calls), a noncooperative game-theory-based compensation mechanism was introduced that compliments the interworked framework. Here, the Nash equilibrium solutions derived for individual subscription classes are treated as fair recompense for the outage loss. While the Nash equilibrium is not the optimal solution, nevertheless it offers the best possible solutions where neither player benefits by changing their strategy.

References

[1] K. S. Munasinghe and A. Jamalipour, "A 3GPP-IMS based Approach for Converging Next Generation Mobile Data networks," *Proc. IEEE ICC*, Glasgow, U.K., June 2007.

[2] M. Buddhikot *et al.*, "Integration of 802.11 and Third-Generation Wireless Data Networks," *Proc. IEEE INFOCOM*, San Francisco, CA, Apr. 2003.

[3] A. K. Salkintzis, C. Fors, and R. Pazhyannur, "WLAN-GPRS Integration for Next-Generation Mobile Data Networks," *IEEE Wireless Commun.*, vol. 9, 2002, pp. 112–24.

[4] 3GPP, "IP Multimedia Subsystem (IMS)," 3GPP TS 23.228 v. 6.10.0 Rel. 6, 2005.

[5] F. G. Marquez *et al.*, "Interworking of IP Multimedia Core Networks Between 3GPP and WLAN," *IEEE Wireless Commun.*, vol. 12, 2005, pp. 58–65.

[6] W. Wei *et al.*, "SIP-based Vertical Handoff Between WWANs and WLANs," *IEEE Wireless Commun.*, vol. 12, 2005, pp. 66–72.

[7] S. Pal, S. K. Das, and M. Chatterjee, "User-Satisfaction Based Differentiated Services for Wireless Data Networks," *Proc. IEEE ICC*, vol. 2, May 2005, pp. 1174–78.

[8] J. Rosenberg *et al.*, "SIP: Session Initiation Protocol," IETF RFC 3261, 2002.

[9] R. Sparks, "The Session Initiation Protocol (SIP) Refer Method," IETF RFC 3515, 2003.

[10] A. H. Enrique Vazquez and J. I. Fernandez, "SIP-IMS Model for OPNET Modeler," OPNET Univ. Program Contrib. Models, 2005.

[11] C. Perkins, "IP Mobility Support for IPv4," IETF RFC 3344, 2002.

[12] K. S. Munasinghe and A. Jamalipour, "A Unified Mobility and Session Management Platform for Next Generation Mobile Networks," *Proc. IEEE GLOBECOM*, Washington, DC, Nov. 2007.

[13] M. J. Osborne and A. Rubinstein, *A Course in Game Theory*, MIT Press, 1999.

[14] F. P. Kelly, "Models for a Self-Managed Internet," Phil. Trans. Royal Soc., vol. A358, 2000, pp. 2335–48.

[15] D. Niyato and E. Hossain, "Radio Resource Management Games in Wireless Networks: An Approach to Bandwidth Allocation and Admission Control for Polling Service in IEEE 802.16," *IEEE Wireless Commun.*, vol. 14, no. 1, Feb. 2007, pp. 27–35.

Biographies

KUMUDU S. MUNASINGHE [S'03] (kumudu@ee.usyd.edu.au) received his B.Sc. in computer science from the University of Delhi in 1998 with a First Class followed by his M.Comp. and M.Sc. (Hons.) from the University of Western Sydney, Australia, in 2001 and 2005, respectively. He is currently attached to the School of Electrical and Information Engineering at the University of Sydney, Australia, as a research scholar reading toward his Ph.D. in telecommunications engineering. His current research focuses on the interworking architectures between heterogeneous mobile data networks. His research is funded by the Australian Postgraduate Award Industry (APAI) and Norman I Price Scholarships.

M. RUBAIYAT KIBRIA [S'04, M'08] (rkibria@ee.usyd.edu.au) graduated with a B.Sc. Engineering in electrical and electronics from Khulna University of Engineering and Technology, Bangladesh, and an M.Eng. (research) and Ph.D. in telecommunications engineering from the University of Sydney, Australia, in 1997, 2002, and 2007, respectively. He is currently working as a research rellow at the School of Electrical and Information Engineering, University of Sydney, and is actively pursuing research in various areas of next-generation networks, wireless mesh networks, wireless sensor networks, and mobile ad hoc networks. He has served as a TPC member and reviewer on many IEEE journals and conferences.

ABBAS JAMALIPOUR [F] (a.jamalipour@ieee.org) holds a Ph.D. from Nagoya University, Japan. He is the author of the first book on wireless IP and two other books, has co-authored seven books and over 190 journal and conference papers, and holds two patents, all in the field of mobile networks. He is an IEEE Distinguished Lecturer and a Fellow of Engineers Australia. He was Chair of the Satellite and Space Communications TC (2004-2006), and is currently Vice Chair of the Communications Switching and Routing TC and Chair of the Asia-Pacific Board, Chapters Coordinating Committee. He is a Technical Editor of *IEEE Communications Magazine*, Wiley's *International Journal of Communication Systems*, and several other journals. He is a voting member of the IEEE GITC and has been a Vice Chair of IEEE WCNC '03–'06, Chair of IEEE GLOBECOM '05 (Wireless Communications), and a symposium Co-Chair of IEEE ICC '05–'08 and IEEE GLOBECOM '06–'07, among many other conferences. He is the recipient of several international awards including the Best Tutorial Paper Award and Distinguished Contribution to Satellite Communications Award, both from the IEEE Communications Society in 2006.