Wireless Netw (2006) 12:733–745
DOI 10.1007/s11276-006-6055-y

A new architecture for 3G and WLAN integration and inter-system handover management

S. Mohanty

Published online: 27 April 2006

Abstract WLAN has strong potential to provide a perfect broadband complement to the 3G wireless systems. This has raised much interest in their integration. In this paper, a novel architecture using the Network Inter-operating Agent (NIA), and Integration Gateway (IG) is proposed to integrate the 3G systems and WLANs of various providers that may not necessarily have direct service level agreement (SLA) among them. The proposed architecture is scalable as it eliminates the need for the creation of bilateral SLA among the 3G and WLAN operators. In addition, inter-system handover (ISHO) protocols using the concept of the dynamic boundary area is proposed to support seamless roaming between 3G and WLAN. The dynamic boundary area is determined based on the speed of the user and WLAN cell size. The ISHO procedures are initiated when a mobile user enters the boundary area of the WLAN and are completed before the user leaves the coverage area of the serving WLAN. This ensures that the roaming from WLAN to 3G is transparent to the applications. The performance evaluation shows that the proposed boundary area based ISHO algorithm outperforms the existing 3G/WLAN ISHO algorithms.

Keywords Integration of 3G wireless systems and WLANs · Inter-system handover · Handover failure probability

S. Mohanty (✉)
Broadband & Wireless Networking Laboratory,
School of Electrical & Computer Engineering,
Georgia Institute of Technology,
Atlanta, GA 30332
e-mail: shanti@ece.gatech.edu

1. Introduction

Third generation (3G) wireless systems and WLAN technologies are becoming the integral part of the wireless communications. Currently, both technologies are operating independently within their inherent limitations. For example, 3G with ubiquitous coverage supports maximum data rate of only 2 Mbps at a higher cost and WLAN provides data rate up to 100 Mbps at extremely low cost, but only for low mobility users and has local coverage.

The complementary nature of 3G and WLAN [3] has attracted industry, academia, and standard bodies [1, 10] for their integration. The integrated 3G/WLAN system keeps the best features of both 3G and WLAN, i.e., global coverage of 3G, and high-speed and low-cost of WLAN [21]. At the same time, it eliminates the weaknesses of either system. For example, the low data rate limitation of 3G can be overcome when a WLAN coverage is available, through handover of the user to the WLAN. Similarly, when the user moves out of WLAN coverage area, it can be handed over to the overlaying 3G system. The basic idea is to use small-coverage area high-bandwidth WLAN whenever possible else to use 3G.

In the literature several architectures have been proposed to interconnect 3G and WLAN. These can be broadly classified into *tight coupling* (also known as *emulator approach*), *loose coupling* (also known as *Mobile IP approach*), and *no coupling* (also known as *gateway approach*) [24, 26]. In the *tight coupling* architecture, the WLAN network appears to the 3G system as either a radio access network (RAN) in case of GPRS [1, 21, 24] or as a Packet Control Function (PCF) in case of cdma2000 [7]. This approach has the advantages of low handover delay and reduced packet loss [1, 24]. However, because both systems are tightly coupled, it is not flexible and also independently operated WLANs cannot be integrated [7, 24]. Moreover, in this architecture all the packets go through the 3G network. Hence, the 3G network becomes a bottleneck [24], and needs to be redesigned to sustain the increased load [7].

In case of *loose coupling* architecture, mobility management in the integrated 3G/WLAN system is handled using Mobile IP (MIP) protocols [1, 7, 21]. This approach has several advantages such as independent data path for WLAN and 3G traffic, and independent deployment and traffic engineering of WLAN and 3G [7]. However it suffers from many shortcomings including triangular routing if route optimization is not performed [24], high handover delay [16], packet loss, high update latency. Multi-tunnel technology in Mobile IP is used in [16] to reduce the handoff delay and packets loss. Loose coupling architecture requires the authentication, billing, and mobility management mechanisms of 3G and WLAN to inter-operate [7]. It also requires that the 3G and WLAN systems have roaming agreement [7].

The *No coupling* architecture treats 3G and WLAN as peer-to-peer networks. In this case, the legacy mobility management schemes are used to handle intra-system roaming, whereas, the inter-system roaming between two networks having roaming agreement, is performed by a gateway. The gateway converts control signals and routes data packets between two networks for roaming users [23]. In [23] a gateway called virtual GPRS support node (VGSN) is used to integrate 3G and WLAN.

All of the above architectures require the existence of bilateral service level agreement (SLA) between the 3G and WLAN operators. However, architectures requiring bilateral SLA between different 3G and WLAN providers are not feasible because of the following reasons. First, operators have reservations to open their network interfaces to every other operators. Secondly, each time a new operator deploys its WLAN service, it has to be integrated to every other existing operators networks separately. This requires changes to the network infrastructures of all the existing operators. Moreover, schemes requiring bilateral SLA are not scalable [13].

Therefore, a new architecture is required to integrate the 3G systems and WLANs of different providers who may not necessarily have bilateral SLA among them. Once such an architecture is designed the next challenge is to support seamless roaming between the 3G and WLAN networks.

In this paper, we propose a novel architecture to integrate the 3G and WLANs of different providers with or without bilateral SLA among them. We propose the use of a third party, Network Inter-operating Agent (NIA), to integrate these networks. Our architecture is scalable, i.e., can incorporate any number of 3G and WLANs of different service providers. We describe the security and billing mechanisms for our architecture. We analyze both client assisted and network assisted approaches to provide seamless roaming between 3G and WLAN; and advocate the latter as the preferred choice. Then we propose a novel network assisted seamless roaming algorithm using the concept of dynamic boundary area. We define steps for both WLAN to 3G and 3G to WLAN inter system handover (ISHO), and design the associated protocols. In addition, we derive the mathematical formulation of the dynamic boundary area size and carried out the performance evaluation of the proposed network assisted ISHO algorithm.

The rest of this paper is organized as follows. In Section 2, we describe our architecture for the integration of 3G and WLAN. We propose a network assisted algorithm to implement seamless roaming from WLAN to 3G in Section 3. We present a detailed description of the ISHO protocols in Section 4, followed by performance analysis in Section 5. Finally, the advantages of our architecture and boundary area based 3G/WLAN ISHO algorithm are summarized in Section 6.

2. The NIA based 3G/WLAN integrated architecture

Our 3G/WLAN integrated architecture is shown in Fig. 1, consisting of cdma2000[1] networks of two different providers (A and B), their WLANs, and WLAN deployed by a wireless Internet service provider (WISP). It may be noted that our architecture can integrate any number of cdma2000 networks of different providers and their WLANs; other 3G networks of different operators and their WLANs; and also any number of WLANs of different WISPs. We define two new entities **Network Inter-operating Agent** (NIA) and **Interworking Gateway** (IG) that are shown in Fig. 1.

Architectures requiring bilateral SLA among different 3G and WLAN providers are not feasible because of the reasons mentioned earlier. We propose the use of a third party to integrate the 3G and WLANs of different service providers.The NIA in our architecture is the third party and it resides in the Internet. A WLAN provider does not have to create separate bilateral SLA with every other 3G operators. Instead it offers roaming service to users of several 3G operators with only one SLA with the NIA. The NIA handles the authentication, billing and mobility management issues of inter-system roaming. Currently, the Authentication, Authorization, and Accounting (AAA) broker networks support authentication and billing for users belonging to different service providers. But they can not handle the mobility management issues, and hence, can not be used as the third

[1] We use cdma2000 as the reference 3G network to explain our architecture. Our architecture can also integrate other 3G networks such as UMTS. We use the terms 3G and cdma2000 interchangeably in the rest part of this paper.

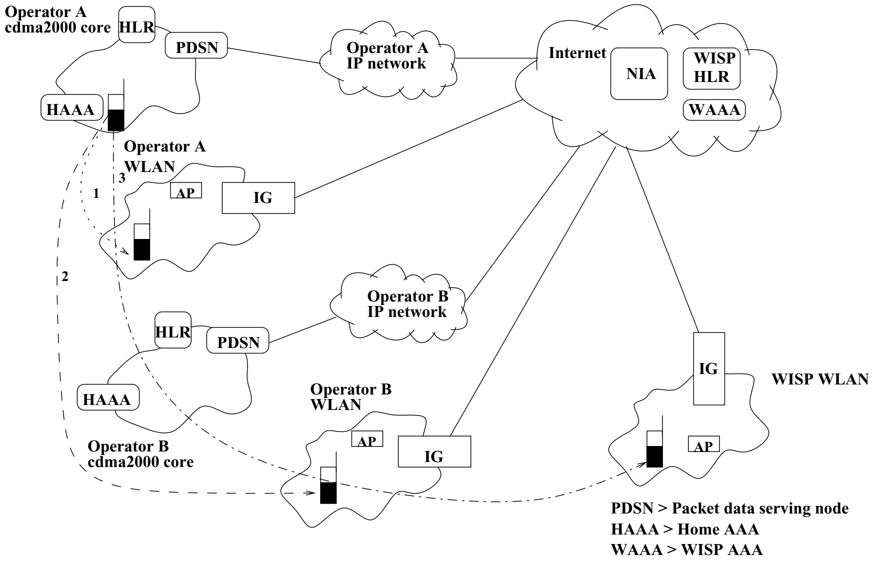

Fig. 1 NIA based integrated 3G/WLAN architecture

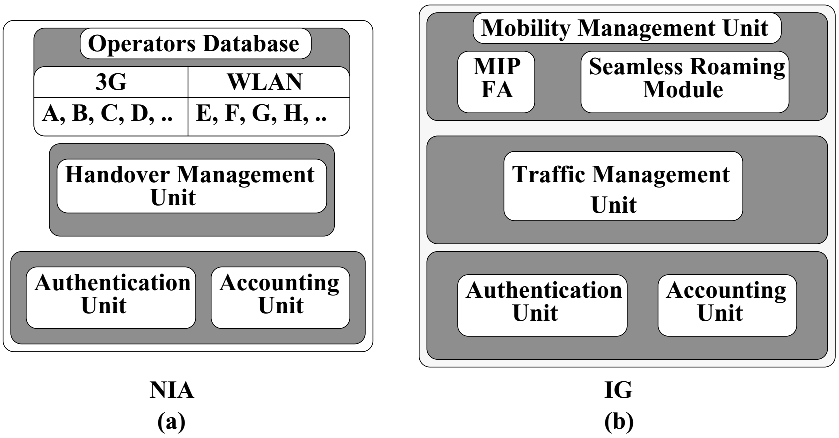

Fig. 2 Logical diagram showing the subsystems of NIA and IG

party. NIA has service level agreements (SLAs) with 3G and WLAN operators. In this work we focus on NIA's role related to mobility management in the integrated 3G/WLAN environment. We also describe how authentication and billing are supported. The sub-systems of the NIA are shown in Fig. 2 (a).

- The *authentication unit* is used to authenticate the users moving between 3G and WLAN belonging to two different service providers (refer to Section 2.1.1).
- The *accounting unit* handles the billing issues between 3G and WLAN as discussed in Section 2.1.2.
- The *operators database* stores information about the 3G and WLAN operators who have SLAs with the NIA.
- The *handover management unit* decides if the MT's[2] ISHO request should be granted or not. For this, it derives the *Network Access Identifier* (NAI) from the Mobile IP *Registration Request* message and verifies with the *operators database* for the existence of SLA with the home operator of the MT. When applicable it also acts as the mediator between 3G and WLAN, e.g., for transfer of user service profile from the 3G to WLAN. Moreover, it stores the locations of the WLANs of various providers and assists the MTs to learn about the available WLANs in their vicinity.

The **Integration Gateway** (IG) functions as the gateway between the WLAN domain and the Internet. Its sub-systms are shown in Fig. 2(b).

- The *mobility management unit* implements the mobile IP [17] (MIP) functionalities using the MIP foreign agent (FA).It also has a *seamless roaming module* which implements the network based mobility management for seamless roaming of users between 3G and WLAN networks as discussed in Section 3.

[2] We use the terms mobile user and mobile terminal (MT) interchangeably in this paper.

- IG implements traffic monitoring function in its *traffic management unit* by discarding the packets coming from unauthorized users.
- The *authentication unit and accounting unit* provide authentiction service and billing support, respectively, to the roaming users (refer to Section 2.1.

The sub-systems of IG other than the *seamless roaming module*, shares functionalities of IOTA gateway proposed in [7].

2.1. Security and billing

The proposed 3G/WLAN integrated architecture provides WLAN operators means to verify the legitimacy of the roaming users. It also provides the operators with suitable billing mechanisms.

2.1.1. Security

Our proposed security architecture for the third party based 3G/WLAN integration is shown in Fig. 3, where the Foreign Network (FN) is a WLAN network and MT's Home Network (HN) is a 3G network. This architecture glues the security architectures of WLAN and 3G through *Authentication Unit* (AU) of NIA (AU_NIA). The use of AU_NIA eliminates the need for any direct security association/agreement between WLAN and 3G networks. Both WLAN and 3G networks have separate security association/agreement with AU_NIA. Thus, AU_NIA functions, in essence, as a trusted third party for authentication dialogs between WLAN and 3G, which do not have security agreement with each other. The working principle of this third party based security architecture is as follows. When a mobile user requests service from a foreign WLAN network and the WLAN determines that it has no SLA with user's home 3G provider, it forwards the request to AU_NIA to authenticate the user. Then, AU_NIA talks to user's home 3G provider and mediates between 3G and WLAN for authentication message exchanges. Once the user is authenticated, AU_NIA also creates security associations/keys required between different network entities. Finally the 3G and WLAN networks will be mutually authenticated, and will have session keys for secured data transfer.

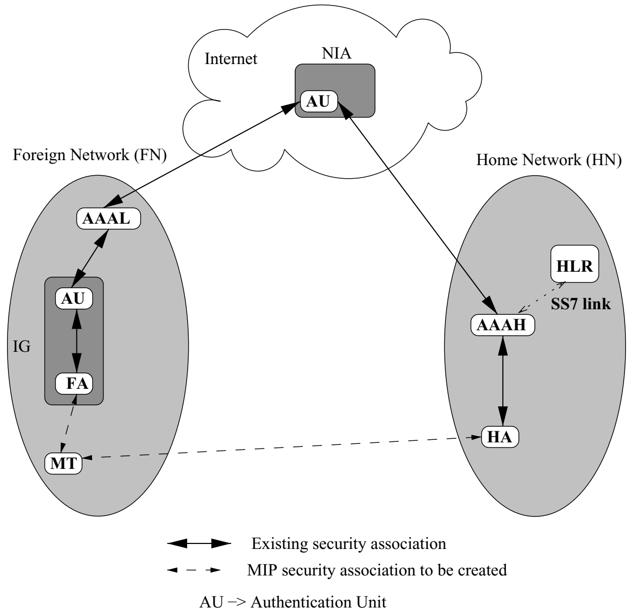
Internet
NIA
AU
Foreign Network (FN)
Home Network (HN)
AAAL
HLR
AU
SS7 link
AAAH
IG
FA
HA
MT
Existing security association
MIP security association to be created
AU –> Authentication Unit

Fig. 3 The proposed security architecture for NIA based 3G/WLAN integrated architecture

We integrate the authentication and Mobile IP registration processes as defined in [13]. The architecture in Fig. 3 shows the existing security associations along with the required MIP security associations so that the Foreign Network (FN) will be able to deliver services to the roaming MT. We use IEEE 802.1x port access control standard [15] for end-to-end mutual authentication between a MT and its home AAA server (AAAH). IEEE 802.1x uses a special frame format known as Extensible Authentication Protocol (EAP) over LAN (EAPOL) for transportation of authentication messages between a MT and an access point (AP). EAP [6] over RADIUS [19] or Diameter [8] is used for the transportation of authentication messages between other entities. When the MT roams into a foreign WLAN domain authentication and MIP registration are carried out as described below. The signaling messages for this are shown in Fig. 4. Here, we use EAP-SIM [12] to illustrate the authentication process. Note that any other authentication schemes, e.g. EAP-AKA [4], EAP-SKE [20], EAP-TLS [2] etc. can also be used.

1) When the MT hears Mobile IP (MIP) *Agent Advertisement* containing *Mobile IP Challenge/Response* extension [18], it sends MIP *Registration Request* including *Mobile IP Challenge/Response* extension and *Mobile-AAA Authentication* extension (as defined in [18]) to the FA located in IG. The MT also includes a *SIM Key Request* extension [11] and a *Network Access Identifier* (NAI) [9], e.g. *MT@relam*, in its MIP *Registration Request*. The *SIM Key Request* extension contains a random number (NONCE_MT) picked up by the MT, which is used for new authentication key generation as discussed later in this Section.
2) When the FA receives the MIP *Registration Request* and finds the *Mobile-AAA Authentication* extension, it learns that the MT is a roaming user and forwards the MIP *Registration Request* to the Authentication Unit of IG (AU_IG). Based on the NAI in the MIP *Registration Request*, the AU_IG recognizes that the WLAN operator does not have direct SLA with the MT's Home Network (HN) and

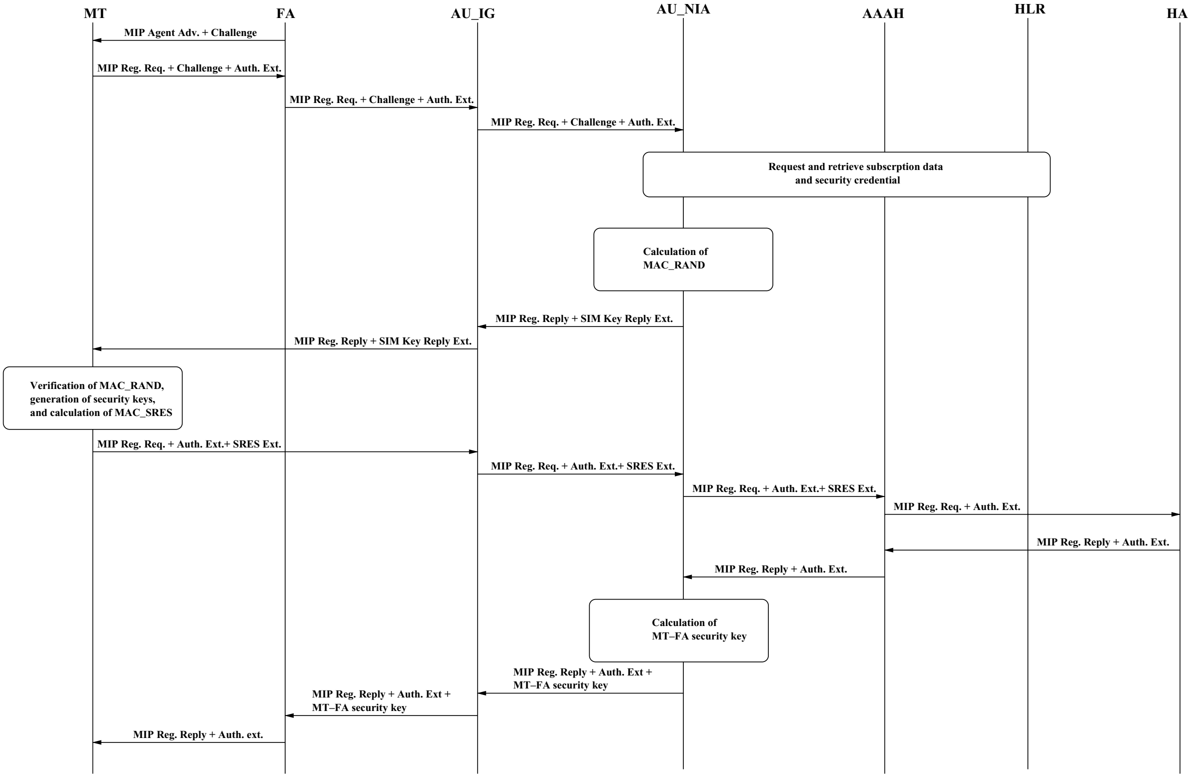
MT
FA
AU_IG
AU_NIA
AAAH
HLR
HA
MIP Agent Adv. + Challenge
MIP Reg. Req. + Challenge + Auth. Ext.
MIP Reg. Req. + Challenge + Auth. Ext.
MIP Reg. Req. + Challenge + Auth. Ext.
Request and retrieve subscrption data and security credential
Calculation of MAC_RAND
MIP Reg. Reply + SIM Key Reply Ext.
MIP Reg. Reply + SIM Key Reply Ext.
Verification of MAC_RAND, generation of security keys, and calculation of MAC_SRES
MIP Reg. Req. + Auth. Ext.+ SRES Ext.
MIP Reg. Req. + Auth. Ext.+ SRES Ext.
MIP Reg. Req. + Auth. Ext.+ SRES Ext.
MIP Reg. Req. + Auth. Ext.
MIP Reg. Reply + Auth. Ext.
MIP Reg. Reply + Auth. Ext.
Calculation of MT–FA security key
MIP Reg. Reply + Auth. Ext + MT–FA security key
MIP Reg. Reply + Auth. Ext + MT–FA security key
MIP Reg. Reply + Auth. ext.

Fig. 4 The authentication signaling messages for 3G/WLAN integrated architecture

forwards the MIP *Registration Request* to the Authentication Unit of NIA (AU_NIA), either directly or through other AAA proxies.

3) The AU_NIA examines the NAI of the received MIP *Registration Request* message and forwards it to MT's Home AAA server (AAAH). Once, AAAH receives the MIP *Registration Request* containing the *SIM Key Request* extension, first it verifies the *Mobile-AAA authentication* extension. If the authentication is successful, it contacts MT's home 3G network elements over SS7 network and obtains n number of triplets (RAND, SRES, Kc). Then it forwards a copy of these triplets to AU_NIA.
4) When AU_NIA receives n triplets it derives a MT_AAAH key ($K_{\text{MT\_AAAH}}$) and calculates message authentication code (MAC) for the RANDs (MAC_RAND) using [11]

$$K_{\text{MT\_AAAH}} = h(n * Kc|NONCE\_MT)$$

$$\text{and} \quad MAC\_RAND = PRF(K_{\text{MT\_AAAH}},\ \alpha) \qquad (1)$$

where α is $n * RAND$ | key lifetime; and $h()$ and $PRF()$ denotes a one-way hash function and a keyed pseudorandom function, respectively. Then, AU_NIA sends the RANDs and MAC_RAND to AU_IG, which forwards those to FA. The FA sends a MIP *Registration Reply* message to the MT containing a *SIM Key Reply* extension. The MIP *Registration Reply* reply message also contains the RANDs, MAC_RAND, and the remaining key lifetime. The MT derives the corresponding SRES and Kc values using its SIM card and the received RANDs. It also calculates ($K_{\text{MT\_AAAH}}$) and MAC_RAND using (1). It validates the authenticity of RANDs by comparing the calculated MAC_RAND with the received MAC_RAND. Thus, confirming that the RANDs are generated by its HN. If the MAC_RAND is valid, the MT calculates a MAC for its SRES values using [11]

$$MAC\_SRES = PRF(K_{\text{MT\_AAAH}}, n * SRES) \qquad (2)$$

The MAC_SRES is used by AU_NIA to know if the SRES values are fresh and authentic. The MT also generates security association keys; ($K_{\text{MT\_FA}}$) for the FA and ($K_{\text{MT\_HA}}$) for the HA using [11]

$$K_{\text{MT\_FA}} = PRF(K_{MT\_AAAH},\ Add_{\text{FA}})$$

$$\text{and} \quad K_{\text{MT\_HA}} = PRF(K_{\text{MT\_AAAH}},\ Add_{\text{HA}}) \qquad (3)$$

where Add_{FA} and Add_{HA} are the IP address of FA and HA, respectively. These keys are used to authenticate subsequent Mobile IP registrations until the key lifetime expires.

5) Now, the MT resends MIP *Registration Request* message to the FA containing *SRES* extension [11] and *Mobile-AAA Authentication* extension. When FA detects the presence of *Mobile-AAA Authentication* extension, it forwards the MIP *Registration Request* message to AU_IG, which forwards it to AU_NIA. AU_NIA calculates MAC_SRES and compares that with the received MAC_SRES. If valid, it forwards the MIP *Registration Request* message to the AAAH. After successful authentication AAAH forwards the MIP *Registration Request* containing $K_{\mathrm{MT\_HA}}$ (calculated using (3)) to the HA. The HA carries out the registration for the MT as defined in [17] and sends MIP *Registration Reply* to AAAH, who forwards it to AU_NIA. AU_NIA calculates MT-FA security key, $K_{\mathrm{MT\_FA}}$, and forwards the MIP *Registration Reply* (containing $K_{\mathrm{MT\_FA}}$ and the Kc keys) to AU_IG. AU_IG forwards it to FA. FA extracts $K_{\mathrm{MT\_FA}}$ and the Kc keys and send a MIP *Registration Reply* to the MT. The Kc keys are used for secure data transfer between the MT and FA providing confidentiality and integrity to the data traffic. If necessary a FA-HA security association key can be generated by AU_NIA using (4) and distributed to the FA and HA as a part of authentication process.

$$K_{\mathrm{FA\_HA}} = PRF(K_{MT\_AAAH}, Add_{\mathrm{FA}}, Add_{\mathrm{HA}}) \tag{4}$$

2.1.2. Billing

Once the MT is authorized by the WLAN, *Accounting Unit* of Integration Gateway (IG) (ACU_IG) maintains a per user accounting record based on the charging policy of the WLAN provider (e.g., connection duration, amount of data transfered etc.). It transfers the accounting information either on per session basis or in real-time to the AAAL server of the WLAN domain. The AAAL server collects and consolidates the accounting information for the MT and forwards it as WLAN access call detail records (WLAN CDRs) to the *Accounting Unit* of NIA (ACU_NIA), which converts it to the CDR format supported by MT's home network and forwards the final CDRs to the AAAH for billing the user.

2.2. Hierarchical NIA

In our architecture, the NIA is involved only during the ISHO process and transfers the control signals between 3G and WLAN. Once the ISHO is over, the data traffic of the roaming users do not go through NIA as discussed in Section 4. Therefore, the load on NIA is limited. We propose hierarchical NIA structure to integrate the 3G and WLAN networks globally. In this hierarchical structure, first the 3G and WLAN networks of various providers are integrated at the regional (e.g. city) level through first tier NIAs. These regional NIAs of a particular country or several countries are then integrated through second tier NIAs, followed by the integration of second tier NIAs through third tier NIAs to realize global 3G and WLAN integration. Exact number of tiers and number of NIAs at each tier depend on several factors, such as number of 3G and WLAN providers in that tier, number of roaming user etc. Determination of the number of NIAs required for a particular deployment scenario can be carried out. This is beyond the scope of this paper. In this hierarchical NIA structure, a 3G or WLAN operator only need to have SLA with the nearest first tier (aka regional) NIA operator to be able to provide its subscribers with global WLAN access.

3. Network assisted algorithm for seamless roaming from WLAN to 3G

3G coverage overlaps the coverage area of WLANs. This means that there is no possibility of connection loss during a 3G to WLAN ISHO. On the other hand, during a WLAN to 3G ISHO if the MT moves out of WLAN coverage before the successful completion of ISHO procedures, it will encounter a connection loss. A client assisted algorithm using the received signal strength (RSS) and the priority of the 3G/WLAN interfaces, is proposed in [7] to implement seamless roaming from WLAN to 3G. The mobile client monitors the RSS of WLAN and switch to 3G when it goes below a threshold. A FFT-based technique is proposed in [27] to trigger a handover from WLAN to 3G when the RSS goes below a threshold value. In these approaches the ISHO procedures must be completed before the WLAN RSS goes from the threshold value to RSS_{min}, i.e., the minimum RSS required for successful communication with a WLAN AP. Else the ISHO process will be unsuccessful and MT will loose its connections. This can happen when the MT is near the boundary of a WLAN and drives out the WLAN coverage area very fast before it is handed over to the 3G system (for example when a user is driving away from the office/airport/campus parking lot, while accessing the WLAN over there). Moreover, in these algorithms, the mobile client always monitors the RSS of the WLAN and 3G interfaces to decide about a possible ISHO. This adds significant unnecessary processing especially when the MT stays inside the WLAN for a long time. This extra processing is costly for power constrained devices such as PDAs, 802.11 phones etc. A typical WLAN user stays inside a WLAN for a long time, especially in offices, universities, airports, shopping malls etc. Therefore, it is unnecessary to carry out the extra processing of monitoring the RSS of the available 3G interface unless the MT is anticipated to move out of the serving WLAN.

We propose a network assisted approach to carry out seamless roaming between 3G and WLAN that eliminates the shortcomings of the above client assisted approaches. We implement this in the *seamless roaming module* of IG. When the *seamless roaming module* learns that the MT starts getting served by a boundary access point (AP), an AP serving a boundary cell of WLAN, it anticipates that the MT may move out of the WLAN coverage area in the near future. Then it estimates the right time to initiate the ISHO process to ensure a successful handoff from WLAN to 3G. The IG determines the right time for WLAN to 3G ISHO initiation using the concept of dynamic boundary area as shown in Fig. 5. The WLAN to 3G ISHO is initiated when an MT enters the boundary area. The size of the boundary area (L_{BA}) is a function of MT's QoS requirement (q), speed (v) and network state (s) as shown in (5).

$$L_{BA} = f(q, v, s) \qquad (5)$$

For simplicity, in this work we consider only MT's speed to estimate the size of the boundary area. It may be noted that this estimate can be easily extended to incorporate the QoS and network state information. We estimate L_{BA} such that ISHO procedures are completed before the MT crosses the WLAN coverage area. Let the time required to complete the ISHO process be τ. During this time an MT with high speed, will travel more distance compared to a slow moving MT. Hence, the ISHO process must start from a farther distance from the boundary of WLAN for a fast moving MT compared to a slow moving one. Therefore, when the speed of the MT is higher the size of boundary area is larger compared to a lower speed case. Detailed procedure to calculate the size of this boundary area is described in Section 5.1. We extend the boundary area beyond the WLAN coverage, such that it is symmetric around the WLAN coverage area. The boundary area beyond the WLAN coverage area is used to avoid the ping-pong effect as described in Section 4.

We use Inter-Access Point Protocol (IAPP) (IEEE Std 802.11f/D5) [14] to detect the association of a MT with a boundary AP. When a MT enters the coverage area of a new AP, it initiates a handover to the new AP. Then, the APME (Access Point management entity) sends the *IAPP-ADD.request* message to the IAPP entity of that AP. When the IAPP receives an *IAPP-ADD.request* message, it sends an IAPP *ADD-notify* packet and a *Layer 2 Update* Frame to the IAPP IP multicast address. This multicast group consists of APs and Layer 2 interworking devices, e.g. bridges and switches of the WLAN domain. The *Integration Gateway* (IG) is a part of this multicast group and hence, it receives the IAPP *ADD-notify* packet and the *Layer 2 Update* Frame. Upon their receipt, in addition to the functions defined in IAPP, the IG also determines if the new AP to which the MT moved, is a boundary AP. For this, the IG maintains a table containing the BSSIDs of the boundary BSSs and the IP address of the corresponding APs. The BSSIDs of the boundary BSSs are available during the WLAN deployment. The mapping between the boundary BSSIDs and the IP addresses of the corresponding APs is done using a RADIUS exchange or locally configured information as defined in [14]. When the IG receives an IAPP *ADD-notify* packet, it checks its Boundary BSSIDs Table. If there is an BSSID in this table with the IP address of the AP received in the IAPP *ADD-notify* packet, then learns that the MT has moved to a boundary AP.

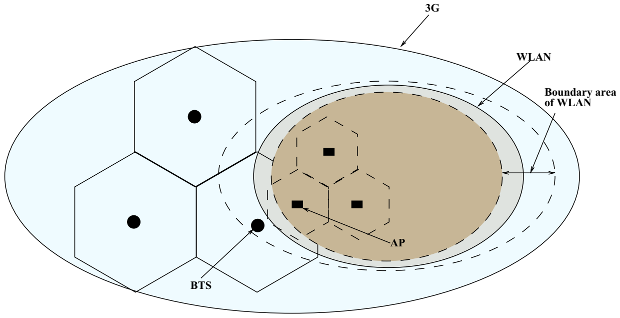

Fig. 5 Dynamic boundary area between WLAN and 3G

4. Inter-system handover protocols

In 3G/WLAN integrated system ISHO can be from WLAN to 3G (henceforth referred as WG_ISHO) or from 3G to WLAN (henceforth referred as GW_ISHO). We divide the entire ISHO process into four phases: *Initiation, Preparation, Start, and Completion*. In the *Initiation* phase, the ISHO process is initiated. Once initiated, the *Preparation* phase prepares the MT for a possible ISHO. Resource allocation in the next system, and alternative route set up are carried out in the *Preparation* phase. Finally, the network decides when to begin the handover and executes the *Start* phase, which is followed by the *Completion* phase. We describe the ISHO protocols in reference to Fig. 1. The ISHO protocols for GW_ISHO are described first followed by those for WG_ISHO.

4.1. ISHO protocols for 3G to WLAN handover

When the MT is served by 3G (e.g. cdma2000), its WLAN interface goes to passive scan mode (where the MT spends only little power) to search for an available WLAN coverage. The MT can avoid the use of passive scan mode to save power and learn about the available WLANs in its vicinity using the *handover management unit* of the NIA. When an MT served by 3G detects the presence of a WLAN, it initiates a handover to the WLAN, i.e., GW_ISHO. The GW_ISHO protocols are illustrated in Fig. 6. These are explained below.

(1) *Initiation*. When an MT detects the presence of a WLAN, it listens for Mobile IP (MIP) *Agent Advertisement* message or sends an MIP *Agent Solicitation* message [17]. It initiates a GW_ISHO by sending an MIP *Registration Request* message to the FA, located in the IG of the corresponding WLAN domain.

(2) *Preparation*. The *Preparation* phase starts once FA receives the *Registration Request*. It carries out the mobile IP registration along with the authentication and authorization operations as discussed in Section 2.1.

(3) *Start*. The *Start* phase is started after the successful registration of the MT with the WLAN. In this phase the MT maintains simultaneous registrations [17] with 3G and WLAN networks as long as it is in a boundary cell of the WLAN. The MT starts receiving packets from its correspondant nodes (CNs) through both 3G and WLAN. But it sends all its traffic through WLAN to take advantage of the higher data rate of WLAN [25]. The CNs communicate with the MT using the MIP procedures [17]. The packets from the CNs are first intercepted by the HA. The HA encapsulates the packets destined for the MT and tunnels those at its care-of-addresses (CoAs). When the MT moves into a non-boundary cell of WLAN, it deregisters from the 3G network. The simultaneous registration during MT's stay in the boundary WLAN cell eliminates the need for a WLAN to 3G handover if the MT moves back to 3G network. Hence, ping-pong effect during ISHO is reduced.

(4) *Completion*. When the IG learns that the MT is no longer in a boundary cell of WLAN, it sends a *release* message (Release) to the MT. The MT acknowledges to this using Release_confm message and deregisters from the 3G network. Then MT's 3G interface goes to the off state.

4.2. ISHO protocols for WLAN to 3G handover

We describe different phases of WLAN to 3G handover (WG_ISHO) below using Fig. 7.

(1) *Initiation*. When an MT moves into a WLAN boundary cell, the *seamless roaming module* of IG anticipates a possible ISHO of the MT into the overlaying 3G system. It estimates the boundary area length (L_{BA}) for the MT as discussed in Section 5.1 and starts to monitor the RSS on both 3G and WLAN interfaces. When the WLAN RSS goes below a dynamically selected threshold value (S_{dth}) (as discussed in Section 5.1), the IG sends an *Inter-system handover warning* (ISHO_warn) message to the MT. Upon the receipt of this message the MT starts the ISHO procedures for its possible handover to 3G while continuing its ongoing connections with the WLAN.

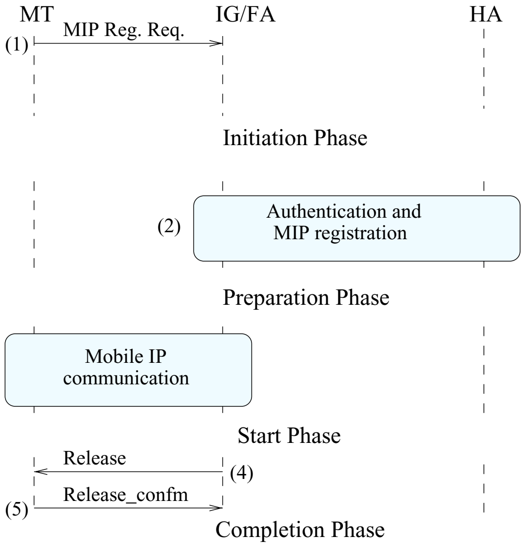

Fig. 6 Signaling messages for GW_ISHO

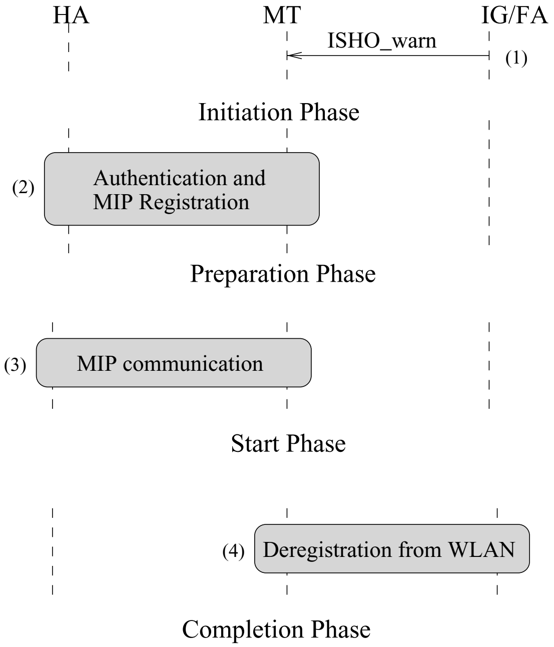

Fig. 7 Signaling messages for WG_ISHO

(2) *Preparation*. In the *Preparation* phase, the MT registers with the 3G network using MIP registration procedures. If the 3G network does not belong to MT's home provider, then 3G roaming protocols are used for this registration. MT also maintains its registration with the WLAN using simultaneous mobility binding to both 3G and WLAN networks.

(3) *Start*. After successful registration with 3G, the MT starts receiving packets from its CNs through both 3G and WLAN. But it sends all its traffic through WLAN as

long as it is within the WLAN coverage area. As the MT is registered with the 3G, its ongoing communications can be immediately switched to 3G when it moves out of WLAN. This ensures a seamless ISHO.

(4) *Completion*. Once the MT moves out of the WLAN coverage, it uses 3G. The IG keeps MT's registration with WLAN active for a timeout duration equal to $\frac{L_{BA}}{v}$, where L_{BA} is the boundary area length and v is the user speed. In this way IG virtually extends the boundary area beyond WLAN's coverage area. If the MT moves back to the WLAN coverage within this time, there is no need to call for GW_ISHO procedures.

5. Performance evaluation

We derive the mathematical formulation for the length of the dynamic boundary area as a function of users' speed. We compare the WG_ISHO failure probability of our dynamic boundary area based ISHO algorithm with the fixed RSS based ISHO algorithm that monitors the RSS of both WLAN and 3G interfaces and initiates a WG_ISHO when the difference of RSS of the interfaces goes below a threshold value. Moreover, we evaluate power consumption and the cost associated with false WG_ISHO initiation for these algorithms. We refer to the boundary area based ISHO algorithm and fixed RSS based algorithm by BA_ISHO and FRSS_ISHO algorithm, respectively.

5.1. Dynamic boundary area length estimation

We assume that while being served by a boundary AP, an MT may enter the boundary area at any point P_1 along the line AC (as shown in Fig. 8) with equal probability. We further assume that user's speed (v) and direction of motion (θ) are uniformly distributed in $[v_{\min}, v_{\max}]$ and $[-\pi, \pi]$, respectively. As the WLAN coverage area is usually much larger than the size of a WLAN cell, we assume that the region ABCD is rectangular. It may be noted that this assumption does not introduce any noticeable error to our analytical model. As θ is uniformly distributed the user may move out of the WLAN coverage at any point P_2 along the cell boundary BD (as shown in Fig. 8) with equal probability. Therefore, probability density function (pdf) of the locations of P_1 and P_2 are given, respectively, by

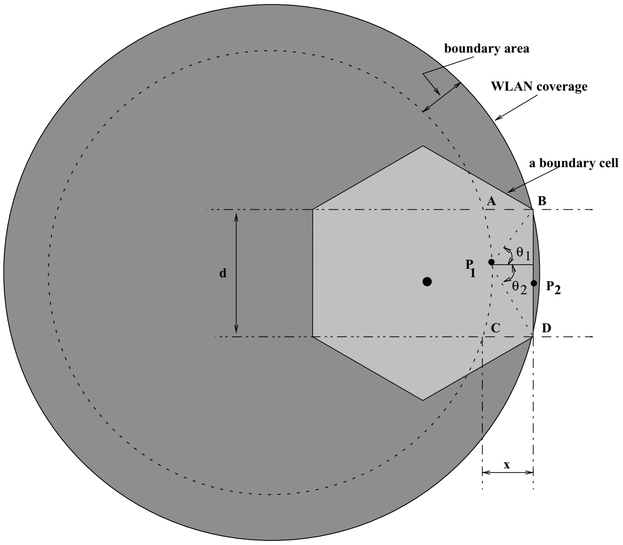

Fig. 8 The boundary region of a WLAN network

$$f_{P_1}(y_1) = \begin{cases} \frac{1}{d} & \text{for } 0 \le y_1 \le d \\ 0 & \text{otherwise,} \end{cases}$$

$$\text{and} \quad f_{P_2}(y_2) = \begin{cases} \frac{1}{d} & \text{for } 0 \le y_2 \le d \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

where d is the length of WLAN cell as shown in Fig. 8. Since the locations of P_1 and P_2 are independent from each other, their joint pdf is given by,

$$f_{P_1} f_{P_2}(y_1, y_2) = \begin{cases} \frac{1}{d^2} & \text{for } 0 \le y_1, y_2 \le d \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

We denote the distance between two random locations of P_1 and P_2 by $L = |Py_1 - P_{y2}|$. The probability that $L \le l$ can be derived using the following integral [5],

$$P(L \le l) = \iint_{\Omega} f_{P_1} f_{P_2}(y_1,\ y_2)\, dy_2\, dy_1 \tag{8}$$

where Ω is the space of locations of P_1 and P_2 such that $L \le l$ and $x \le l \le \sqrt{x^2 + d^2}$, where x is the length of the boundary area. $P(L \le l) = 0$ for $l < x$ and $P(L \le l) = 1$ for $l > \sqrt{x^2 + d^2}$. We can observe that (8) can be rewritten as

$$\begin{aligned} P(L \le l) = \frac{1}{d^2}\Bigg[& \int_0^{\sqrt{l^2-x^2}} \int_0^{\sqrt{l^2-x^2}+y_1} \\ & + \int_{\sqrt{l^2-x^2}}^{d-\sqrt{l^2-x^2}} \int_{-\sqrt{l^2-x^2}+y_1}^{\sqrt{l^2-x^2}+y_1} \\ & + \int_{d-\sqrt{l^2-x^2}}^{d} \int_{-\sqrt{l^2-x^2}+y_1}^{d} \Bigg] dy_2\, dy_1 \\ = & \frac{2}{d}\sqrt{l^2 - x^2} - \frac{l^2 - x^2}{d^2} \\ & \text{for} \quad x \le l \le \sqrt{x^2 + d^2} \end{aligned} \tag{9}$$

The pdf of l can be derived by taking the derivative of (9) and is given by

$$f_L(l) = \begin{cases} \dfrac{2l}{d^2}\left(\dfrac{d}{\sqrt{l^2-x^2}} - 1\right) & \text{for } x \le l \le \sqrt{x^2+d^2} \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

The amount of time a user will take to travel the distance between the points P_1 and P_2 is $T = \frac{L}{v}$. The pdf of T is given by

$$\begin{aligned} f_T(t) &= v f_L(vt) \\ &= \begin{cases} \dfrac{2v^2 t}{d^2}\left(\dfrac{d}{\sqrt{v^2t^2-x^2}} - 1\right) & \text{for } \dfrac{x}{v} \le t \le \dfrac{\sqrt{x^2+d^2}}{v} \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{11}$$

Using (11), the WG_ISHO failure probability is given by

$$p_f = \begin{cases} 1 & \text{for } \tau > \dfrac{\sqrt{x^2+d^2}}{v} \\ p_T(t < \tau) = \dfrac{\sqrt{v^2\tau^2 - x^2}}{d} & \\ \left(2 - \dfrac{\sqrt{v^2\tau^2-x^2}}{d}\right) & \text{for } \dfrac{x}{v} \le \tau \le \dfrac{\sqrt{x^2+d^2}}{v} \\ 0 & \text{for } \tau < \dfrac{x}{v} \end{cases} \tag{12}$$

where τ is the WG_ISHO signaling delay and p_T $(t < \tau)$ is the probability that $t < \tau$. The Eq. (12) shows that zero probability of WG_ISHO failure is achieved for $x > v\ \tau$. Moreover, to guarantee a non-zero WG_ISHO failure ($0 < p_f < 1$) the required value of x can be estimated using,

$$x = [\tau^2 v^2 + d^2(p_f - 2 + 2\sqrt{1-p_f})]^{\frac{1}{2}} \tag{13}$$

(13) is derived from (12) for a particular value of p_f such as $0 < p_f < 1$.

The value of x that is estimated in (13) is the required size of the boundary area, L_{BA}. We determine the WLAN RSS at the entrance of the boundary area, i.e., the RSS at a distance L_{BA} from the boundary of the WLAN coverage, using the path loss model given by [22]

$$\begin{aligned} RSS(x)\ [dBm] = {} & RSS(x_0)\ [dBm] \\ & -10\beta \log_{10}\left(\frac{x}{x_0}\right) + \epsilon\ [dB] \end{aligned} \tag{14}$$

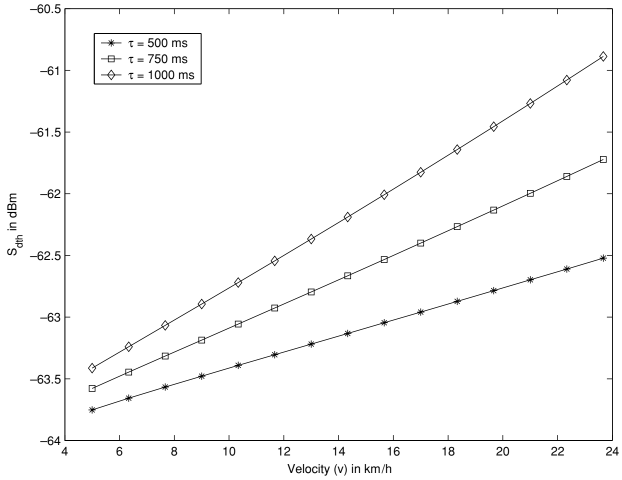

Fig. 9 The value of S_{dth} vs. speed for different value of WG_ISHO signaling delay

where β is the path loss co-efficient, $RSS(x)$ and $RSS(x_0)$ are the RSS at distance of x and a reference distance (x_0), respectively, from an AP. ϵ $[dB]$ is a zero-mean Gaussian random variable with standard deviation σ (typical value of σ is 6 to 8 dB) that represents the statistical variation in $RSS(x)$ caused by shadowing. Using (14) the RSS at the entrance of boundary area that we refer as dynamic RSS threshold (S_{dth}) is given by

$$\begin{aligned} S_{dth}\ [dBm] = {} & RSS_{\min}\ [dBm] \\ & +10\beta \log_{10}\left(\frac{d}{d - L_{\text{BA}}}\right) + \epsilon\ [dB] \end{aligned} \tag{15}$$

where RSS_{min} is the minimum RSS required for the MT to communicate with an AP, i.e., the RSS at the boundary of a WLAN cell. In BA_ISHO the WG_ISHO is initiated when WLAN RSS goes below S_{dth}. We use 914 MHz Lucent WaveLAN DSSS radio interface for which RSS_{min} (i.e., RXThresh) is −64 dBm and $\beta = 4$ for our simulation. Figure 9 shows that the value of S_{dth} increases as the speed increases for a particular value of τ. This is because ISHO must be started earlier for the fast moving users. Moreover, for a particular speed value S_{dth} is increases as τ increases as shown in Fig. 9 as the WG_ISHO procedures must be initiated earlier for higher values of τ.

5.2. WG_ISHO failure probability

To analyze the WG_ISHO failure probability, we assume that the target WG_ISHO failure probability is $p_f = 0.02$. The WG_ISHO failure probability for the BA_ISHO is given by (12) for different values of speed. FRSS_ISHO uses a fixed value of RSS threshold (RSS_f). Therefore, the WG_ISHO is initiated effectively at a distance L_f from the boundary of the

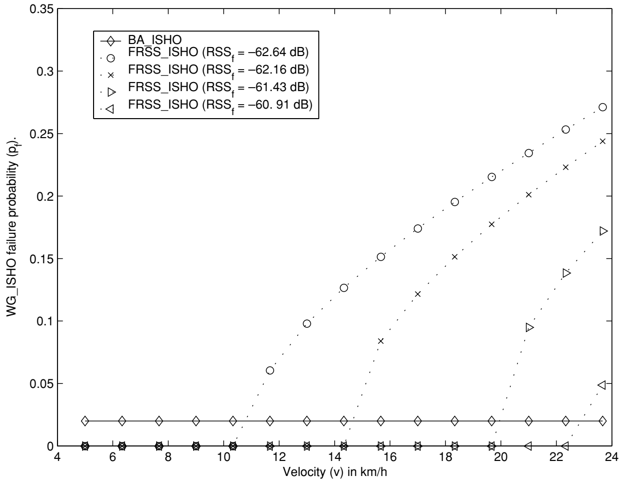

Fig. 10 WG_ISHO failure probability of BA_ISHO algorithm vs FRSS_ISHO algorithm

WLAN coverage where L_f is given by

$$L_f = d\left[1 - 10^{-\left(\frac{\Delta RSS_f + \epsilon}{10\beta}\right)}\right] \tag{16}$$

where $\Delta RSS_f = RSS_f - RSS_{\min}$. Therefore, the WG_ISHO failure probability for FRSS_ISHO algorithm can be calculated using $x = L_f$ in (12). The WG_ISHO failure probability for BA_ISHO and FRSS_ISHO algorithms is shown in Fig. 10 for $\tau = 0.5$ seconds The results show that for FRSS_ISHO algorithm p_f depends on the speed and the value of the RSS_f used. Therefore, the target p_f is achieved only for certain speed values. On the other hand, for BA_ISHO algorithm, p_f is always limited to the target p_f of 0.02 and is independent of speed. Figure 10 shows that for FRSS_ISHO algorithm a higher value of RSS_f reduces p_f. However, in this case the false handoff initiation probability (p_a) increases as discussed in next.

5.3. False WG_ISHO initiation probability

In BA_ISHO and FRSS_ISHO algorithms the WG_ISHO is initiated when the RSS goes below a certain threshold value (either fixed threshold or dynamically selected threshold). This implies that the WG_ISHO is initiated from a particular distance from the boundary of the WLAN coverage. The distance is dynamically chosen for BA_ISHO algorithm and is a fixed value for FRSS_ISHO algorithm. From Fig. 8, it is clear that when started from a distance x from the boundary of the WLAN the need for WG_ISHO arises only if the MT's direction of motion from P_1 is in the range $[-\theta_2, \theta_1]$. Otherwise, the WG_ISHO initiation is a false one. Therefore, the probability of false WG_ISHO (p_a) is given by

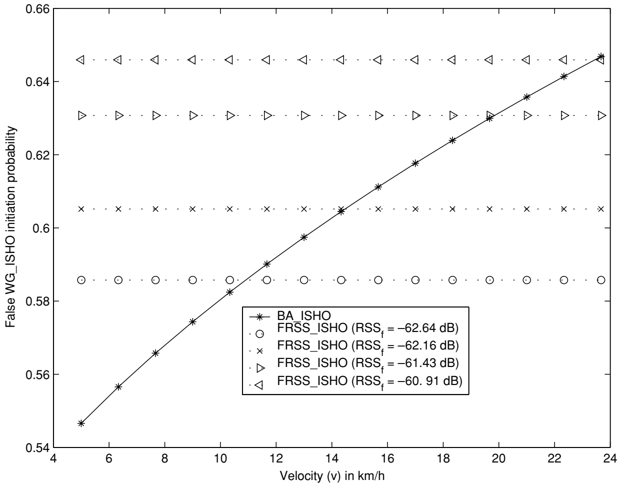

Fig. 11 WG_ISHO false initiation probability

$$\begin{aligned} p_a &= 1 - \frac{1}{d}\int_0^d \left(\frac{\theta_1 + \theta_2}{2\pi}\right) dy \\ &= 1 - \frac{1}{\pi}\arctan\left(\frac{d}{x}\right) + \frac{x}{2\pi d}\ln\left(1 + \frac{d^2}{x^2}\right) \end{aligned} \tag{17}$$

The value of p_a can be calculated for BA_ISHO and FRSS_ISHO algorithms by using $x = L_{BA}$ and $x = L_f$, respectively. Figure 11 shows that for FRSS_ISHO p_a depends on the value of RSS_f. p_a is higher for larger value of RSS_f. Therefore, it is not a good idea to use a unnecessarily large value of RSS_f in a hope to reduce p_f for higher speed values as this will increase the value of p_a for lower speed. This is because for higher RSS_f threshold, the ISHO is initiated too early even when the speed of the user is low. This leads to the wastage of limited wireless network resources. Moreover, this increases the load on the network that arise because of the handoff initiation. On the other hand BA_ISHO algorithm initiates the WG_ISHO in such a way that just enough time is there for successful execution of ISHO procedures for a particular speed. Therefore, the ISHO is neither started too early nor too late. The former limits the high cost associated with unnecessarily large value of false handoff initiation for low speed value. The later ensures that ISHO procedures are smooth even for high speed. Thus, the BA_ISHO algorithm optimizes the WG_ISHO false initiation probability through the dynamic selection of S_{dth} as shown in Fig. 11.

5.4. WG_ISHO Power Consumption

In case of the existing FRSS_ISHO algorithms while inside a WLAN an MT always monitors the RSS on the 3G interface to decide about a possible WG_ISHO initiation. On the other hand in BA_ISHO the terminal monitors the RSS on the

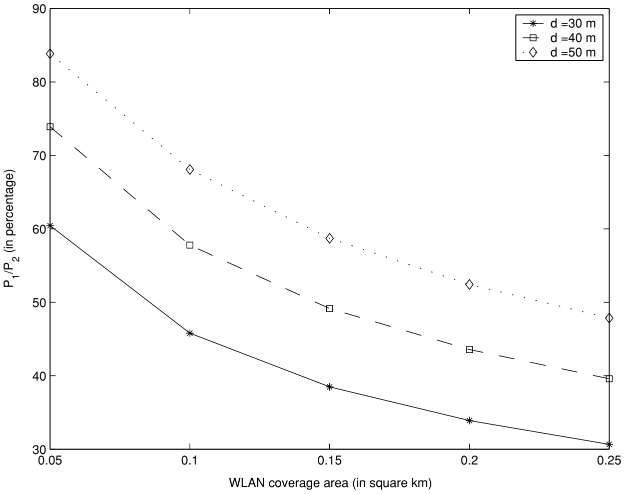

Fig. 12 Comparison of power consumption for RSS monitoring on the 3G interface

3G interface only after moved into a boundary WLAN cell. Therefore, if we assume that the WLAN has a coverage area of A_w and hexagonal cell size of d, then ratio of power consumption because of the RSS monitoring on the 3G interface for BA_ISHO (P_1) and FRSS_ISHO (P_2) is given by

$$\begin{aligned}\frac{P_1}{P_2} &= \frac{(\text{number of boundary WLAN cells})\ \ t_r P}{(\text{total number of WLAN cells})\ \ t_r P}\\ &= \left(\frac{A_c}{A_w}\right)\left[-3+3\sqrt{1+\frac{4}{3}\left(\frac{A_w}{A_c}-1\right)}\right] \qquad (18)\end{aligned}$$

where A_c is the area of each WLAN cell, t_r is the mean WLAN residence time, and P is the power required to monitor the RSS on 3G interface. Figure 12 shows that BA_ISHO achieves significant power saving. The amount of power saving is more for larger WLAN coverage areas.

6. Conclusions

In this paper, we proposed a novel 3G/WLAN integrated architecture using the third party, Network Inter-operating Agent (NIA), to integrate 3G and WLANs of different providers. The proposed architecture does not require the existence of direct SLAs among the network providers. Therefore it is scalable. We developed a novel algorithm using the concept of dynamic boundary area to support seamless ISHO between the 3G and WLAN. In addition, signaling protocols were designed for WG_ISHO and GW_ISHO. The boundary area based 3G/WLAN ISHO algorithm selects a dynamic RSS threshold to initiate the WG_ISHO in such a way that the ISHO procedures are completed before the MT moves out of the WLAN coverage area. Thereby always ensures a successful handoff from WLAN to 3G. Moreover, it optimizes the cost associated with the false handoff initiation. In addition as it does not require the MT to monitor the RSS of the 3G system interface when the MT is served by a WLAN unless the need for WLAN to 3G ISHO arises. Thus it reduces the power consumption associated with the monitoring of 3G interface significantly.

References

[1] 3GPP, 3GPP System to WLAN Interworking; functional and architectural definition, Tech. rep. 3GPP TR 23.934 v0.3.0 (June 2002).

[2] B. Aboba and D. Simon, PPP EAP TLS authentication protocol, IETF RFC 2716 (Oct. 1999).

[3] J. Ala-Laurila, J. Mikkonen, and J. Rinnemaa, Wireless LAN access network architecture for mobile operators, *IEEE Communications Magazine* (Nov. 2001).

[4] J. Arkko and H. Haverinen, EAP AKA authentication, IETF Internet draft, draft-arkko-pppest-eap-aka-09.txt (Feb. 2003), work in progress.

[5] C. Bettstetter, H. Hartenstein, and X. Pérez-Costa, Stochastic properties of the random waypoint mobility model, *ACM/Kluwer Wireless Networks, Special Issue on Modeling and Analysis of Mobile Networks*, 10 (5), (Sept 2004) 555–567.

[6] L. Blunk and J. Vollbrecht, PPP extensible authentication protocol (EAP), RFC 2284, IETF (March 1998).

[7] M. Buddhikot *et al*., Integration of 802.11 and third-generation wireless data networks, In *Proceedings of IEEE INFOCOM 2003*, (Mar–Apr. 2003).

[8] P. Calhoun et al., Diameter base protocol, RFC 3588, IETF, (Sept. 2003).

[9] P. Calhoun and C. Perkins, Mobile IP network access identifier extension for IPv4, RFC 2290, IETF (March 2000).

[10] K. Ahmavaara, H. Haverinen and R. Pichna, Interworking architecture between 3GPP and WLAN systems, *IEEE Communications Magazine*, vol. 41, no. 11, (Nov. 2003).

[11] H. Haverinen, N. Asokan and T. Maattanen, authentication and key generation for mobile IP using GSM authentication and roaming,'in *Proceedings of IEEE ICC (2001)*, vol. 8, pp. 2453–2457.

[12] H. Haverinen and J. Salowey, EAP SIM Authentication, IETF Internet draft, draft-haverinen-pppest-eap-sim-10.txt (Feb. 2003), work in progress.

[13] S. Glass et al., RFC2977-mobile IP authentication, authorization, and accounting Requirements, RFC 2977, (Oct. 2000).

[14] IEEE 802.11f, Draft 4 recommended practice for multi-vendor access point interoperability via an inter-access point protocol across distribution systems supporting IEEE 802.11 Operation, *IEEE Draft 802.11f/D5* (Jan. 2003).

[15] IEEE Std 802.1X-2001, IEEE standard for local and metropolitan area networks - port-based network access control (2001).

[16] Y. Min-hua, L. Yu and Z. Hui-min,"The mobile IP handoff between hybrid networks, in: *Proceedings of IEEE PIMRC 2002* (Sept. 2002).

[17] C. Perkins, IP Mobility Support for IPv4, RFC 3220, IETF, (Jan. 2002).

[18] C. Perkins et al., Mobile IPv4 challenge/response extensions (revised), draft-ietf-mobileip-rfc3012bis-05.txt, Internet Draft, work in progress (May 2003).

[19] C. Rigney et al., Remote Authentication Dial In User Service (RADIUS), RFC 2865 (June 2000).

[20] L. Salgarelli, EAP SKE authentication and key exchange protocol, Internet Draft, draft-salgarelli-pppext-eap-ske-03.txt (May 2003), work in progress.

[21] A.K. Salkintzis, C. Fors and R. Pazhyannur, WLAN-GPRS Integration for Next-Generation Mobile Data Networks, *IEEE Wireless Communications* (Oct. 2002).

[22] G.L. Stuber, *Principles of Mobile Communication*, Kluwer Academic Publishers.

[23] S. Tsao and C. Lin, VGSN: A gateway approach to interconnect UMTS/WLAN networks, in: *Proceedings of IEEE PIMRC 2002* (Sept. 2002).

[24] S. Tsao and C. Lin, Design and evaluation of UMTS-WLAN interworking strategies, in: *Proceedings of VTC 2002* (Sept. 2002).

[25] V. K. Varma, S. Ramesh et al., Mobility management in integrated 3G/WLAN networks, In: *Proceedings of ICC '03*, (May 2003).

[26] D. Wisely and E. Mitjana, Paving the Road to Systems Beyond 3G-The IST BRAIN and MIND Projects, *Journal of Communications and Networks*, 4, (4) (Dec. 2002) 292-301.

[27] Q. Zhang, C. Guo, Z. Guo, and W. Zhu, efficient mobility management for vertical handoff between WWAN and WLAN, *IEEE Communications Magazine*, 41 (11) (Nov. 2003).

Shantidev Mohanty (SM'04) received his B. Tech. (Hons.) degree from the Indian Institute of Technology, Kharagpur, India and the M.S. degree from the Georgia Institute of Technology, Atlanta, Georgia, in 2000 and 2003, respectively, both in electrical engineering. He is currently a graduate research assistant with the Broadband and Wireless Networking Laboratory and a Ph.D. candidate at the School of Electrical and Computer Engineering, Georgia Institute of Technology. His current research interests include wireless networks, mobile communications, mobility management, ad-hoc and sensor networks, and cross-layer protocol design. From 2000 to 2001 he worked as a mixed signal design engineer for Texas Instruments, Bangalore, India. He worked as a summer intern for Bell Labs, Lucent Technologies, Holmdel, New Jersey, during the summers of 2002 and 2003 and for Applied Research, Telcordia Technologies, Piscataway, New Jersey, uring the summer of 2004.