Adaptive handoff scheme for heterogeneous IP wireless networks

Christian Makaya *, Samuel Pierre

Abstract

Recent technological advances allow mobile devices to be equipped with multiple wireless interfaces. Moreover, the coexistence of diverse but complementary architectures and wireless access technologies is one of the major trends in 4G or next-generation wireless networks (4G/NGWN). In this context, the selection of an appropriate interface to ensure that a mobile node (MN) remains connected to the network is a challenging issue for seamless roaming. Furthermore, mobility management as well as the integration and interworking of existing wireless systems are complex due to their specific characteristics. This paper proposes an efficient handoff protocol, called Handoff Protocol for Integrated Networks (HPIN), which alleviates services disruption during handoff in 4G/NGWN. HPIN is based on a novel handoff decision function and carries out localized mobility, fast handoff and access networks discovery. Performance evaluation based on numerical results shows that the proposed protocol performs better than existing schemes.

© 2008 Elsevier B.V. All rights reserved.

Keywords: Mobility management; Quality of service; Next-generation wireless networks; Vertical handoff; Seamless roaming

1. Introduction

Next-generation wireless networks (NGWN) are expected to exhibit heterogeneity in terms of wireless access technologies, personalized and user-oriented services, application requirements, high usability and increased capacity. With NGWN, users will have greater demands for seamless roaming across different wireless networks, support of various services (e.g., multimedia applications) and quality of service (QoS) guarantees. The advantages of 3G cellular networks, such as UMTS and CDMA2000, reside in their global coverage while their weaknesses lie in their bandwidth capacity and operation costs. On the other hand, WLAN technology, such as IEEE 802.11, offers higher bandwidth coupled with low operation costs, although it covers a relatively short range. These existing wireless networks have been subject of extensive individual investigations. Moreover, technological advances in the evolution of portable devices made possible the support of different radio access technologies (RATs) under multi-homing concepts. This has raised much interest for the integration and interworking of 3G wireless networks and WLAN capable of providing integrated authentication, billing and global roaming. Users will have exactly one service subscription with one service provider in order to benefit connection anytime and anywhere, known as always best connected [1].

The integration of these existing systems seems unavoidable due to the potential benefits of their complementarity and will be the basis of NGWN design rather than invest efforts into developing new radio interfaces and technologies [2]. An integrated and interworking architecture for NGWN should handle specific requirements and satisfies the following main features [3]: being economical and scalable, provisioning seamless mobility and security. Conceptually, a typical NGWN framework can be viewed as many overlapping wireless access domains, as shown in Fig. 1. Heterogeneity in terms of RATs and network protocols in NGWN requires a common interconnection element. Since the Internet Protocol (IP) technology enables the
support of applications in a cost-effective and scalable way, it is expected to become the core backbone network of NGWN [3]. Thus, current trends in communication networks evolution are directed towards an all-IP principles in order to hide heterogeneity and to achieve convergence of various networks. For example, third generation wireless initiatives, 3GPP and 3GPP2, adopted IPv6 as the sole IP version for the IP-based Multimedia Subsystem (IMS) [4].

Mobility management, with provisioning of seamless handoff and QoS guarantees to mobile nodes (MNs), is one of the key topics in 4G/NGWN. It is crucial to provide seamless mobility and services continuity (i.e., minimal service disruption during roaming) support based on intelligent and efficient techniques. This means that seamless handoff schemes should have following features: minimum handoff latency, lower packet loss, limited handoff failure or blocking and lower signaling overhead. The handoff latency refers to the time interval during which an MN cannot send or receive any data traffic during handoffs. It is composed of L2 (link switching) and L3 (IP layer) handoff latencies. The overall handoff latency may be sufficiently long and lead to packet loss, which is inappropriate for real-time applications such as voice over IP (VoIP). The signaling traffic overhead is defined as the total number of control packets (for registration, binding update and binding refresh procedures) exchanged between an MN and a mobility agent (e.g., home agent).

The handoff process in NGWN is composed of three phases: network discovery, handoff decision and handoff execution. The most simple way for a multiple interfaces MN to discover reachable wireless networks is to keep all air-interfaces on at all times. However, keeping an air-interface active all the time consumes battery power and bandwidth even when the device unit is not sending or receiving any packets. The handoff decision refers to the process of selecting the right moment when to perform the handoff. It is thus critical to avoid keeping idle air-interfaces perpetually on. Moreover, in order to avoid the ping-pong effect, an MN must observe if the new network is consistently better than the current one before performing a handoff.

In homogeneous networks, the handoff decision is typically driven by metrics which are strictly related to the received signal strength (RSS) level and resources availability. However, in NGWN, the RSS from different networks do not have the same meaning since each network is composed of its specific characteristics and there is no common pilot signal. Then, RSS comparisons are insufficient for handoff decision and may be inefficient or impractical. A more complex decision criterion that combines a large number of parameters or factors such as monetary cost, bandwidth, power consumption and user profile is necessary.

This paper proposes a novel mobility management scheme, called Handoff Protocol for Integrated Networks (HPIN), that enables QoS guarantee for real-time applications in heterogeneous IPv6-based wireless environments. HPIN is a one-suite protocol that performs network selection based on our proposed handoff score function approach. Moreover, HPIN performs fast handoff, localized mobility management, context transfer and access network discovery. The aim of HPIN is to allow seamless roaming and services continuity across various access networks. The remainder of this paper is organized as follows. The following section offers an overview of basic concepts and work related to interworking and mobility management in heterogeneous IP wireless networks. After that, an interworking architecture, called Integrated InterSystem Architecture (IISA) is presented along with HPIN. Subsequently, an analytical framework to derive a signaling traffic cost, handoff latency and total packet loss is described. The performance analysis based on this analytical framework is carried out before concluding remarks are drawn in the last section.

2. Background and related work

Mobility management enables systems to locate roaming terminals in order to deliver data packets (i.e., location management) and maintain connections with them when moving into new subnet (i.e., handoff management). With the coexistence of various wireless access technologies, two kinds of handoffs are possible in NGWN: horizontal and vertical handoffs. Horizontal or intrasystem handoff occurs when an MN moves between the access points (APs) or base stations (BSs) of a same network technology. When AP/BSs belong to different networks (e.g., IEEE 802.11 and UMTS), such movement refers to vertical or intersystem handoff. NGWN characteristics make the implementation of vertical handoff more challenging than horizontal handoff. In fact, maintaining an uninterrupted session while the physical interface changes is very complex.

2.1. Mobility management protocols

An evident way to achieve roaming among networks of different domains or service providers consists of using Service Level Agreements (SLAs). However, due to several reasons, this approach is not always feasible. In fact, the increasing number of wireless networks make it impractical
for network operators to have direct SLAs with every single operator. Moreover, the SLAs can only provide static information. Furthermore, network operators are reticent to the idea of opening their databases to others. With characteristics of mobility in 4G/NGWN, the user profile seems to be important when performing handoff decision. More complex metrics combining a large number of parameters such as monetary cost or price, bandwidth, power consumption, service types, network conditions and user preferences should be defined for handoff decision in NGWN [7]. Designing handoff decision function to evaluate these various metrics simultaneously is crucial in NGWN and remains a challenging research issue.

Various schemes for horizontal handoff have been proposed in the literature [6]. Recently, research on vertical handoff in 4G/NGWN attracted more attention and some works have been presented in the literature with their strength and weakness [9]. Several of these related papers use an handoff decision based on RSS and bandwidth. Moreover, other proposals focus on the design of an architecture for heterogeneous networks such as the IPv6-based mobility schemes proposed by the Internet Engineering Task Force (IETF). Mobile IPv6 (MIPv6) [11] was proposed for mobility management at the IP layer and allows MNs to remain reachable in spite of their movement within IP-based mobile environments. Each MN is always identified by its home address, regardless of its current point of attachment to the network. While away from its home network, an MN is also associated with a care-of address (CoA), which provides information about its current location. However, MIPv6 has some well-known drawbacks such as signaling traffic overhead, high packet loss and handoff latency, thereby causing a user-perceptible deterioration of real-time traffic [12,13]. Simultaneous mobility is another problem MIPv6 faces due to route optimization, which can occur when two communicating MNs have ongoing session while both are moving simultaneously [14].

These weaknesses led to the investigation of other solutions to enhance MIPv6. Two main MIPv6 extensions proposed by the IETF are the Hierarchical MIPv6 (HMIPv6) [15] and the Fast Handovers for MIPv6 (FMIPv6) [16]. These protocols tackle intra-domain or micro-mobility while MIPv6 is used for inter-domain or macro-mobility. HMIPv6 handles handoff locally through a special node called Mobility Anchor Point (MAP). The MAP acts as a local home agent (HA) in the network visited by an MN, limits the amount of MIPv6 signaling outside its domain and reduces the location update delay. However, HMIPv6 cannot meet the requirements of delay-sensitive traffic such as voice over IP (VoIP), due to packet loss and handoff latency [12,13]. FMIPv6 was proposed to reduce handoff latency and minimize services disruption during handoff pertaining to MIPv6 operations such as movement detection, binding update and addresses configuration. In other words, FMIPv6 allows an MN to receive data before the binding is done at the HA and correspondent nodes (CNs). The link layer information (L2 trigger) is used either to predict or respond rapidly to handoff events.

Although FMIPv6 paves the way for improving MIPv6 performance in terms of handoff latency, it does not efficiently reduce signaling overhead (due to new messages introduced and exchanged for handoff anticipation) nor does it prevent packet loss (due to buffer space requirement). This may lead to unacceptable service disruption for real-time applications. Combining HMIPv6 and FMIPv6 motivates the design of Fast Handover for HMIPv6 (F-HMIPv6) [17] to increase network bandwidth usage efficiency. However, F-HMIPv6 may inherit drawbacks from both FMIPv6 and HMIPv6, for example synchronization and signaling overhead issues. In fact, in F-HMIPv6, when an MN performs a handoff immediately after sending a fast binding update (FBU) message to the MAP, all packets transferred to the previous on-link care-of address (PLCoA) during the period that the FBU needs to reach the MAP, are lost [12]. Moreover, F-HMIPv6 provides fast handoff and localized mobility management although, it does not provide context transfer, access router and network discovery in the same way as for FMIPv6 and HMIPv6.

2.2. Policy-based and integrated architectures

An architecture for next generation all-IP-based wireless systems is proposed in [3], called Architecture for Ubiquitous Mobile Communications (AMC). Two new entities, the Network Interworking Agent (NIA) and the Interworking Gateway (IG), are introduced in order to allow the integration of several wireless networks while supporting MN roaming. Moreover, an intersystem handoff protocol at the IP layer is designed for mobility management in this new architecture. However, the AMC architecture provides no appropriate handoff decision mechanism to take heterogeneity into account. The deployment of IG entity in all networks may require excessive economical costs and require changes in individual networks. Furthermore, the AMC architecture is based only on SLAs which can provide only static information. On the other hand, AMC may inherit certain drawbacks of loose coupling. The handoff decision is based on RSS criterion, which is inappropriate for NGWN as stated above. Also, air-interfaces always on approach is used in AMC architecture, QoS provisioning and guarantees are not taken into account in AMC.

Other works have been presented in [18–20] for intersystem mobility management and interworking of heterogeneous 3G cellular wireless networks, yet not for IP-based heterogeneous wireless networks. Often, proposed integration schemes are based on the deployment of a gateway, which solves interworking issues between each pair of networks. Adding a gateway at the boundaries of both systems would increase deployment costs. Moreover, these studies seem to integrate only 3G cellular networks. To reduce energy consumption of MNs without degrading...
throughput, an approach called WISE (Wise Interface SElection) [21] for 3G/WLAN vertical handoff has been proposed. With WISE, the handoff decision is performed according to the network load and the energy consumption of the air-interfaces. However, requirements such as services and applications security are not considered. In [8], an integrated architecture and interface selection schemes are proposed based on signal strength and radio interfaces priorities. As aforementioned, these parameters are not appropriate for handoff decision in NGWN. Moreover, an MN must passively evaluate handoff conditions, even when the application is running well in the current network. This introduces unnecessary power consumption and usage of network resources. In [23,22], decision-making procedures are proposed for roaming management between heterogeneous access networks technologies based on finite-state transducers.

The IETF proposed a policy-based architecture in order to implement a set of rules to manage and control access to network resources which is particularly useful for QoS management [24]. Two main logical entities for policy control-based architecture are the Policy Decision Point (PDP) and the Policy Enforcement Point (PEP). To enable judicious choice for vertical handoff, several papers have proposed a utility or cost function to measure the network quality. A policy-enabled handoff decision algorithm proposed in [25] is based on a cost function approach that considers several factors (e.g., bandwidth, power consumption and monetary cost). This cost function is very simple and cannot handle more sophisticated scenarios. Moreover, the cost function evaluation could require high processing time and power. A vertical handoff decision algorithm has been proposed in [7,26] and metrics that characterize NGWN have been identified. However, the proposed cost function could lead to singularity problems if connections are free of charge. Furthermore, handoff instability problem and mobility management at the IP layer are ignored. The factors considered in the above cited papers are insufficient. In fact, information about authentication types, access network types and the support of roaming partners are not taken into account. Moreover, these studies do not provide a viable architecture framework for selection mechanisms, nor business models for prospective deployment.

2.3. Media independent handover (MIH)

The IEEE 802.21 or Media Independent Handover (MIH) [10] standard has been proposed in order to provide interoperability, generic link layer intelligence and to enable efficient handoff management between heterogeneous access networks, that includes both 802 and non-802 based networks (e.g., 3GPP and 3GPP2). A new logical network entity called the MIH Function (MIHF) is defined between link layer (L2) and network layer (L3) which provides three main services: Media Independent Command Service (MICS), Media Independent Information Service (MIIS) and Media Independent Event Service (MIES). The MICS provides a set of commands that enables MIH users to issue commands for handoff control and mobility. In fact, the MICS commands are used to determine the status of the connected links and also to execute connectivity and mobility decisions of the higher layers to the lower ones. While the MIES provides event classification, event reporting and event filtering corresponding to dynamic changes in link characteristics, status and quality. It reports both local and remote events to the upper layers.

Some of the events that have been specified by MIH are Link Down, Link Up, Link Going Down, Link Parameters Report, and Link Detect. Finally, the MIIS defines a utility or cost function to discover available neighboring network and for obtaining the necessary information to make effective handoff decisions. Note that, the MIIS defines a set of information elements (IEs) to provide access to dynamic and/or static information and higher layer services supported by the networks. Such information include details on the services and characteristics of the serving and neighboring networks, roaming agreements, available networks, access costs, QoS parameters (e.g., delay, jitter, data rate), security capabilities. The MN can access the relevant information via its current active network interface and the other interfaces do not need to be turned on simultaneously.

3. Interworking architecture for NGWN

As stated in [5], no use cases have been identified for the access to 3G wireless system circuit-switched based services scenario. Thus, for further development, it is not considered worthwhile. Hence, we focus on two main scenarios: service continuity and seamless services provisioning. Based on 3GPP/3GPP2-WLAN interworking models, an interworking architecture, called Integrated InterSystem Architecture (IIASA) is proposed in [27] and it is shown in Fig. 2. For the sake of simplicity, only UMTS, CDMA2000 and WLAN networks are illustrated. Although IISA is designed to integrate any number of radio access technologies (RATs) and mobile devices may be equipped with any number of interfaces. Instead of developing new infrastructures, IISA extends existing infrastructures to tackle integration and interworking issues and provide mobile users with ubiquity or always best connected [1].

The serving GPRS (general packet radio service) support node (SGSN) and packet control function (PCF) are enhanced with the access router (AR) functionalities and called Access Edge Node (AEN). Similarly, the gateway GPRS support node (GGSN) and packet data serving node (PDSN) are extended with MAP or HA functionalities (to enable message format conversion, QoS requirement mapping, etc.) and is called Border Edge Node (BEN). The WLAN Interworking Gateway (WIG) acts as a route policy element, ensuring message format conversion. Extended functionalities can be integrated into the existing networks’ entities or implemented separately. We advocate the first
choice as it is more easily managed and implemented. Mapping between the home location register or the home subscriber server (HLR/HSS) in 3G wireless networks and AAA server in WLAN is required to execute authentication and billing when user roams across both technologies.

A novel entity, Interworking Decision Engine (IDE), illustrated in Fig. 3 is introduced to enable the interworking and handoff between various networks by reducing signaling traffic, services disruption during handoff and it also handles authentication, authorization and accounting (AAA) as well as mobility management. The usage of the IDE could be seen as a value-added service that network operators offer to their subscribers to allow roaming to other networks. To avoid additional signaling overhead due to the execution of AAA procedure every time an MN performs a handoff and requests registration, a token-based approach is proposed. This token is obtained from the IDE after the MN first successful registration in the foreign network. The token includes security association parameters to setup a secure tunnel between an MN and AR/AENs. IISA allows the separation of control and transport plane. In fact, data packet traffic bypasses the IDE. In other words, the IDE is in a control plane while the MAP/BEN handles the actual data traffic, thus it is in the transport plane.

The Authentication Module (AuM) is used to authenticate users moving across different wireless networks and avoids the required direct security agreements or association between foreign networks and home network. The AuM stores information such as the subscribers’ identities, users’ preferences/profiles and terminal mobility patterns. The Accounting Module (AcM) enables billing between different wireless networks and stores billing information associated with the resource usage. It acts as common billing/charging system between various network operators.

Usually, different administrative domains have different QoS policies for resources allocation. Thus, when an MN moves from one administrative domain to another, QoS re-negotiation may be required. Such re-negotiation will be based on service level agreements (SLAs) between both domains. The Resource Management Module (RmM) enables QoS mapping and re-negotiation. Furthermore, the RmM allows fast transfer of user profiles and QoS parameters between two administrative domains during handoff. The SLRA Module stores information about service providers or network operators which have SLAs and roaming agreements (RAs) with the IDE manager. The Handover Decision Module (HdM) is used when intersystem or inter-domain handoff should be granted or not. In other words, it provides support for roaming and handoffs. For further details about the IISA architecture, please refer to [27].
4. Proposed handoff protocol

With the coexistence of various access networks, the selection of subsystem that allows better service provisioning to subscriber is crucial and depends on several factors, for example defined in user’s profile and preferences. Then, a more complex handoff decision criterion combining a large number of parameters or factors such as monetary cost, bandwidth, priority, power consumption, service types, system performance, user preferences, MN moving speed, security, resource availability, network accessibility and MN conditions, must be defined in NGWN [7]. The design of handoff decision function which evaluates these various factors simultaneously is crucial in NGWN and remains a research challenge. A hybrid vertical handoff decision function in heterogeneous wireless networks is proposed in this work to provide satisfactory overall performance, based on the aforementioned criteria.

4.1. Handoff score function

In NGWN, the selection of the best access network is important during connection or handoff request. Handoff triggering is performed either in the access network by an MN, AP/AR, while the whole handoff process may require several entities located either in a home or access network and foreign network, particularly for vertical handoffs. In fact, handoff triggered by an MN and/or AP/AR could be conducted only with locally available information such as link quality, signal strength, AR’s capabilities, and subnet load and can lead to inefficient system performance. Other information recorded in the home and foreign networks such as operator policies, access network’s load and user preferences, can be relevant for network selection and handoff decision. This paper proposes a novel handoff decision function for this purpose.

The usage of network \( n \) at a certain time is associated to a score which is a function of several of the aforementioned parameters. Those parameters can co-relate, interact and may have conflicting objectives. The presence of such conflicts makes it difficult to find an effective solution that optimizes all criteria simultaneously. In this case, network selection issue can be formulated as a multicriteria optimization problem. Several approaches are available in the literature in order to solve multicriteria optimization problem. Amongst them, we use a weighted sum approach introduced in [28] and combine different conflicting criteria into a single criterion.

For a given user \( u \), the score function \( f_{n}^{u} \) is evaluated for each network \( n \) that can provide user services. In other words, the score function quantifies the QoS provided by a wireless network to handle running application on an MN device. The target network that results in the least highest computed score function value among all candidates is the network that would provide significant benefits (i.e., QoS level) to the user. More specifically, let \( n_{c} \) be the current serving network, \( N \) denotes the set of neighbor networks of \( n_{c} \), and \( F_{u} \) represents a set defined by:

\[
F_{u} = \{ f_{n}^{u} : f_{n}^{u} > f_{n}^{u}, \forall n \in N \}
\]

The optimal target network, \( n^{*} \), for a mobile user \( u \) is obtained as follows:

\[
f_{n}^{u} = \inf (F_{u})
\]

where \( f_{n}^{u} \) is expressed by:

\[
f_{n}^{u} = \sum_{s} p_{n,s}^{u} f_{n,s}^{u} \quad \forall n, u.
\]

The score function of unreachable network always equals zero. \( p_{n,s}^{u} \) is the priority of service or session \( s \) for network \( n \) based on user \( u \) profile, i.e., the probability that an MN prefers network \( n \) for a connection of service \( s \) and \( f_{n,s}^{u} \) is the per-service score function for network \( n \). In other words, it represents a QoS factor and is computed as follows:

\[
f_{n,s}^{u} = \sum_{i} w_{i,s}^{u} f_{n,i}^{u,s} \quad \forall s, n, u
\]

where \( f_{n,i}^{u,s} \) is the normalized QoS function or factor provided by network \( n \) for parameter \( i \) to carry out service \( s \) and \( w_{i,s}^{u} \) stands for a weight indicating the impact of the QoS parameters on either user or network and sum to one, i.e., \( \sum_{i} w_{i,s}^{u} = 1 \) and \( w_{i,s}^{u} \in [0,1] \). The assignment of weights \( w_{i,s}^{u} \) plays a key role in the network selection. Hence, it is assumed that the assignment of these weights is based on the subscribers’ home network policy or users profile. The target network is the network which provides just enough consistently higher QoS level than current network. Due to dynamic network conditions of wireless environments, the score function of target and current wireless networks may vary considerably. Then, a dwell timer or stability period should be adjusted according to the current measurements of the handoff score function.

In order to reflect the inability of candidate networks to guarantee the desired QoS requirements and to speed up score function evaluation, several constraints are considered depending on each factor such as the MN speed, bandwidth threshold, ARs load and delay. Hence, it is necessary to define maximal and minimal requirements for each parameter to enable the application provisioning. Then, if an available network cannot guarantee a minimum requested QoS (e.g., delay for real-time applications or bandwidth), it should be immediately discarded as a candidate network when there are several networks available. Otherwise, the network which allows best effort as QoS level is selected. The processing time and power are then reduced during the computation of the score function.

The normalized QoS function \( f_{n,i}^{u,s} \) is given by:

\[
f_{n,i}^{u,s} = \begin{cases} 0 & \text{if } Q^{u,s} \leq L_{n,i}^{s} \\ \left( \frac{Q^{u,s} - L_{n,i}^{s}}{U_{n,i}^{s} - L_{n,i}^{s}} \right)^{y_{i}} & \text{if } L_{n,i}^{s} < Q^{u,s} < U_{n,i}^{s} \\ 1 & \text{if } Q^{u,s} \geq U_{n,i}^{s} \end{cases}
\]

where \( Q^{u,s} \) is the real value of parameter \( i \) in wireless network \( n \) associated to application \( s \), measured by the MN or announced by a mobility agent. \( L_{n,i}^{s} \) and \( U_{n,i}^{s} \) respectively
express the minimal and maximal requirement of parameter \( i \) associated with wireless networks \( n \) for application \( s \). These boundaries make it possible to check if the serving network satisfies the application’s requirements. \( x_i \) is a constant that can take different values in order to specify different normalized QoS functions for each parameter \( i \). The values of \( x_i \) greater than 1 result in a slow increase from the unacceptable required boundaries and fast near the maximal required boundary. If \( x_i \) equals 1, the normalized QoS function is strictly proportional between the required boundaries. Values of \( x_i \) lower than 1 result in a fast increase from the unacceptable required boundaries and slow near the maximal required boundary. Normalization is needed to ensure that the sum of the values, measured with different units, is meaningful.

### 4.2. Handoff decision algorithm

The proposed score function for handoff decision may be computed at MN side or at the IDE. In fact, assuming that mobile devices will become increasingly powerful, intelligent and sensitive to link layer changes, we adopt a network-assisted and mobile-controlled handoff strategy. The proposed handoff scheme combines mobile-monitored and network-probed information to provide reliable handoff management. Prior to handoff, an MN can obtain information from wireless network candidates to which it is likely to handoff, and use such information to optimize handoff performance. On the other hand, if mobile device capabilities are limited, the handoff score function is computed at the IDE. In this case, handoff strategy turns to mobile-assisted and network-controlled. According to MIH standard, an MN which implements our proposed handoff decision function is able to obtain periodically the information about available networks in its moving coverage area by using its current network interface. In other words, the information provided by the MHS is used to estimate the parameters of the QoS capabilities and requirements.

The sequential execution of the system discovery and handoff decision steps for vertical handoff may be inappropriate. In fact, if the ongoing session of an MN runs with satisfactory QoS level with the current network, there is no need to discover another better network. Performing unnecessary handoff operations will waste network resources and energy of MN battery. We propose a handoff scheme based on a cross system discovery and handoff decision steps. In this proposed scheme, the current network conditions are checked first after an impending handoff event is generated if they can satisfy the ongoing session requirements. If they do, there is no need to perform the system discovery process, which will be launched only if the network conditions cannot satisfy an MN session requirements. Otherwise, if possible an MN tries to perform horizontal handoff. However, if no other AR/AEN of current network or technology exists, a vertical handoff is needed for this MN and it tries to find another more suitable network.

To avoid all air-interfaces always on approach for system discovery, we propose an adaptive scheme. An MN requests neighbors networks information from its serving network to the IDE. Through information reported periodically to the IDE, it maintains a global view of the connection state of roaming MNs and access networks conditions in its coverage area. The IDE replies by sending information about neighbor networks, if any, to the MN through its point of attachment. Then, the MN will compute the handoff score function for each reachable network using the information received in order to determine candidate networks. If candidate networks are available, the MN sets up a waiting timer to assess the stability of these candidate networks. If the QoS level remains better until the waiting timer expires, the MN selects the candidate network which offers a QoS level that is slightly better than the current serving network and can start the handoff execution step. Otherwise, it will remain in the current network as the target network is unstable and cannot maintain better QoS level during the waiting time.

The choice of candidate network that offers a slightly improved QoS level overcomes inefficient usage of network resources. The handoff decision algorithm flow chart is illustrated in Fig. 4, where RSS\(_T\) is a predefined or adaptive received signal strength threshold value. Although it is possible to select two different networks to access two ongoing services (in case of multiservice/session) simultaneously, in practice such a choice can create several problems. For example, authentication with two networks simultaneously, turning on two radio-interfaces at the same time and routing of service data appropriately within the device and the network represent some of these challenges. It is thus recommended to avoid such network selection approach.

### 4.3. Operation mode of HPIN

Usually, handoff schemes assume that an MN monitors periodically neighbor AR/AENs signal strength by keeping all of its interfaces always on. However, keeping on all interfaces continuously drains the MN battery energy. This problem becomes worst when the number of RATs supported by the mobile devices increases and are available in an MN’s moving area. To achieve seamless mobility across various access technologies and networks, an MN needs the information about the wireless network to which it could attach. It is also necessary to transfer information (context transfer) related to the MN from the current AR/AEN to the next one. The proposed Handoff Protocol for Integrated Networks (HPIN) implements access router and network discovery based on message exchanged between the IDE and mobility agents, minimizes the usage of limited wireless resources, provides fast mobility and secure transfer. The key objective of HPIN is to reduce ser-
vices disruption and to avoid the re-initiation of signaling to and from an MN during handoff from the beginning.

4.3.1. Overview of HPIN

The Router Information eXchange (RIX Request/Reply) messages are used to allow the MAP/BEN and the IDE to maintain a global view of their coverage area. The RIX messages exchange is quite similar to the routing information protocol (RIP) [29] which allow neighboring routers to exchange their routing tables with one another. The information about the global view may be defined in terms of system parameters, subnet load, QoS status and information (e.g., supported data rate, video coding rate, maximum delay), bandwidth availability, and MN’s signal strength.

Four main messages are introduced for handoff management:

- **Handoff Preparation Request** (HPReq) message sent from an MN to the MAP/BEN for a handoff request. It contains information about user preferences/profiles, applications required QoS capabilities, L2 information of AR/AENs, IP address of MNs, signal strength of the MN, AR/AEN’s ID for an MN location tracking.

- **Handoff Preparation Reply** (HPRep) message sent by the MAP/BEN to an MN. It contains network prefixes, the list of candidate AR/AENs (CAR/AENs), their capabilities and the QoS status.

- **Handoff Preparation Notification** (HPN) message sent by an MN to the MAP/BEN to notify the possibility of an impending handoff. It contains the information about the selected new AR/AEN where an MN will handoff. The HPN includes the request to verify the pre-configured new on-link care-of address (NLCoA) and establish a bi-directional tunnel between the MAP/BEN and the NAR/AEN in order to prevent routing failures during handoff.

- **New Link Attachment (NLA)** message sent by an MN to the NAR/AEN to announce its presence on the new link and confirm usage of the NLCoA.

Moreover, there are also two optional messages:

- **Handoff Preparation Acknowledgment** (HPAck) message which contains information about the current capabilities that an AR/AEN can support.

- **New Link Attachment Acknowledgment** (NLAck) message to notify an MN to use another NLCoA rather than its prospective NLCoA, in case of address collision. This message is also sent to the current MAP/BEN to allow it to bind previous on-link care-of address (PLCoA) and to validate the NLCoA.

The MN decides whether to send a request message (HPReq) for handoff preparation depends on the generation of anticipated triggers (AT). The high bit error rate, link going down, weak signal strength, security risks, monetary cost and geographical location can be used as anticipated triggers. To allow seamless service continuity, requirements specified in the request message need to be set consistently with the QoS negotiated in the previous subnet. QoS consistency remains a very challenging issue and is crucial for real-time applications. This consistency is handled by the IDE, which allow QoS mapping between various networks. Mapping is needed to translate the QoS guarantees and specifications provided for a session across

---

**Fig. 4. Flow chart of handoff decision algorithm.**
heterogeneous networks. The QoS mapping performed by the IDE is for example the resources about reservation and delay threshold according to SLAs.

4.3.2. Roaming scenarios

Upon receipt of HPReq message, the MAP/BEN checks its local CAR/AEN table to retrieve information about their capabilities. The MAP/BEN performs a pre-filtering process, based on the requirements specified in the HPReq message and the available network conditions to obtain the CAR/AEN list. Note that, if the context information of this MN is not available at the candidate MAP/BEN, the latter sends Context Transfer Request (CTReq) message [30] to the IDE in order to get session management parameters of the MN for establishment of traffic bearers on the new path. In response to a CTReq message, the IDE transmits a Context Transfer Data (CTD) message that includes the MN’s feature contexts. When the new MAP/BEN receives a CTD message, it installs the contexts as received from the IDE.

The MAP/BEN responds to an MN by sending a HPRep message. If the MAP/BEN lacks information about the user profile, it requests this information to the IDE rather than to the MN’s home network, which may be away of the current MAP/BEN domain. When the MN receives a HPRep message, it performs stateless address configuration [11] to get new on-link CoAs (NLCoA) and knows L2 technologies provided by CAR/AENs to which it is likely to handoff. A primary NLCoA will be associated to the selected network. The MN will activate only the air-interface associated to the CAR/AEN list, rather than setting all air-interfaces always on. This selective air-interface activation enables a better tradeoff between system discovery time, power consumption efficiency and allows shorter scanning delay.

Whenever an MN receives a L2 trigger, it initiates a target AR/AEN selection among CAR/AENs. This selection is based on the handoff score function, \( f_n \), proposed above. After the target AR/AEN selection process, an MN notifies the MAP/BEN that it is moving into a new subnet by sending a HPN message to allow the MAP/BEN to establish a binding between PLCoA and NLCoA and to buffer all incoming and in-flight packets (this avoid synchronization issue identified in basic IP mobility schemes) having PLCoA as destination address. The MAP/BEN forwards the HPN message to the new access router (NAR/AEN) and includes the CTD message. Upon reception of acknowledgment from the NAR/AEN, the MAP/BEN starts tunneling any packets (buffered and incoming) addressed to PLCoA towards NLCoA. When the NAR/AEN receives the HPN, it performs the duplicate address detection (DAD) procedure on the NLCoA and application requirements validation. Then, the NAR/AEN responds to the HPN with a HPAck message. When the CAR/AEN processes the CTD message, it can optionally generate a CTD Reply (CTDR) message to report on the status of processing the received contexts and piggybacks this message in HPAck message.

After transmitting a HPN message, the MN performs a link layer switching and announces its presence on the new link by sending the NLA message to the NAR/AEN. Then, the NAR/AEN will start delivering the buffered packets to the MN. The packets forwarding procedure remains active until the binding update (BU) procedure is completed. Note that, if the HPN is not sent before the L2 handoff, the MN sends a HPN piggybacked in the NLA message (NLA[HPN]) over the new link. This situation corresponds to the reactive mode of HPIN in contrary to its predictive mode, i.e., the HPN is sent through the previous link. When the NAR/AEN receives the NLA[HPN] message, it processes the NLA message part, extracts the HPN message part and forwards it to the serving MAP/BEN. At this time, the serving MAP/BEN starts buffering all incoming and in-flight packets having PLCoA as destination address and forwards them toward the NLCoA. If an address collision occurs when the NAR/AEN processes the NLA message, it changes the prospective NLCoA to a valid NLCoA and includes it in the HPN message before forwarding it to the MAP/BEN and simultaneously the NAR/AEN sends a NLAck to the MN. Fig. 5 illustrates the messages sequence during intra-BEN roaming while Fig. 6 represents the messages sequence flow during inter-BEN roaming.

The binding update (BU) procedure is performed by the NAR/AEN on the behalf of MNs. In fact, the AR/AEN acts as a proxy and copies the BU list of an MN in its cache and manages this list (e.g., lifetime entries) in the same way as the original is managed by the MN. The copy in the AR/AEN cache must be updated periodically according to the original BU list of the MN. As soon as an MN attached to the NAR/AEN, the copy of the BU list is used by the NAR/AEN to inform the MAP/BEN about the NLCoA. When the lifetime of the BU list cached in the AR/AEN is about to expire, the AR/AEN can send a request for a BU list renewal to an MN. The BU list renewal is per-
The notation used in this paper is given in Table 1.

Hence, analyzing of these metrics are most useful to evaluate the performance of mobility management protocols.

5. Analytical model for HPIN

In IP-based wireless networks, the QoS may be defined by packet loss, handoff latency and signaling overhead. Hence, analyzing of these metrics are most useful to evaluate the performance of mobility management protocols. The notation used in this paper is given in Table 1.

5.1. User mobility and traffic models

User mobility and traffic models are crucial for efficient system design and performance evaluation. Usually, an MN mobility is modeled by the cell residence time and various types of random variable are used for this purpose [31]. Two commonly used mobility models in wireless networks are random-walk and fluid-flow models [32]. We consider the random walk, i.e., an MN moves at constant speed $v$ with uniformly distributed angular directions belonging to $[0, 2\pi]$ as mobility model. Let $d_{s, u}$ be the distance between AP/BS $i$ and mobile user $u$. We assume that the path loss or link gain is given by $h_{s, u} = 10^{-m d_{s, u}^\alpha}$, where $\alpha$ is the path loss exponent, $s_{r, u}$ is the log-normal shadowing with zero mean and standard deviation $\sigma$.

The exponential distribution provides an acceptable tradeoff between complexity and accuracy. Thus, most cost analyses adopt exponential assumption [31]. We consider a traffic model with two levels, session and packets. The session duration follows an exponential distribution with inter-session rate $\lambda_s$ while packet generation and arrival rate follow a Poisson process. Let $\mu_c$ and $\mu_d$ be the border crossing rate for an MN out of a subnet (i.e., AR/AEN domain) and a MAP/BEN domain, respectively. When an MN crosses a MAP/BEN domain border, it also crosses an AR/AEN border. Then, let $\mu_b$ be the border crossing rate for which an MN still stays in the same MAP/BEN domain, $\mu_b = \mu_c - \mu_d$.

If we assume that all subnets have circular shape and form together a contiguous area and that each MAP/BEN domain is composed of $M$ equally subnets, the domain border crossing rate is $\mu_d = \mu_c / M$. The roaming probability depends on an MN’s movement pattern in its original network but not in its destination network. Hence, the probabilities that there are at least one local binding update ($P_c$) and one global binding update ($P_d$) between two consecutive sessions of an MN are:

$$P_c = \Pr(t_s > t_c) = \frac{\mu_c}{\mu_c + \lambda_s}$$

$$P_d = \Pr(t_s > t_d) = \frac{\mu_d}{\mu_d + \lambda_s}$$

The average number of location binding updates during an inter-session time corresponding to subnet crossings, $E(N_c)$, and MAP/BEN domain crossings, $E(N_d)$, are given by:
With the same time variables assumption, we can obtain the expression of $E(N_i)$, i.e., the average number of subnets (AR/AENs) that an MN crosses and still stay within a given MAP/BEN domain during an inter-session time interval, as follows: $E(N_i) = \mu_i / \lambda_i$.

5.2. Binding update signaling cost

Performance analysis of wireless networks must consider the total signaling cost induced by a mobility management scheme. In NGWN, there are two kinds of location or binding update signaling. One occurs during an MN’s subnet crossing while the other occurs when the binding is about to expire. Depending on the type of movement, two kinds of binding update can be performed: global and local. Global binding update occurs when an MN moves out of its MAP/BEN domain. In this case, the MN registers its new regional CoA (RCoA) to the HA and the CNs. On the other hand, if the MN changes its current address (LCoA) within a MAP/BEN domain, it only needs to register this new LCoA to the MAP/BEN. Hence, the average binding update signaling cost during inter-session time interval heavily depends on the computation of numbers of binding updates and is given by:

$$C_{BU} = E(N_i)C^l + E(N_d)C^s = \frac{1}{\text{SMR} \sqrt{M}} \left[ C^s + (\sqrt{M} - 1)C^l \right]$$

where SMR is the session-to-mobility ratio and represents the relative ratio of session arrival rate over user mobility rate: $\text{SMR} = \lambda_i / \mu_i$.

Anticipated trigger and link layer information (L2 trigger) are used either to predict or rapidly respond to handoff events. Hence, the HPIN signaling cost depends on the probability that the handoff anticipation is accurate. If there is no real handoff after the L2 trigger, all messages exchanged for handoff anticipation can be unnecessary. Thus, global and local binding update signaling cost for HPIN are expressed as follows:

$$C^s = \mu_s S^s + (1 - \mu_s)(S^s + S^g) + C_{ru}$$

$$C^l = \mu_s S^s + (1 - \mu_s)(S^l + S^g) + C_{mu}$$

where $C_{ru}$ and $C_{mu}$ represent the binding update cost at HA/CNs and MAP/BEN, respectively. Their expressions are given in Table 2.

5.3. Handoff latency and packet loss

Since the number of packets lost is proportional to handoff latency, only the expression for handoff latency is derived in this section. The following parameters are defined to compute handoff latency and packet loss: $t_{L2}$ the L2 handoff latency and $t_{X,Y}$ one-way transmission delay between nodes X and Y for a message of size $s$. If one of the endpoints is an MN, $t_{X,Y}$ is computed as follows:

$$t_{X,Y}(s) = \frac{1 - q}{1 + q} \left( \frac{s}{B_w} + L_{ul} \right) + (d_{X,Y} - 1) \left( \frac{s}{B_w} + L_u + \sigma_q \right)$$

where $q$ is the probability of wireless link failure and $\sigma_q$ the average queuing delay for each router on the Internet [33], $B_w$ (resp. $B_u$) represents the wireless (resp. wired) link bandwidth and $L_{ul}$ (resp. $L_u$) denotes the wireless (resp. wired) link delay.

The HPIN handoff latency depends on the information available as well as the link where fast handoff messages are exchanged. The average handoff latency of HPIN for intra-MAP/BEN roaming is then given as follows:

$$D_{HPIN} = P_{CMN} O_{HPIN} + (1 - P_{CMN})N_{HPIN}^l$$

where $O_{HPIN} = t_{L2} + 2t_{MN,AEN}$ is the handoff latency if the information about the NAR/AEN and impending handoff are available before the L2 handoff. Otherwise, this handoff latency is given by $N_{HPIN}^l = t_{L2} + 2t_{MN,AEN} + 2t_{AEN,BEN}$ associated to the HPIN reactive mode. For inter-MAP/BEN, $N_{HPIN}^l$ becomes $N_{HPIN}^l = t_{L2} + 2t_{MN,AEN} + 2t_{AEN,BEN} + t_{BEN,BEN}$ while $O_{HPIN}^l = O_{HPIN}$. In fact, handoff procedure only depends on intra-MAP/BEN communication delay, since the inter-MAP/BEN signaling is completed before the L2 handoff. The average handoff latency of HPIN for inter-MAP/BEN roaming is computed similarly as in (10).

6. Performance evaluation

The performance analysis is conducted by examining several metrics such as throughput, handoff latency, packet loss and signaling traffic overhead. The parameter and default values used in the performance evaluation are listed in Table 3. An analytical framework to evaluate performance of IPv6-based handoff schemes proposed by the IETF (i.e., MIPv6, HMIPv6, FMIPv6 and F-HMIPv6) is presented in [34]. Such evaluation method is used to compare the performance of the IETF’s protocols with HPIN. Traditional handoff protocols based on received signal strength (RSS) are compared with a handoff score function-based approach used in HPIN. For the sake of sim-
Compliance, four parameters are used for network selection: power consumption ($p$), bandwidth ($b$), latency ($l$) and usage cost ($c$). Values used for those parameters and application requirements are given in Table 4.

We assume that 3G/UMTS wireless network fully overlap WLAN (i.e., IEEE 802.11) networks and MNs give more weight to bandwidth and latency requirements, $w^p_n = w^l_n = 0.35$, followed by power consumption, $w^p_n = 0.20$ and less weight for usage cost $w^c_n = 0.10$ for all $n$ and $x = 0.3$ for all $i$. The network topology considered for the analysis is illustrated in Fig. 7. It is assume that the distance between different domains is equal, i.e., $c = d = e = f = 10$ and set $a = 1, b = 2, and g = 4$. All links are supposed to be full-duplex in terms of capacity and delay. Parameter values used to compute signaling cost are defined as follows: $M = 2, \tau = 1, \kappa = 10, P_{AEN} = 8, P_{HA} = 24, P_{CN} = 4, P_{IDE} = 15$ and $P_{BEN} = 12$. The performance analysis is conducted through MATLAB and OPNET.

### 6.1. Throughput and signaling traffic overhead

Fig. 8 shows the average throughput (in packets/time slot) for both handoff decision schemes (RSS and HPIN) when the average arrival rate of packets is 5 packets per second per user. This result refers to one target MN (with two air-interfaces: UMTS and IEEE 802.11/WLAN) which moves from left to right. Handoff occurs at the following instants: 200, 380, 550 and 700 s. A significant gain in throughput can be achieved with HPIN comparatively to the RSS scheme. The target MN is initially connected to UMTS, then, it moves towards the first WLAN, after it enters in overlapping area of all networks and moves into the second WLAN before returning to UMTS. When the target MN is located in the overlapping area, we can see how HPIN allows an increasing throughput compared to RSS scheme. In fact, with the RSS scheme, UMTS is chosen more often as a target network, since it provides highest signal strength and wide coverage area. This leads to negative side

### Table 3

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Symbols</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wired link bandwidth</td>
<td>$B_w$</td>
<td>100 Mbps</td>
</tr>
<tr>
<td>WLAN bandwidth</td>
<td>$B_{wl}$</td>
<td>5.5 Mbps</td>
</tr>
<tr>
<td>UMTS bandwidth</td>
<td>$B_{wl}$</td>
<td>384 kbps</td>
</tr>
<tr>
<td>Wired link delay</td>
<td>$L_{wl}$</td>
<td>2 ms</td>
</tr>
<tr>
<td>Wireless link delay</td>
<td>$L_{wl}$</td>
<td>10 ms</td>
</tr>
<tr>
<td>Prediction probability</td>
<td>$p_i$</td>
<td>0.90</td>
</tr>
<tr>
<td>Wireless link failure probability</td>
<td>$q_i$</td>
<td>0.50</td>
</tr>
<tr>
<td>Control packet size</td>
<td>$s_k$</td>
<td>96 bytes</td>
</tr>
<tr>
<td>Data packet size</td>
<td>$s_d$</td>
<td>200 bytes</td>
</tr>
<tr>
<td>MN’s average speed</td>
<td>$v$</td>
<td>5.6 km/h</td>
</tr>
<tr>
<td>Time slot length</td>
<td>TS</td>
<td>5 s</td>
</tr>
<tr>
<td>Path loss exponent</td>
<td>$\gamma$</td>
<td>4</td>
</tr>
<tr>
<td>Shadowing standard deviation</td>
<td>$\sigma_s$</td>
<td>8 dB</td>
</tr>
</tbody>
</table>

### Table 4

<table>
<thead>
<tr>
<th>Network parameters</th>
<th>UMTS</th>
<th>UMTS</th>
<th>Idle</th>
<th>Voice (CBR)</th>
<th>Data (VBR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power (h)</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Latency (ms)</td>
<td>100</td>
<td>5</td>
<td>150</td>
<td>5</td>
<td>250</td>
</tr>
<tr>
<td>Bandwidth (kbps)</td>
<td>250</td>
<td>800</td>
<td>20</td>
<td>200</td>
<td>50</td>
</tr>
<tr>
<td>Usage cost ($/min)</td>
<td>0.3</td>
<td>0.1</td>
<td>0.5</td>
<td>0.4</td>
<td>0.6</td>
</tr>
</tbody>
</table>
effects such as lower achievable data rate and imbalanced load. However, with the HPIN scheme, the subnet load can be efficiently distributed amongst all networks, leading to higher throughput guarantees. After the session switching from UMTS to WLAN, the throughput increases since the WLAN provides better network conditions and a higher packet rate. With varying packet arrival rate, Fig. 9 shows throughput ratio which refers to the ratio of the actual data rate over the requested rate. The HPIN scheme provides a better performance than the RSS, except when networks usage is low or congested.

To alleviate packet loss, fast handoff schemes should support packet buffering and forwarding during handoff execution. Fast handoff schemes (FMIPv6 and F-HMIPv6) require more buffer space than MIPv6 and HMIPv6 since they start packets buffering and forwarding early. HPIN requires less buffer space than F-HMIPv6 as illustrated in Fig. 10. In this analysis, the required buffer space for one MN during the handoff procedure is considered. The required buffer space increases according to the number of MN performing handoff and in proportion with the packet arrival rate. On the other hand, the buffering time may affect real-time applications. For example, if certain packets are stored in a buffer for a longer period of time than acceptable end-to-end delay, they may become useless. Hence, it is crucial to manage buffer efficiently in order to minimize overhead and provide better QoS to delay-sensitive applications.

Fig. 11 illustrates the signaling overhead cost during handoff as a function of the SMR. When the SMR is small, the mobility rate is larger than the session arrival rate. Then, an MN changes its point of attachment frequently due to its mobility, which results into several handoffs and increased signaling overhead. However, when the session arrival rate is superior to the mobility rate (i.e., SMR > 1), the binding update is less often performed and results into lower signaling overhead. FMIPv6 uses the wireless bandwidth more often than MIPv6 due to the additional messages it introduces for the handoff anticipation. For lower subnet residence time, the signaling overhead reduces considerably from FMIPv6 to HPIN. Furthermore, since the reactive mode of F-HMIPv6 correspond to HMIPv6, when an acknowledgment is not received by an MN through the previous link, the messages exchanged during router discovery step becomes unnecessary. However, such messages exchange results in an increased in signaling overhead with F-HMIPv6 compared to HPIN. In fact, for F-HMIPv6, more messages are exchanged after the L2 trigger generation, which is not the case with HPIN. The RIX messages exchange introduces additional signaling similarly as with the routing information protocol (RIP). However, this signaling increment occurs only in the wired part of networks. Compared to the wireless part, the wired one has far much bandwidth and resources.
6.2. Handoff latency and packet loss

According to Fig. 12, the handoff latency increases proportionally with the wireless link delay. We can observe that MIPv6 and HMIPv6 have the worst results among all protocols, followed by FMIPv6 and F-HMIPv6, while HPIN provides the lowest delay. In F-HMIPv6, the synchronization problem mentioned above is not solved and causes packet loss as well as increased data delay. This issue is solved in the HPIN, which allows a lower delay compared to F-HMIPv6. It is well known that the maximal tolerable delay for interactive conversation is approximately 200 ms. Hence, HPIN can meet this requirement when the wireless link delay is set below to 50 ms. Since packet loss is proportional to handoff latency, similar results and behaviors are observed.

Fig. 13 shows the average packet loss versus the packet arrival rate. Packet loss is far lower for fast handoff schemes than for MIPv6 and HMIPv6. HPIN allows lower packet loss compared to other protocols. Due to the lack of any buffering and anticipated handoff mechanisms, all in-flight packets will be lost when the handoff is executed in MIPv6 and HMIPv6. However, in fast handoff schemes (FMIPv6, F-HMIPv6 and HPIN) packet loss begins from the moment the L2 handoff is detected until the buffering mechanism is initiated or if buffers overflow. This time interval is shorter for HPIN than for F-HMIPv6 due to its ability to solve the synchronization issue. Moreover, in HPIN, when the MN attaches to the new link, the redirected packets are already waiting in the NAR/AEN.

7. Conclusion

Mobility management and systems interworking are crucial in NGWN or 4G. Several IPv6-based mobility management schemes have been proposed in the literature. However, they cannot guarantee seamless roaming and services continuity for real-time applications. On the other hand, interworking architectures described in the literature cannot fulfill all requirements for sensitive (e.g., delay and packet loss) applications. This paper proposes an efficient handoff management protocol, called Handoff Protocol for Integrated Networks (HPIN) to enable a better network performance in heterogeneous IPv6-based wireless environments.

HPIN is a one-suite protocol that performs access network discovery, context transfer, fast handoff and localized mobility mechanisms. An adaptive handoff decision scheme based on the score function derived by combining various criteria such as bandwidth, power consumption, latency and monetary cost is proposed. The HPIN provides guarantees for seamless roaming, services continuity and alleviates services disruption during handoff as required for 4G/NGWN. Analyses of results from the performance evaluation indicate that HPIN improves performance in terms of throughput, handoff latency, packet loss and signaling overhead cost compared to other existing protocols, such as MIPv6, HMIPv6, FMIPv6 and F-HMIPv6. Plans for future work consist of validating numerical results using intensive simulation and prototype.

References


