

An intelligent agent-based scheme for vertical handover management across heterogeneous networks

Atiq Ahmed · Leïla Merghem-Boulahia ·
Dominique Gaïti

Received: 1 February 2010 / Accepted: 30 July 2010 / Published online: 8 January 2011
© Institut Télécom and Springer-Verlag 2011

Abstract Advances in technology have enabled a proliferation of mobile devices and a broad spectrum of novel and outbreking solutions for new applications and services. The increasing demand for all time and everywhere services requires the network operators to integrate different kinds of wireless and cellular networks. To enable this integration, it is important that users can roam freely across networks. As different technologies are involved in the current infrastructure, the problem of vertical handover needs to be addressed. To cope with the problem of seamless connectivity, several solutions have been presented. But most of them either lack intelligence or are not adaptable for reducing the packet loss and delay involved in the handover procedure. An intelligent technique is needed in order to perform the service continuity in the heterogeneous environment. This paper presents a cooperative agent based approach for the vertical handover using a knowledge plane. We propose to introduce the agents in the mobile nodes and access points to collect the necessary information from the environment. Based on this information, agents take a handover decision. A selection function is also introduced in this work which helps in choosing a best network from the available ones for handover. Finally, the proposed approach is validated with the help of simulations.

Keywords Vertical handover · Heterogeneous wireless networks · Multi-agent systems · Knowledge plane

1 Introduction

Mobile and ubiquitous computing have given users the expectation that they can access whatever information and services they want, whenever, and wherever they are. Current telecommunication market comprises of heterogeneous wireless networks providing varying coverage and quality of service (QoS). To satisfy the QoS constraints of the applications, the mobile devices need to allow seamless and fast switching, also known as handover or handoff, among various wireless network interfaces [37] while maintaining uninterrupted services during user's movement from one location to another.

In new generations of wireless networks, vertical and seamless mobility across heterogeneous networks is supported. A well-known vision of the next generation networks includes coexistence of current wireless technologies such as wireless local area network (WLAN), Worldwide Inter-operability for Microwave Access (WiMAX), General Packet Radio Service, and Universal Mobile Telecommunications System (UMTS) [19]. Different access technologies will be combined into a single network, and IP might be acting as a glue among them. This heterogeneous interworking is linked to many technical challenges like the development of new network architectures [8], new solutions for a seamless handover [17], and enhanced network management functionality for the heterogeneous wireless networks [10]. At the same time, interworking issues are receiv-

A. Ahmed (✉) · L. Merghem-Boulahia · D. Gaïti
ICD/ERA, STMR UMR 6279, Université de Technologie de
Troyes, 12, rue Marie Curie 10010 Troyes, France
e-mail: atiq.ahmed@utt.fr

L. Merghem-Boulahia
e-mail: leila.merghem@utt.fr

D. Gaïti
e-mail: dominique.gaiti@utt.fr

ing a lot more attention in standardizations [14] as well like the Third Generation Partnership Project (3GPP), 3GPP2, Internet Engineering Task Force, WiMAX Forum, and Media Independent Handover [31]; the new IEEE 802.21 standard seems to be a clear candidate of such an effort.

Nowadays, a mobile user is able to communicate through various wireless networking architectures and to roam within these architectures [4]. If the access network has to be altered, for example, due to insufficient signal strength, in many cases the connection is first broken and then re-established via a new network. This wastes time and is inconvenient for the user. For better service provision and efficient use of a network, switching from one network to another should be made seamless and/or fast, while guaranteeing a certain level of QoS during the mobility.

While designing a handover mechanism, several critical aspects have to be kept in mind. One of the most important aspects is to avoid packet loss. Packet loss in this context means that while performing a handover to another system, the terminal loses parts of the data stream that belong to the currently consumed service. Typically, such loss implies a negative effect on the QoS being delivered to the user, who is running some applications on its terminal. This effect may be very severe (e.g., while performing an important file download) or just annoying to the user (e.g., when watching a TV program).

Packet loss may basically occur due to two different reasons in relation with handovers. First, the handover process itself may cause data loss. This may happen because of the switching time from one network region to another in which the terminal is unable to receive the signal from either the previous region or the new one. Such data loss may possibly be avoided by a suitable design of the handover mechanism. Second, data loss may occur due to unexpected breakdowns (because of interference, weather conditions, etc.) in signal strength, which are typical for mobile scenarios in the border regions of cells [36]. Such data loss may be reduced by using soft handover techniques, known, for example, from UMTS, where the terminal employs a diversity reception to receive the signals of two or more cells and chooses the better signal among the available ones.

Apart from the problem described above, handover must be performed with the minimum latency as it is also a very important metric for networks. Handover latency is defined as the time interval from when the handover process starts to when the mobile node can send and receive data packets [44].

To satisfy application QoS constraints, mobile devices need to allow seamless or efficient handover among various wireless network interfaces. Thus, it is very important to develop intelligent and efficient techniques in order to provide the continuity of service and seamless mobility, i.e., service disruption should be minimal. To meet these objectives, we propose to use multi-agent systems (MAS) for the handover process. MAS has some very interesting characteristics (defined in the later sections) that have driven us to see the impact of their utilization in the above defined domain. An MAS is a group of intelligent agents, which are embedded in the network devices and are autonomous, adaptive, and distributed in nature as well. Agents have a goal in common (which no agent could achieve while being isolated) and their actions tend to achieve that goal. The agents in an MAS perform actions that enable or achieve not only their own goals, but also the goals of agents other than themselves [16, 32].

This paper starts with an overview of existing handover approaches in the literature in order to better view the problem scope (Section 2). Afterward, our agent-based proposal is discussed in detail along with the proposed agent behaviors (Sections 3 and 4). Proposed technique is evaluated, in terms of decision processing, network selection, handover blocking rate, throughput, and packet loss, with the help of simulations (Section 5). The paper closes with a short conclusion and an outlook over the future works.

2 Mobility management: a literature review

Handover is one of essential issues in mobile wireless communications since it is needed to maintain uninterrupted services during user's movement from one location to another. Handovers are mainly classified into horizontal and vertical handovers [47] as depicted in the Fig. 1. Handover within same access networks (e.g., WLAN-to-WLAN) is referred to as horizontal handover or intra-domain handover, while handover across heterogeneous access networks (e.g., GSM-to-WiMAX) is referred to as the vertical or Inter-domain handover. Vertical handover occurs when a network node changes the type of connectivity it is using to access a supporting infrastructure usually to assist node mobility.

Making a handover decision is the first step in any handover process. It plays a vital part in fulfilling users needs, balancing resource utilization, and maximizing network performance. In order to keep users always best connected [39], various QoS parameters are con-

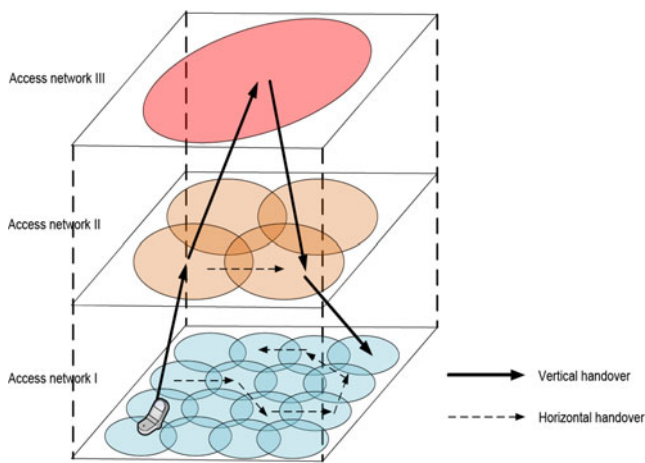


Fig. 1 Horizontal and vertical handovers

considered that affect the vertical handover decision. Received signal strength (RSS) is a traditional and unavoidable factor for making handover decisions. RSS decreases when a user moves far away from the currently accessed networks' access point (AP). The user must handover to another available network before the connection is completely lost. Ylianttila et al. [54] show that the optimal value for the dwelling timer is dependent on the difference between the available data rates in both networks. A handover is triggered if RSS is continuously below a threshold and smaller than the sum of the new RSS and a hysteresis. However, it is observed in [11] that only RSS is not sufficient enough to make a handover decision in a heterogeneous wireless environment because the RSS of different networks cannot be compared directly, and also, network conditions cannot be reflected adequately with RSS. In order to make vertical handover decisions, some of the following QoS parameters are suggested:

- Network load: Background services (e.g., FTP and e-mail) or streaming (e.g., real-time video) services perform better if higher bandwidth is provided by the network.
- Monetary cost: if two networks provide the same QoS, the network with lower cost is usually preferred by service users.
- Delay: Real-time services (e.g., voice call) are sensitive to delay.
- User preferences: User preferences can be defined on the basis of preferred network from the available ones, in order to execute the applications. Preferences can also be defined on application priority executed by the user that can either be high or

low (e.g., users usually prefer a connection with high bandwidth, low cost and reliable in terms of coverage).

- Bit error rate (BER): BER is the number of received bits that have been altered due to noise and interference, divided by the total number of transferred bits during a time interval. BER of a network may be improved by choosing a network with stronger signal in order to improve the QoS.

In [8], Wei et al. propose a policy-enabled handover across a heterogeneous network environment using a cost function. A cost function considering the parameters of RSS and bandwidth is used as a measurement of the benefit obtained by handing over to a particular network. The cost function presented in [50] is estimated for the available networks and later, used in the handover decision of a mobile node. Using a similar approach, a cost function-based vertical handover decision algorithm for multi-services handover is proposed in [53]. The available network with the lowest cost function value becomes the handover target. However, only the available bandwidth and the RSS of the available networks are considered in the handover decision performance comparisons in both solutions.

In [38], Nasser et al. present a neural network-based approach to detect signal decay and making handover decision. It is based on predefined user preferences (using the parameters of cost, power, and security) on an MN for a better network selection. But it gives high latency during the handover execution due to the large data size of neural packages, which could become an issue when implementing such a system in real time. Similarly, Horrich et al. [24] propose neural networks and fuzzy-based technique for the handover decision. They propose to use RSS, velocity, and network load as the decision parameters. But this scheme is applicable for the small environment variations and not adaptable to the new network conditions. Hence, if the network load increases due to an increase in the number of mobile nodes, then the overall performance of the proposed scheme deteriorates.

An integrated network handover decision algorithm using one of the multiple attribute decision making methods, i.e., analytic hierarchy process (AHP), is presented in [43]. They consider user preferences, service applications, and network conditions as the decision factors and use a merit function to evaluate the performance of a network in terms of revenue. They have developed a user–operator negotiation scheme for screening the network candidates according to network operator preferences in terms of long-term revenue.

The major drawback of this scheme is that it cannot guarantee completely successful handovers if operator benefits are violated.

In [22], Hasswa et al. propose a vertical handoff decision function (using the parameters of cost, security, velocity, and power consumption) that provides handoff decision while roaming across heterogeneous wireless networks. Makaya et al. [34] propose a handoff protocol, called as Handoff Protocol for Integrated Networks, to enable the better network performance in heterogeneous IPv6-based environments. It performs access network discovery and location-aware mobility mechanisms. A score function is derived by combining various criteria for the handover decision. However, these solutions lack computational intelligence techniques like adaptation and flexibility. Also, these schemes do not completely guarantee successful handovers because of the changing network conditions.

For the provision of adequate service (anytime and anywhere) to the users, applications and services should be aware of their contexts and should automatically adapt to their changing contexts.¹ In this regard, Ahmed et al. [3] propose an AHP-based context aware handover decision algorithm that deals with the session transfer of the multi-mode mobile devices. It takes into account the device capability parameters like processor speed, memory size, and battery life. A simulation software is developed that is user friendly as it is easily configurable by the users. It works only if the capabilities of the neighboring networks are being input by the user or are known in advance. This approach is applicable only if more than one network are present in the vicinity. Moreover, the proposed approach lacks QoS-related issues and deals with only non-real time applications.

In [49], Wei et al. propose a context management-based handover solution using user profile repository and a traffic monitor. It describes a mobile agent²-based decision mechanism in which, information is downloaded at the decision point and later invoked at the handover time. This scheme minimizes the handover decision time, but a single point failure may crash the whole system as all the information is stored at a single point and there is no distribution of information.

¹Context is any information that is pertinent to the situation of an entity (person, place, or object.) [23].

²Mobile agents are programs, typically written in a script language, which may be dispatched from a client computer and transported to a remote server computer for execution [6].

Similarly, Bellavista et al. propose a mobile agent-based middleware solution for the horizontal handover management in [6]. However, a large number of agents are used, which is already a load over the network. Deployment of the proposed prototype is also an issue as it incorporates only application layer protocols for session management (e.g., RTP and SIP).

After reviewing this literature, we can delegate that the need for designing an efficient vertical handover mechanism is imperative. In contrast, in this paper, we develop a novel QoS-aware and knowledge-based vertical handover approach in order to balance against user satisfaction and network efficiency. We consider network conditions including RSS, user preferences, cost, and application requirements in terms of the QoS parameters (bandwidth, delay, and BER) in order to best satisfy the users. We chose multi-agent systems as a tool for an intelligent handover decision process, in order to reduce the data loss and to improve the handover decision time. Next section describes in detail our proposed approach.

3 Knowledge plane conceptualization for handover management

Based on our synthesis of the existing handover strategies in the previous section, we have built our scheme by taking into considerations the most interesting solutions. In our view, a multi-criteria solution is needed. We take into account the contextual information from the network and MN sides, as it can be useful to avoid wrong handover decisions and, hence, bad performance of the overall handover strategy. Due to the distributed and heterogeneous nature of the information in the networks, we need a distributed knowledge base that we will refer to as the knowledge plane (KP) [13]. We introduce a novel agent-based technique to perform make-before-break handover with the help of knowledge plane. Our contribution in this paper is the deployment of multi-agent systems as a tool for the construction of a KP for handover procedure. The KP is proposed to facilitate network management by permeating the intelligence across network. We opt to analyze its efficiency in the domain of mobility and also study to what extent it optimizes the handover process.

3.1 Multi-agent systems

According to [32], the agent is an entity that performs a specific activity in an environment of which it is aware and that can respond to changes. An MAS is composed

of multiple interacting intelligent agents. These agents are able to communicate together, possess their own resources, perceive their environment, have a behavior, and have a partial representation of their environment [16]. In a typical MAS, each agent has incomplete information or capabilities. The agents cooperate with each other in order to achieve a global objective based on distributed information and control. MAS has the important characteristics of decentralization, reactivity, pro-activity and sociability, etc. [9].

Decentralization means that no agent has a global vision of the system, and the decisions are taken in a totally decentralized way. An agent is a part of an environment and its decisions are based on what it perceives from its environment and reacts accordingly. It takes a local view (also called as situated view) of its environment. Pro-activity is the ability of setting goals and realizing them. Sociability is the ability of intelligence distribution between the different agents and cooperation with other agents in the system.

3.2 Knowledge plane requirements

To build a distributed system such as the knowledge plane, we believe that the system should exhibit the following features:

- Scalability. As the amount of knowledge about the Internet is huge and distributed, we need to organize the agents in a scalable manner to better support knowledge collection and dissemination.
- Responsiveness. The KP is supposed to respond quickly to requests and distribute knowledge, where it is needed. Furthermore, in many cases, knowledge, such as the available bandwidth of a network, may become outdated quickly. Therefore, we need efficient knowledge collection mechanisms.
- Distribution. The KP needs to collect, share, and distribute knowledge, without adding too much burden to the network and to reduce network traffic overhead.
- Others. There are other properties we should consider. We should design the KP for trustworthiness and longevity, as they are important issues [13]. But they are out of the scope of this paper.

After looking at the MAS characteristics and system requirements for the handover management, we delegate that by having these characteristics, agents are very well suited for the construction of a decentralized decision system for the handover in order to maintain the service continuity. In our proposal, distributed in-

telligent agents will cooperate with each other in order to exchange the relevant information. There is a certain kind of information that would be needed by the agents in order to choose the best network for the handover. This information will be stored in the KP, and each time a handover would be needed, this KP will be consulted for the availability of networks.

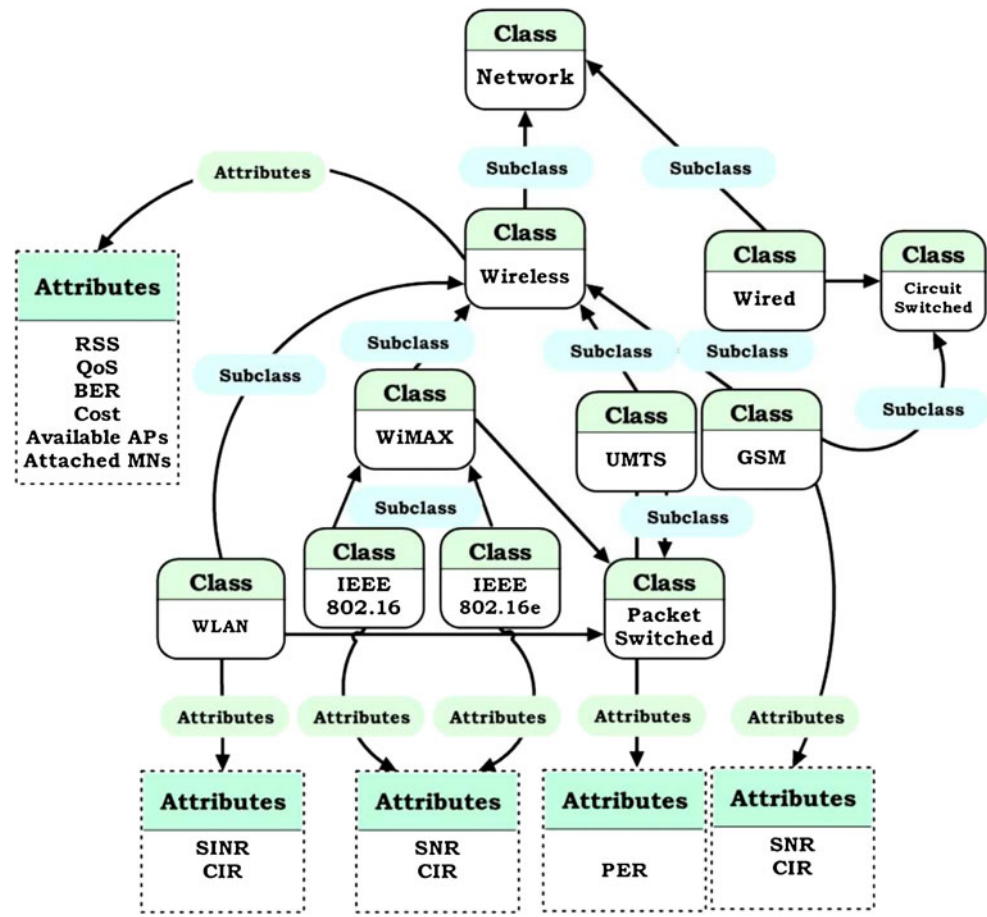
3.3 Knowledge plane design

In this paper, knowledge refers to any useful information in the Internet, including the information about individual objects in the network (MNs and APs) and the relationships between these objects. There are two ways to categorize knowledge. First, knowledge about network that may originate from any layer in the protocol stack (cross layering). Second, based on the form of the knowledge, we can classify it into facts and relationships. Sources of knowledge include measurements, reasoning, and monitoring from the environment. For designing a knowledge plane, knowledge management is a very important task. Knowledge management is hard to define precisely and simply. According to our understanding and context, it is an agent activity that takes care of integration of heterogeneous information (knowledge) in order to improve the handling of knowledge at the agent (individual) and multi-agent (cooperative) levels [30]. Therefore, we have divided this task into three subtasks as depicted below:

1. Knowledge representation addresses the question of how agents represent their local knowledge and exchanged knowledge with other agents in the network in order to maintain a distributed knowledge base and a rapid access to the stored knowledge (Section 3.3.1).
2. Knowledge plane organization refers to how agents organize themselves together in a heterogeneous environment and how they manage a distributed KP (Section 3.3.2).
3. Request propagation and knowledge dissemination deal with how to propagate requests and responses so that requests can be solved quickly and efficiently and also how to disseminate knowledge so that agents needing that specific knowledge can receive it in a timely manner (Section 3.3.3).

In order to accomplish these tasks, we propose to deploy intelligent agents in the mobile nodes and access points. These agents are responsible for collecting, storing, disseminating knowledge, and resolving requests coming from their one hop neighbors. Agents also manage a distributed knowledge base. This knowledge

Fig. 2 A network ontology model for vertical handover



base is distributed among agents since the knowledge is distributed in the Internet and is managed by different parts of the network (MNs and APs). The next two tasks (i.e., representation and organization) are very closely related to each other because agent organization largely determines how requests and knowledge can be propagated.

3.3.1 Knowledge representation using network ontology

Knowledge representation is one of the main issues in the context-aware systems. In this work, we opt to define an ontology to accomplish this task. According to Gruber et al. [20], an ontology is a specification of a conceptualization represented in some formal language. Some markup languages and information models have been proposed as a basis of context representation, such as Extensible Markup Language (XML), Resource Description Framework, or Web Ontology Language [20, 48]. We have opted to use XML because of its simplicity, flexibility, and ease of use. Information is represented in the form of tags and structural relationships between them. For this purpose, information

is stored in the form of classes, sub-classes, and their attributes. The ontology that has been proposed in this paper is inspired by the ones created during the course of the Situated and Ubiquitous Networks³ project [45] in which we are one of the partners. Figure 2 gives a brief description of the network ontology schema.

In the Fig. 2, “Network” represents the main class while others are the sub-classes of it. Network has the wired and wireless sub-classes. As we are in a heterogeneous wireless environment, we extend this concept to WLAN, WiMAX, UMTS, and GSM networks. These subclasses are extended from the wireless class. All the subclasses have some common attributes, which are added in their main class Wireless. Each of these subclasses also contains the attributes associated with a specific network. This ontology is flexible enough to add more attributes, if necessary. These attributes (RSS, bandwidth, cost, and QoS) are very important for the handover decision and selection process. These

³<http://projet-sun.it-sudparis.eu>

attributes represent the preferences of an application in terms of QoS and cost, for example. The user can choose a network on the basis of the values of these attributes. A user may prioritize a network by inserting its preferences, i.e., if a user requires a high bandwidth network, then our proposed scheme will select a network from the knowledge base that provides high bandwidth indifferent from the offered cost.

3.3.2 Situated view of an agent

The collected information can be relative to the network (APs) and the MN. The information collected by an MN is different from that of an AP. This information is in:

- AP context: Intuitively, it represents the social context of an AP and its awareness of the existence of other APs and MNs in the surroundings. That is, a social context gathers parameters representing the information provided by other access points. In some sense, this context adds a communication dimension to the individual context. This communication allows, for instance, a node to detect the presence of other APs, estimate their proximity, exchange information about individual contexts, or services they can offer. It also helps in collecting important information related to QoS parameters (available bandwidth, packet loss, etc.), network coverage, monetary cost, link quality as RSS of the current access network and its neighbors, and number of attached MNs.
- MN context: It represents the local view of the environment of a mobile node. That is, it gathers all parameters representing information about its environment that are directly accessible to it. Such context parameters typically capture the state of the device or some other values measured by the device in its surrounding environment, e.g., running applications, terminal status, RSS, available APs, and velocity.

An AP can detect the other APs that might not be in the signaling range of an MN; hence, an AP may inform MNs about the presence of those APs. Locally collected information and the exchanged information with the other APs and MNs constitute the local (or situated) view of the agent (Fig. 3). We have a situated view limited to one hop only; MN needs help of the network in order to update its local view, and for this purpose, it exchanges its parameters with the neighboring APs. This assistance can be demanded specifically from the current or the target attachment point.

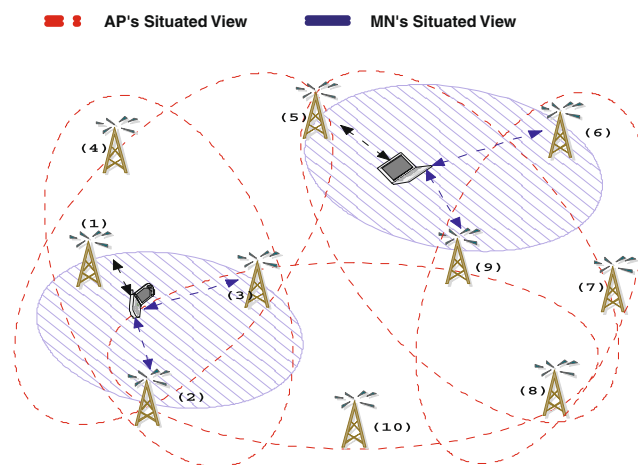


Fig. 3 Situated view of an agent

3.3.3 Agent cooperation

As we said earlier that each agent in a multi-agent system has a partial information about its environment, so it needs to cooperate with the other agents in the system. Different methods of cooperation are presented in the literature, such as voting, contract-net protocol, and negotiation [21, 27, 46, 51, 52]. The contract net is a protocol for task sharing among agents. It is done by the preparation and execution of a contract between an agent and a manager agent contractor. Negotiation is the process by which a group of agents come to a mutually acceptable agreement on some matter using the shared information present in the system. Beside this, information sharing may be conducted using a blackboard.

We have chosen to send and receive simple cooperation messages between the agents residing in APs and MNs. By doing so, agents residing in APs and MNs create their situated views for taking a handover decision on the basis of communicated knowledge. At the node level, the available interfaces are monitored (layer 2 and layer 3 monitoring) to reach access networks where the user preferences are already defined (fast connection, preferred network, etc.). At the AP level, QoS parameters or cost can be publicized by the available access networks. For exchanging this information with each other, agents residing in AP and MN send simple messages to the neighboring networks for inquiring about their parameters. A standard message format with the meaningful structure and semantics and a mechanism for agents to exchange ontologies and message interpreters have become key issues. Furthermore, the message format should be accepted not only by the agent research community but also by all

information providers. That is why in our case, agents send and receive information through XML-encoded messages. We have chosen to use a KQML/FIPA ACL-like format, encoded in XML [18]. A frame format of this message is presented in the Fig. 4. Java Document Object Model [26] API has been used to parse through the XML-based ontology. In fact, an XML document is a database for storing the network parameters, which can be used by any receiving agent. Further, an agent can carry an XML front-end to a database for data exchange, where both queries and answers are XML encoded. In order to better explain the agent communication, we present a brief scenario.

Example scenario In Fig. 5, we have two autonomous networks (AN1 and AN2) in an overlapping coverage. AN1 contains three access points and a mobile node (MN₁) while AN2 contains two access points and a mobile node (MN₂). MN₁ is connected to AP₅, and it is in the signaling range of two other APs, i.e., AP₁ and AP₄. Similarly, MN₂ is connected to AP₃, but it can sense the signals of AP₂ as it is in its range. For the cooperation purpose, agents exchange three different XML-encoded KQML/FIPA ACL messages that are as follows:

- Request** is a type of message that is periodic among the APs, and it is on-demand between APs and MNs. In order to update the information in the knowledge plane, MN sends requests to the APs about their new collected knowledge.
- Response** is sent to the agents residing in the MNs and the APs for their information requests.
- Inform** is the type of message that is periodically sent from an AP to other APs and to an MN that has requested information in or-

Element	Description
Sender	MN_id
Receiver	AP_id
Reply-to	MN_id
Content	RSS, avail_bw, cost, BER, Net_Type, Id, Neigh_APs
Language	xml
Encoding	<?xml version="1.0" encoding="UTF-8"?>
Ontology	Handover.xml

Fig. 4 A cooperation message frame format

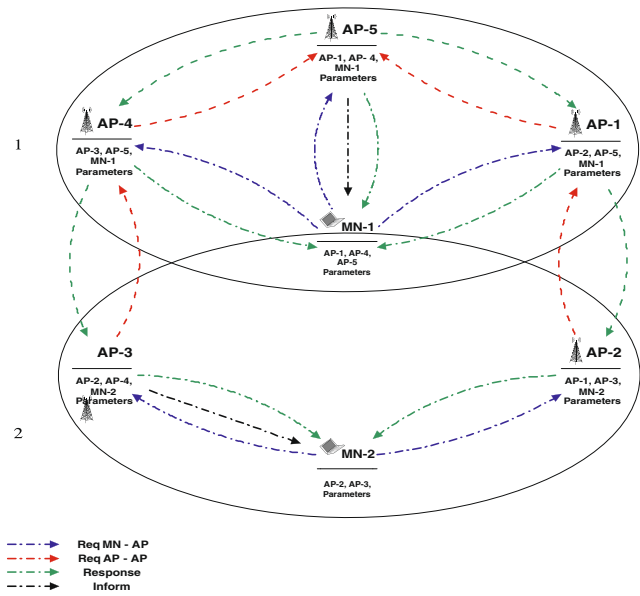


Fig. 5 Agent cooperation in wireless networks

der to update its parameters or about the presence of a new network or about any anomaly in the existing connection.

The situated view of both the MNs and all the APs is different from one another, which means that all of them have different information stored in their knowledge bases. Suppose that a multimedia application is launched by the MN₁, it sends a request message (Req_{MN1→AP5}) to the currently connected AP, i.e., AP₅. This message is sent in order to know the internal state of the connected AP₅, i.e., its RSS, QoS, number of connected MNs, and neighboring APs. Upon receiving this request message, the agent residing in the AP₅ sends a response message (Res_{AP5→MN1}) to the MN₁ by informing it about its parameters. Similarly, AP₅ sends the request message (Req_{AP5→AP}) to all the neighbors that it can detect in order to know their internal states, which reply with a response message (Res_{AP→AP5}). All this information is stored in the agent knowledge base if different from the existing one. An MN can also send requests to all its neighboring APs that it listens to. Upon receiving the responses from the detected APs, MN updates its local view. In this way, an agent only maintains the local view of the network. This is due to the following reasons: (1) it is difficult to get an accurate global view of network topology, but it is easy to obtain the local view; (2) such local network knowledge should be able to satisfy local requests most of the time, and (3) remote network

knowledge can be obtained from other agents through sending requests.

The purpose of this communication is to keep an updated situated view and to anticipate the handover toward a better network if the current connection does not satisfy the application/user QoS needs. AP also has stronger transmission power as compared to an MN, so the information collected by an AP may concern those networks that were not detected by the MN. For example, if on MN_1 , an application is launched requiring a high level of QoS. MN_1 receives the best signal of AP_5 ; however, AP_5 is already very loaded in terms of users connected, and consequently, necessary QoS for MN_1 cannot be assured on AP_5 . A decision strategy consists in guiding the MN_1 toward AP_1 which is empty or has less number of attached clients and which can provide the necessary QoS.

During this communication between the agents of APs and MNs, if all the interfaces in an AP are found filled, the user must be able to use another access technology in order to continue its applications, which corresponds to its needs. APs send information to the MNs periodically while MNs may ask about information on demand when needed in order to save their resources. In this way, agents in APs and MNs cooperate with each other for updating the database entries and anticipate handover for assuring the continuity of service. These entries in the database can be static (such as user preferences, cost of the different access networks) or dynamic (such as access network availability, MN's velocity, RSS, and number of attached clients). These values are very important as decision parameters, for the handover decision, which is explained in the coming sections.

Mobile users would like to avoid inappropriate handover decisions due to the changing costs of different wireless networks (e.g., WLAN, WiMAX, and UMTS). Users may require to choose either a cost effective network or a best performance network (in terms of QoS) from the system even if it is costly as the connectivity is more important, or the user needs a dynamic hybrid approach which exhibits different behaviors in different environments.

4 Handover management

Unlike the common vertical handover algorithms [11, 22, 54], in which only the mobile node or the network does the computation work, in our algorithm, both the mobile node and the network side are responsible for not only the information gathering but also the process-

ing of the collected information. For this purpose, we have defined three agent behaviors: information gathering, decision making, and selecting a proper network from the available ones. We define an agent behavior as an algorithm, which has some intelligence aspects and which is based on some knowledge (present in the knowledge base) for taking some decision and that may be executed concurrently with the other behaviors. This means that multiple agent behaviors can be activated at the same time. A trigger is needed to activate any behavior in the system. For example, in our particular case, degradation in QoS, RSS, or change in user preferences, etc. are the triggers to activate or deactivate any behaviors [1, 2]. In the following, we present the agent behaviors involved in our proposed scheme.

4.1 Network monitoring and information gathering

The use of intelligent agents facilitates the deployment, maintenance, and management of the handover process. MNs and APs collect all the necessary information, i.e., RSS, network load, distance, cost, and BER from the network environment. As we said earlier, every MN and AP has an agent running on it; both MN and AP collect this information and send it to their corresponding agents. Agents are like a virtual layer deployed over the MNs and APs. The agents in the MNs cannot communicate with each other, but the agents residing in APs can communicate with the other agents in other APs as well as the MNs too. Apart from its own collected information, an MN may ask for the processed information related to neighborhood from the APs as processing the information on an MN consumes terminal resources heavily. The new networks in the locality are discovered with the help of RSS, which can be calculated as in [1, 55]:

$$RSS = P_t - L - 10n \log(d) + v \quad (1)$$

where P_t is the transmitted power of an AP, L is the power loss, n is the path loss exponent, d is the distance between an MN and an AP, and v is the velocity of the MN.

In order to collect the information, two methods of scanning are commonly available in the literature: active scanning and passive scanning [42]. Active scanning refers to keeping on all the interfaces continuously active while passive scanning refers to the periodic activation of the interfaces for verifying the availability of new networks. However, both these methods result in unnecessary power utilization when a mobile node is far from an available AP. Hence, it is very important

to adopt a time adaptive scanning process, as described below in Eq. 2 [40]:

$$S_{\text{int}} = S_{\text{int.min}} + (S_{\text{int.max}} - S_{\text{int.min}}) \left(\text{RSS}_{\text{cur}} - \frac{\text{RSS}_{\text{new}}}{\text{RSS}_{\text{th}}} \right) \quad (2)$$

where S_{int} is the scan activation interval time, i.e., the time between two interface activation operations; $S_{\text{int.min}}$ is the lower threshold value of S_{int} ; $S_{\text{int.max}}$ is the upper bound of the S_{int} ; RSS_{cur} is the signal strength of the current network; RSS_{new} is the signal strength of the newly available network; and RSS_{th} is the minimum threshold value of the RSS for adding a new network to the agent knowledge base.

This scheme is chosen in order to reduce the power consumption of the mobile node and signaling overhead produced as a result of agent cooperation messages with the neighboring APs. It is very important to note that different networks have different interface activation intervals, e.g., interface activation interval of a WiMAX cell is different from that of a WLAN network base station. However, the fact of having multiple active interfaces can increase the amount of energy consumed by a mobile device, reduce devices lifetime, and can cause the interferences and other physical effects like weather conditions, etc.

According to Algorithm 1, mobile nodes and access points both scan their environment in order to detect the changes and collect device parameters. This algorithm executes on both mobile node and access point. The agent starts by monitoring the network neighborhood. As the MN detects a new network, the agent measures the received signal strength of the current network RSS_{cur} and compares it to a defined threshold T_{min} . If the RSS_{cur} is lower than the threshold T_{min} , decision behavior is activated. This is where handover anticipation is done in order to stay connected. De-

crease in the RSS below the threshold level triggers and starts the handover anticipation process. On the other hand, if the RSS_{cur} is higher than the value of the threshold T_{min} , RSS of the newly found network RSS_{new} is compared to the T_{max} . The new network is added into the database if the RSS_{new} is sufficient enough for a future handover, based on the value of the T_{max} . Next, we present our handover decision process.

4.2 Distributed vertical handover decision

Handovers are normally classified into two main kinds based on the communication session continuity: imperative and alternative [41]. Imperative handovers are executed because of the low link quality of a network. For example, if the RSS of the serving access node drops below a specific handover threshold, a random network selection is done in order to execute the handover. Alternative handovers are those that are used to provide the users with a better service or performance. These handovers can tolerate longer handover latency as compared to imperative handovers and can be sub-divided into QoS-related, and Authentication, Authorization, Accounting, and Charging-related handovers [41]. For example, a user may ask for more bandwidth to accelerate a data transfer, or need a cheaper network for reducing the service cost. Handover decision could be triggered from many instances, i.e., user demands, coverage limitations, or any other specific requirements. The trigger for handover could originate from any layer in the communication stack, e.g., degradation in the available bandwidth and/or a constant degradation in the RSS of a network, may initiate a handover.

Decision behavior is the preliminary step for the network selection process as it is based on multi-criteria decision and selection commonly known as multi-criteria decision making [12]. In our proposed approach, we consider multiple parameters, i.e., RSS, user preferences, cost, and bandwidth in order to best satisfy the users in terms of coverage, service continuity and QoS, etc. We have divided this mechanism in three logical function blocks: processing the collected information from network monitoring, making decision, and selecting the appropriate network for handover. Normally, the handover initiation evaluates user location changes, i.e., users may enter or leave a particular network coverage. The evaluation of user location changes is based on the distance between user and the serving AP and is carried out based on the RSS [7]. The system adopts the network-assisted approach, where the network assists the terminal in the decision process by performing data collection, processing, and

Algorithm 1 Network monitoring process

- 1: Scan for new networks
 - 2: Measure the RSS Levels
 - 3: **if** $\text{RSS}_{\text{cur}} \leq T_{\text{min}}$ **then**
 - 4: enable decision process
 - 5: **else**
 - 6: **if** $\text{RSS}_{\text{new}} \geq T_{\text{max}}$ **then**
 - 7: add new network to Agent knowledge-base
 - 8: **else**
 - 9: reject the network
 - 10: **end if**
 - 11: **end if**
-

analysis. As our proposal is based on the RSS threshold values, we have the following three cases in this regard as depicted in the Algorithm 2:

1. if $RSS_{cur} \leq T_{min}$: In this specific case, MN performs imperative handover as the value of RSS is very critical. As the selection of the network is random in this case, mobile user mU_i is simply switched to access network AN_j with a probability $Pr[mU_i = AN_j] = 1/n$, where n is the number of available access networks in the agent knowledge base.
2. if $RSS_{cur} \geq T_{max}$: In this case, MN stays connected to the current AP and returns to the network monitoring process. The condition may vary if the velocity of the MN changes; in that case, MN would choose a large coverage area network to stay connected and continue its services. For example, if an MN is moving with a high speed, then it would be unsuitable to choose a WLAN due to its shorter coverage area as compared to a WiMAX network with a larger coverage zone. Velocity threshold adaptation and estimation is not a simple task as it is not easy to predict the velocity variations of a mobile node. We have opted to use a constant velocity of an MN. We have done so, by limiting the MNs' velocity to pedestrian speed during our tests and simulations. We plan to test the high mobility

Algorithm 2 Decision algorithm

```

1: if  $RSS_{cur} \leq T_{min}$  then
2:   Imperative Handover using  $Pr[mU_i = AN_j] = 1/n$ 
3: else
4:   if  $RSS_{cur} \geq T_{max}$  then
5:     stay connected
6:   else
7:     if  $T_{min} < RSS_{cur} < T_{max}$  then
8:       measure user preferences (if any)
9:       if  $AppPriority = High \ \&\& \ Bandwidth_{cur} = High \ \&\& \ Cost_{cur} = Low$  then
10:        stay connected to the current network and
        continue with the network scanning
11:      else
12:        if  $Bandwidth_{cur} = Low \ \&\& \ Cost_{cur} = High$  then
13:          enable network selection
14:        if  $AppPriority = Low \ \&\& \ Bandwidth_{cur} = High \ \&\& \ Cost_{cur} = Low$  then
15:          stay connected
16:        else
17:          enable network selection
18:        end if
19:      end if
20:    end if
21:  end if
22: end if
23: end if

```

scenarios in the future along with the localization aspect of the mobility.

3. In this step, the value of RSS is between the threshold values T_{min} and T_{max} , and further calculations are performed. The system is checked for the availability of user preferences, i.e., if the user has any specific requirements regarding the type of network, bandwidth, cost, QoS, etc. This information is extracted from the agent knowledge base, where all available networks in the neighborhood are stored. Later on, application priority is checked. Here, we have two cases: high priority and low priority applications. If a high priority application is running with high bandwidth and low cost, then agent in the MN stays connected and continues with the networking monitoring process. In contrast, if the same application is receiving low bandwidth and high cost, then network selection process is enabled in order to choose a better network in terms of QoS to better satisfy user needs. The same procedure is applied for low priority applications.

4.3 Access network selection

Network selection for optimized service provision in heterogeneous access networks, when a mobile user has a choice between multiple available networks, is an area of active research [5, 15, 28]. Network selection is a very complex, yet important process before the handover execution. Access network selection is the mechanism of finding the best available network that is best suited for a specific mobile user, taking into considerations the requested/running services. The processed information from the monitoring and decision behavior works as the input data for the selection behavior. Processed information can be relative to:

- A set of measurements reflecting the availability, signal quality, and other parameters (e.g., access technology, number of attached clients, cost, and available bandwidth of the networks) perceived from each of the available APs
- Applications that are running on MN, the corresponding quality levels (e.g., high bandwidth) at which these applications are being provided as well as the set of applications that the user is requesting to use
- The set of user preferences, according to which the parameters of bandwidth and cost are prioritized

The majority of the parameters have a momentous effect on the correct network choice. In our proposal,

proper network selection comprises of the following two steps as depicted in the Fig. 6 as well.

4.4 Weight assignment

The computed access network selection is equivalent to a best possible provision of both the requested and the already running services to appropriate quality levels and also the requested services to network interfaces. These allocations are bound to certain constraints, (a) the quality level demanded by the user should be guaranteed, (b) the user must connect to a comparatively better RSS value than the current network, and (c) the user preferences should be satisfied in the specified order. This functionality can be obtained by defining a selection function (SF) by assigning the weights attributed to the selection criteria. By using this function, we can calculate the network scores of the available APs. Suppose A is the set of available APs in the agent knowledge base monitored by the MN and its connected AP itself,

$$A = \{a_1, a_2, a_3, \dots, a_i\}, i \in \{1, 2, 3, \dots, n\} \quad (3)$$

and $Q(a_i)$ is the quality level at which AP a_i can provide its services (RSS, available bandwidth and cost),

$$Q(a_i) = \{q_1, q_2, q_3, \dots, q_j\}, j \in \{1, 2, 3, \dots, m\} \quad (4)$$

In order to select the best network from the agent knowledge base, we need to compute SF:

$$\text{SF}(a, q) = w_{\text{RSS}} \times \text{RSS}(a, q) + w_{\text{nt}} \times \text{NetType}(a) \\ + w_b \times \text{Bandwidth}(a, q) + w_c \times \text{Cost}(a, q) \quad (5)$$

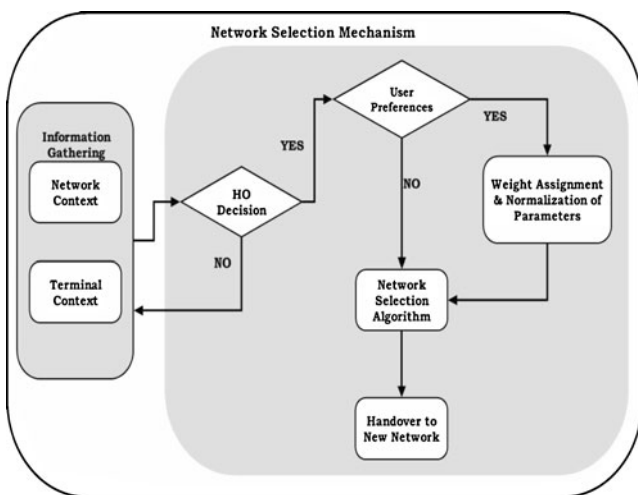


Fig. 6 Network selection process

where $a \in A$ and $q \in Q$. It is assumed that the weight assignment in the SF is based on user's profile. $\text{RSS}(a, q)$ in Eq. 5 represents the perceived signal quality of each AP a . $\text{NetType}(a)$ is the network type and it represents the preferences of a mobile user for the available network. $\text{Bandwidth}(a, q)$ and $\text{Cost}(a, q)$ refer to the available bandwidth and cost of an access network a at quality level q , respectively. The cost of the available networks is already stored in the agent knowledge base along with the other parameters. The choice of the weights of different networks is a very important process as they reflect the user preferences over one network or another. Through this process, the user is able to specify his/her preferences and that allows him/her to give different priorities to parameters that may influence the access network selection process. This prioritization is equivalent to the specification of values for the different coefficients w_{RSS} , w_b , w_{nt} , and w_c , which correspond to parameters RSS (i.e., the RSS of a network), available bandwidth, preferred network technology type, and cost (i.e., the maximum cost the user is willing to pay), respectively. This represents the measure by which each one of these parameters is weighted in the access network selection algorithm. For example, if the user chooses to specify that at any time RSS for him is the most important factor in access network selection, bandwidth comes second, cost comes third, and last comes network type, then the respective coefficients will be assigned values as $w_{\text{RSS}} > w_b > w_c > w_{\text{nt}}$ where $\sum w = 1$. After the weight assignment to the network parameters, normalization of those parameters is very important as normalization helps us to process different network parameters easily. This processed information tells us about the cost of a service being provided at a specific quality level, by a specific wireless network operator and through a specific technology (e.g., WiMAX, UMTS, etc.), per data volume unit (e.g., per Kb or Mb) or per time unit (e.g., per second, per minutes).

4.5 Parameter normalization

Next step in the selection process is the normalization of the available network parameters. By using the simplified Grey Relational Analysis method, we have to sort the parameters in the following sets, larger-the-better and smaller-the-better [29]. The parameter of RSS and bandwidth of a network belong to the larger-the-better set while other parameters belong to smaller-the-better (for example, cost). In the second step, we have to normalize the values of these parameters. In

case of smaller-the-better parameters, the lowest value (l_j) is the best. Hence, the normalization factor can be calculated as [29]:

$$N'_i(j) = \frac{U_j - N_i(j)}{U_j - l_j} \quad (6)$$

where U_j represents the highest while l_j is the lowest value and $N_i(j)$ is the actual while $N'_i(j)$ is the normalized value of the network parameter. If $N_i(j) = U_j$, then the network under consideration gets the lowest normalized value for a QoS parameter, i.e., (0). The network, which has the lowest value, is assigned with the best normalization value according to user preferences, i.e., (1). In contrast, if we assume the case of larger-the-better division, then the best QoS parameter value is the highest one (U_j), i.e.,

$$N'_i(j) = \frac{N_i(j) - l_j}{U_j - l_j} \quad (7)$$

This implies that the normalized value of a given parameter will be higher, when $N_i(j)$ reaches the U_j . If the highest value of the given parameter becomes equal to the $N_i(j)$, then the network will be the best in the agent knowledge base and the normalized value will be (1).

All this processing, i.e., weight assignment to the network parameters and their normalization, needs some time to be accomplished. Once this task is completed, a prioritized list of networks is created in the agent knowledge base. We assume that there is enough time for this, since our proposed algorithm tries to see if it can better satisfy the user by providing access to a more favorite access technology. We know that the time constraints are very tight in case of vertical handover. The highest priority is on the service continuity rather than the fine-tuning of the decision according to all parameters. A vertical handover will be allowed, only if it is acceptable by the user preferences, for example, a mobile user may not be willing to pay a higher price for a UMTS network for an FTP session. So, if it is acceptable, a similar procedure is followed as before. If the selection of a specific access network implicates speed considerations, e.g., a limited range WLAN, then the procedure for their evaluation is started. This is required since it is meaningless to handover to a WLAN AP if the MNs' speed is very high and it will leave the APs' coverage in few seconds. Agent knowledge base will assist in choosing a network with a wider coverage in order to reduce the number of possible unnecessary handovers. Finally, the last trigger is a handover request indicating an imperative handover. We would like to

remind here that the MN has already stored the alternative access networks in its knowledge base in order to continue its services; hence, it selects one of them according to coverage and user preferences to ensure the service continuity with user satisfaction.

5 Evaluation

In this section, we will discuss the results obtained through simulations and their analysis.

5.1 Simulation setup

In order to analyze the performance of our handover decision algorithm, simulations are conducted using the MATLAB and its JAVA deployment tool MATLAB-JA [35]. It is a graphical user interface that comes with MATLAB Compiler. It lets us build our MATLAB applications into Java classes and then package the components for deployment. For the deployment of agents in our proposed approach, we have modified the main control frames, so as to simulate the basic measurement mechanisms and incorporate the appropriate information, for example, cost or bandwidth of a network demanded by a user. This modification does not change the operating methodology of the control mechanism. We simulate a heterogeneous networks environment, containing WLAN, WiMAX, and GSM networks, in order to see the benefits of our proposed scheme.

The simulated wireless network consists of 28 (7×4) hexagonal cells. The diameter of each cell is fixed to 1 km. Every cell contains a minimum of two WLANs and one WiMAX or cellular network in an overlapping fashion. Each cell has a total bandwidth capacity of 10 Mbps. The simulated model of the wireless network is shown in the Fig. 7. In order to avoid the edge effect of the network size, we have applied a wrap-around the edge cells. In this way, every cell in the network has a minimum of three and a maximum of six neighboring cells and also to ensure that there is always a network available for the users to execute their tasks. Some parts of the network are available in the overlapping coverage of two or more networks as well. This wrapping will help us in better understanding the network selection process in the presence of user preferences.

The simulation parameters are presented in the Table 1. For now, we assume that MN is always authenticated in the foreign networks, and it has free access in

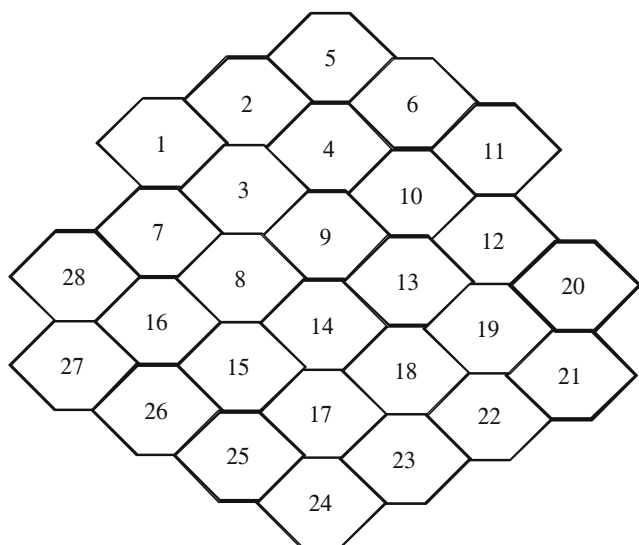


Fig. 7 Wireless network model layout

foreign networks, as we do not take into considerations AAA and admission control mechanisms in the current work. We intend to integrate these features in the future. MN moves with a constant velocity of 2 m/s. We have opted to execute real time (online video) and non-real time (file transfer) applications over an MN. An MN can create multiple sessions depending upon the priority and importance degree of these applications.

5.2 Simulation results and analysis

For the evaluation of our proposed agent-based scheme, we have compared our approach with a classic handover [54], a cost-based function [53], and a context-based solution [49]. Context-based solution has been used only for the decision processing delay and commu-

Table 1 Simulation parameters

Parameter	Value
Simulation time	50 min
Cell radius (cellular and WiMAX)	1,000 m
AP transmit power	38 mW
AP sensitivity	80 dBm
Max AP transmit power	80 dBm
Min AP transmit power	70 dBm
Number of channels	13
Thermal noise power	116 dB
Transmit power	1 mW
Data rate	3 Mbps
Frame rate per second	5
WLAN range	100 m
Traffic type	FTP, video
Maximum number of admitted users	50

nication overhead measurements because they do not consider the parameters of QoS and handover blocking in their implementations. We have taken into account the following parameters [33]:

- Quality of service
- Handover blocking rate
- Decision processing and communication overhead

5.2.1 Quality of service

One of the main objectives of the handover is to guarantee the quality of service. For this experiment, we measure the parameters of packet loss, latency, and throughput in order to evaluate the obtained QoS for our proposed scheme. Figure 8 depicts the amount of packets dropped due to channel errors and handover in the communication sessions. We can see that our handover mechanism achieves lower percentage of dropped packets. It is true that our proposed scheme has obtained very close results to the cost based scheme [53], but we have improved the performance in case of handover delay. Our proposal provides a balanced network operation. MNs that face poor network conditions by having low RSS, low bandwidth, and high cost perform handovers in order to choose a better network while the proposed underlying handover mechanism optimizes the MNs handover decision. Our proposed mechanism decreases further the number of lost packets and manages to keep it low as it does not choose a network which is heavily loaded.

In the Fig. 9, we observe the average handover delay in the voice packet transmission while performing handover. The handover delay is affected by the

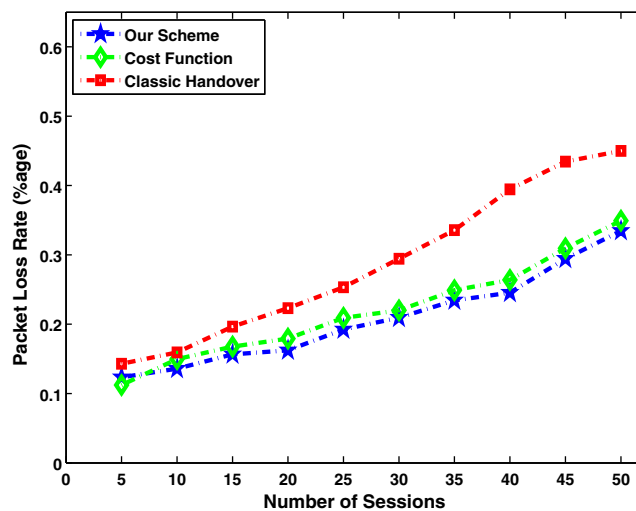


Fig. 8 Packet loss rate

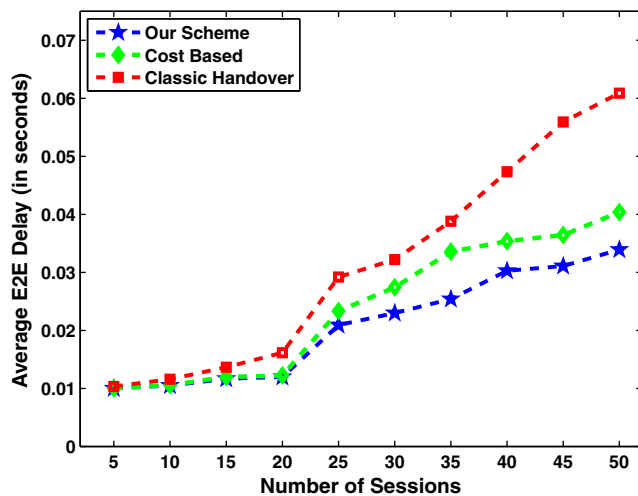


Fig. 9 Average end-to-end transmission delay

routing delay that is introduced in the simulation while changing from one network to another. In our proposal, we achieve low handover delays. This improvement is also due to the make-before-break mechanism that is provided. Therefore, we have a network performance improvement of approximately 11% and 27% as compared to the cost-based and classic handover schemes, respectively.

WLANs have a small coverage range, but they offer much higher throughput than the cellular networks. Throughput of a network depends on the number of active devices sharing the link. Bit rates for file transfers may vary, depending upon the available access network type, its configuration, and load. Different users can have different expectations for file completion times after the handover depending upon their terminal capacity in terms of power, memory, and processor speed, and the same users can change their expectations depending on the current context, or over time as higher bandwidth access networks become available and/or network cost decreases. Network input includes the available bandwidth, which is reported by the network at some cost and is stored in the agent database, and the throughput, which is observed at the terminal after the handover completion. Figure 10 depicts the comparison of our approach with the classic and the cost based approaches in order to reflect the average throughput obtained after the handover execution by transferring a file of 2 MB.

This handover is performed between GSM and WiMAX networks. In the Fig 10, we can see that all the presented schemes change their network between 1 and 7 s. There is a variation of 1 to 2 s in the handover timings. As the handover is executed, the perceived

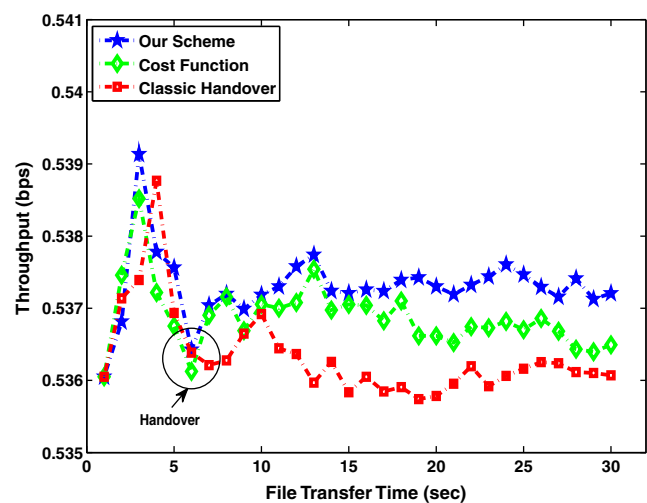


Fig. 10 Average throughput

throughput of our approach stabilizes till the end of the file transfer. This happens because we perform a handover toward the best available network in the agent database. As there was a file transfer, our intelligent selection mechanism gave the priority to the parameter of bandwidth and continued with its activity. Here, we would like to recall that all the networks, which are present in the database, are periodically updated by monitoring behavior. So, all the networks have the RSS well above the defined threshold, and there is no issue of connectivity in terms of RSS. Hence, we have obtained better results with respect to QoS as compared to cost based and classic handover schemes.

5.2.2 Handover blocking rate

Besides the aforementioned objectives, another consideration in handover is to calculate the handover blocking rate, which occurs when the mobile node chooses an inappropriate wireless access network. An incoming call (either new or handover) is blocked when the base station fails to allocate a free channel for it. It represents the percentage of calls that did not finish their services. Lower level of blocking rate reflects the overall system service continuity and the scheme efficiency. We have used a generic mathematical model for mobility management protocols presented in [25], in order to calculate the blocking rate. This model takes into account the network topology and the movement patterns of a mobile node.

Figure 11 depicts the decreased handover blocking rate for our proposal as compared to the classic scheme [54] and cost-based scheme [53]. During our simulations, we have allowed a maximum number of 50 users

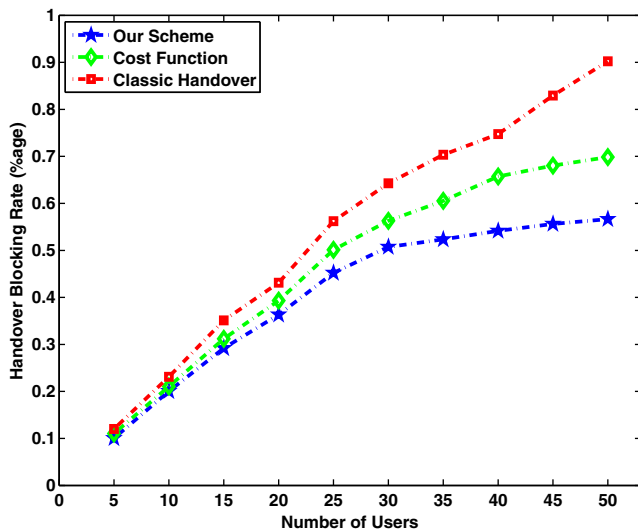


Fig. 11 Handover blocking rate comparison

at a time as described in the Table 1 as well. During the simulations, we have observed that as the number of mobile users stays less than 20–25, all the approaches behave similarly, but as this number starts to increase, our proposed scheme outperforms the classic and cost-based schemes. This can be explained by the fact that the handover resources, in these two approaches, are

not sufficient to satisfy all the handover requests. Our proposed scheme stabilizes even after the admissions of more users.

5.2.3 Decision processing and communication overhead

Decision processing delay is the processing time needed by the mobile node to decide toward which network it has to go. This notion should not be confused with handover delay, which is the transfer time of an MN to leave the current network and get connection to the new network. Our experiments show that by distributing the parameter processing task among APs and MNs instead of the mobile nodes only, the processing time at the mobile node side decreases significantly. As APs do most of the processing, MNs conserve their resources.

On the other hand, in the Fig. 12, we have obtained better performance if we compare our scheme with the context-based solution, where all the information is first downloaded over the terminals and then processed or sent, where demanded. In this context-based solution, every time a service is initiated or the context exchange format is changed, the underlying service modules need to be downloaded and updated. This wastes time and energy of the mobile terminals. This happens due to the platform dependency of the installed modules

Fig. 12 Information processing overhead

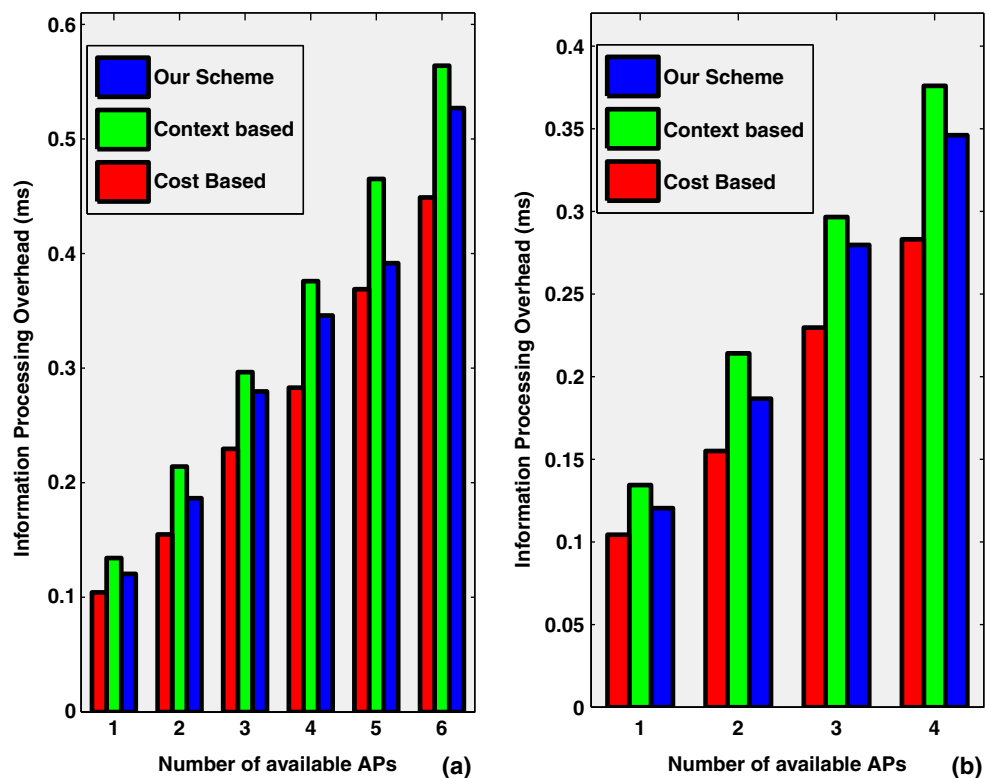
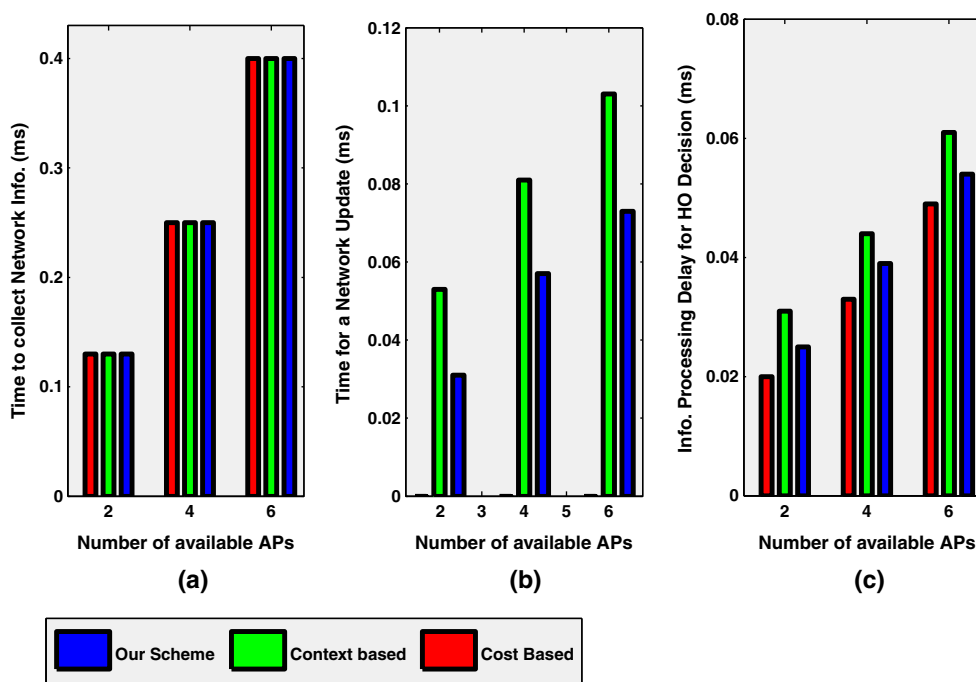


Fig. 13 Decision processing delay



Moreover, we can infer that the number of available target networks in the agent database have an impact on the decision processing delay as shown in the Fig. 13.⁴ The increase of the information processing time is mainly because of the information exchange between APs and MNs. In order to best decide about the time of handover and the network to choose, it is very important to process the collected information. Agents collect this information in raw form from the network and later process it after storing it in the agent knowledge base. Information is locally processed by the agents of MNs and APs.

Information processing and communication overhead (MN + AP) for our proposal and context aware scheme is depicted in the Eq. 8:

$$IPC_{\text{overhead}} = T_c + T_u + T_p \tag{8}$$

where IPC_{overhead} is the total information processing and communication overhead, T_c is the time to collect the information from the network, T_u is the time to get or send an update, and T_p is the time to process the information for a handover decision.

In our scheme, we use a time T_u in order to get the updates from the network about the availability of new networks and/or parameter updates of the existing networks. By doing so, agents residing in MNs and APs update their situated view. Of course, this produces

an additional overhead (0.073 ms when the number of APs = 6), but the importance of getting the fresh information cannot be ignored in this case. A handover decision based on fresh information would be more reliable than the one collected few seconds before.

Similarly, context-based approach uses this time T_u in order to download and deploy new services from the service deployment server. This process involves searching and fetching the appropriate service components and installation of those new services on the network nodes. As the installation procedure completes, a confirmation is sent to the service deployment server and the handover decision module. If the installation process fails, then all the above described process is repeated. This is one of the main reasons that our scheme performs better than the context-based approach.

Information processing and communication overhead for cost-based function is given in the Eq. 9:

$$IPC_{\text{overhead}} = T_c + T_p \tag{9}$$

In the cost function-based solution, decision process and communication overhead is less than that of our proposal and context based solution. It is because in cost-based proposal, information is collected from the network once, without sending or receiving any updates from the network, and a handover decision is made based on that information, unlike our's and context-based approach, where information is exchanged between the MN and the network. This is useful because,

⁴The results presented in this figure are related to Eqs. 8 and 9.

under the dynamic conditions of a network, the collected and exchanged information quite often changes over time.

In our scheme, the time to collect and process the information for the handover decision takes about 0.8 ms compared to the 1.3 ms of the context-based approach, as an average. Agents collect only the relevant information for the handover process in order to save time and simplify the handover process. This activity of the agents helps us in better network selection and handover anticipation, with some level of QoS.

6 Concluding remarks and future directions

In this paper, we presented an agent-based vertical handover scheme for heterogeneous wireless networks. We considered the parameters of RSS, bandwidth, user preferences, and cost for the handover and evaluated the performance of the approach using a selection function. We proposed to introduce intelligent agents in the mobile nodes and access points. Handovers are initiated by intelligent agents depending on the source of handover triggering events in the network. We proposed the use of knowledge plane that is controlled by intelligent agents. The proposed scheme is able to gather information from different network nodes and to use this information on mobile nodes for local processing and a better handover decisions. This is done with the help of cooperation among the distributed knowledge plane of agents residing in APs and MNs. The procedures to exchange, process information, and make an optimal decision are very fast.

The evaluation of the proposed approach has shown its feasibility. Specifically, we showed that the knowledge plane increased the efficiency of handovers. Also, the handover decision and the appropriate network selection on the basis of collected parameters during information gathering behavior are done for the better adaptation to the users and the application preferences. Our agent-based proposal is simple enough to be deployed in dynamic network conditions. Also, it shows noticeable improvements in QoS and overall performance of the handover. Intelligent agents used in the proposed scheme are adaptive. The proposed solution easily supports heterogeneity in technologies without any changes to the existing infrastructure and without any service disruption. However, we have obtained an overhead during the collection and cooperation phases of the proposed approach, but this overhead is very little as compared to the overall handover delay.

Our future work comprises of taking into account more QoS parameters to improve the performance of our proposed approach. We also intend to enrich our approach using some mathematical model for the admission control of the users in a network. Also, location management aspect will be taken into consideration in the future.

Acknowledgements This work has been partially supported by the SUN Project under the grant number and also partially supported by the Higher Education Commission of Pakistan.

References

1. Ahmed A, Boulahia LM, Gaiti D (2010) Cooperative agent based vertical handover scheme for heterogeneous networks. In: AICT'10: The 6th Adv. Intl. Conf. on Telecoms. IEEE, Los Alamitos, pp 410–415
2. Ahmed A, Rahim-Amoud R, Merghem-Boulahia L, Gaiti D (2010) An agent-based approach for vertical handover in heterogeneous wireless networks. In: AICSSA'10: The ACS/IEEE 8th international conference on computer systems and applications. IEEE Computer Society, Los Alamitos, pp 1–8
3. Ahmed T, Kyamakya K, Ludwig M (2006) Design and implementation of a context-aware decision algorithm for heterogeneous networks. In: SAC '06: proceedings of the 2006 ACM symposium on applied computing. NY, USA, pp 1134–1138
4. Akyildiz IF, Lee WY, Vuran MC, and Mohanty. S (2006) Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey. *Comput Netw* 50:2127–2159
5. Andrej V, Roman N, Gorazd K (2009) The impact of network topology on the performance of map selection algorithms. *Comput Netw* 54:1197–1209
6. Bellavista P, Corradi A, Foschini L (2007) Context-aware handoff middleware for transparent service continuity in wireless networks. *Pervasive Mob Comput* 3:439–466
7. onnin JM, Lassoued I, Hamouda ZB (2009) Automatic multi-interface management through profile handling. *Mob Netw Appl* 14:4–17
8. Bou Diab W, Tohme S (2009) End-to-end security and seamless handover solution for real-time communications over 3G networks. In: In Q2SWinet '09: proceedings of the 5th ACM symposium on QoS and security for wireless and mobile networks. ACM, New York, pp 13–22
9. Bullot T, Khatoun R, Hugues L, Gaiti D, Merghem-Boulahia L (2008) A situatedness-based knowledge plane for autonomic networking. *Int J Netw Manag* 18:171–193
10. Çeken C, Yarkan S, Arslan H (2010) Interference aware vertical handoff decision algorithm for quality of service support in wireless heterogeneous networks. *Comput Netw* 54:726–740
11. Chen WT, Shu YY (2005) Active application oriented vertical handoff in next-generation wireless networks. In: Wireless communications and networking conference, 2005 IEEE, vol 3. IEEE Computer Society, Los Alamitos, pp 1383–1388
12. Chu TC, Lin Y (2009) An extension to fuzzy MCDM. *Comput Math Appl* 57:445–454

13. Clark DD, Partridge C, Ramming JC, Wroclawski JT (2003) A knowledge plane for the internet. In: SIGCOMM '03: proceedings of the 2003 conference on applications, technologies, architectures, and protocols for computer communications. ACM, New York, pp 3–10
14. El-Sadek WF, Mikhail MN (2009) Universal mobility with global identity (UMGI) architecture. In: Proceedings of the 2009 international conference on wireless networks and information systems, WNIS '09. IEEE Computer Society, Los Alamitos, pp 389–394
15. Eng HO, Jamil YK (2009) Cooperative radio resource management framework for future IP-based multiple radio access technologies environment. *Comput Netw* 54:1083–1107
16. Ferber J (1999) Multi-agent system: an introduction to distributed artificial intelligence. Addison-Wesley Longman, Boston
17. Ferrus R, Sallent O, Agusti R (2010) Interworking in heterogeneous wireless networks: comprehensive framework and future trends. *Wireless Commun* 17:22–31
18. FIPA (2002) Foundation for Intelligent Physical Agents (FIPA) ACL Message Structure Specification <http://www.fipa.org/>. FIPA TC Communication, sc00061g standard edition
19. Gao X, Wu G (2009) System and method for supporting quality of service in vertical handovers between heterogeneous networks. US patent no. 7613457
20. Gruber TR (1993) A translation approach to portable ontology specifications. *Knowl Acquis* 5:199–220
21. Guerin F (2007) Applying game theory mechanisms in open agent systems with complete information. *J Auton Agents Multi-agent Syst* 15:109–146
22. Hasswa A, Nasser N, Hossanein H (2005) Generic vertical handoff decision function for heterogeneous wireless networks. In: 2nd IFIP international conference on wireless and optical communications networks. IEEE Computer Society, Los Alamitos, pp 239–243
23. Hong J, Suh E, Kim SJ (2009) Context-aware systems: a literature review and classification. *Expert Syst Appl* 36:8509–8522
24. Horrich S, Ben Jamaa S, Godlewski P (2007) Adaptive vertical mobility decision in heterogeneous networks. In: Third international conference on wireless and mobile communications, 2007. ICWMC '07, pp 44–49
25. Jain M (1999) Finite population cellular radio systems with directed retry. *Appl Math Model* 23:77–86
26. JASON H (2002) JDOM in the real world. [Jdom.org](http://www.jdom.org)
27. Jennings NR, Faratin P, Lomuscio AR, Parsons S, Wooldrige MJ, Sierra C (2001) Automated negotiation: prospects, methods and challenges. *Group Decis Negot* 10:199–215
28. Joe I, Kim W, Hong S (2008) A network selection algorithm considering power consumption in hybrid wireless networks. *IEICE Trans Commun* E91-B:314–317
29. Joe I, Kim W, Hong S (2008) A network selection algorithm considering power consumption in hybrid wireless networks. *IEICE Trans Commun* E91-B:314–317
30. Kebede G (2010) Knowledge management: an information science perspective. *Int J Inf Manag* 30:416–424
31. Lampropoulos G, Skianis C, Neves P (2010) Optimized fusion of heterogeneous wireless networks based on media-independent handover operations. *Wireless Commun* 17: 78–87
32. Leon SS, Kuldar T (2009) The art of agent-oriented modeling. MIT, Cambridge
33. Lin YB, Chlamtac I (2001) Wireless and mobile network architectures. Wiley, New York
34. Makaya C, Pierre S (2008) Adaptive handoff scheme for heterogeneous IP wireless networks. *Comput Commun* 31:2094–2108
35. The MathWorks, Inc (2002) MATLAB the language of technical computing, 13th edn. MathWorks, South Natick
36. Mohanty S, Akyildiz IF (2007) Performance analysis of handoff techniques based on mobile IP, TCP-migrate, and SIP. *IEEE Trans Mob Comput* 6:731–747
37. Mohanty S, Jiang X (2007) Performance analysis of a novel architecture to integrate heterogeneous wireless systems. *Comput Netw* 51:1095–1105
38. Nasser N, Guizani S, Al-Masri E (2007) Middleware vertical handoff manager: A neural network-based solution. In: IEEE International conference on communications, ICC '07, pp 5671–5676, 24–28
39. Passas N, Paskalis S, Kaloxylas A, Bader F, Narcisi R, Tsontsis E, Jahan AS, Aghvami H (2005) Enabling technologies for the 'always best connected' concept: research articles. *Wireless Commun Mob Comput* 5:175–191
40. Qingyang v, Abbas J (2008) A quality of service negotiation-based vertical handoff decision scheme in heterogeneous wireless systems. *Eur J Oper Res* 191:1059–1074
41. Quoc-Thinh NV, Nazim A, Yacine GD (2008) A user-centric and context-aware solution to interface management and access network selection in heterogeneous wireless environments. *Comput Netw* 52:3358–3372
42. Rizvi SS, Khan MA, Riasat A (2009) Active scanning: a better approach to reduce handover time at MAC layer for wireless networks. In: Computer, control and communication, IC4 2009, 2nd international conference, pp 1–4
43. Song Q, Jamalipour A (2008) A quality of service negotiation-based vertical handoff decision scheme in heterogeneous wireless systems. *Eur J Oper Res* 191:1059–1074
44. Stevens-Navarro E, Lin Y, Wong VWS (2008) An MDP-based vertical handoff decision algorithm for heterogeneous wireless networks. *IEEE Trans Veh Technol* 57(2):1243–1254
45. SUN (2010) Situated ubiquitous networks project. <http://projet-sun.it-sudparis.eu>
46. Tait RJ, Schaefer G, Hopgood AA (2008) Intensity-based image registration using multiple distributed agents. *Know-Based Syst* 21:256–264
47. TalebiFard P, Wong T, Leung VCM (2010) Access and service convergence over the mobile internet—a survey. *Comput Netw* 54:545–557
48. Wang XH, Zhang DQ, Gu T, Pung HK (2004) Ontology based context modeling and reasoning using OWL. In: Proc. of the 2nd IEEE annual conf. on pervasive computing and communications workshops, pp 18–22
49. Wei Q, Farkas K, Prehofer C, Mendes P, Plattner B (2006) Context-aware handover using active network technology. *Comput Netw* 50:2855–2872
50. Wei S, Qing AZ (2006) A novel decision strategy of vertical handoff in overlay wireless networks. In: Network computing and applications, IEEE international symposium. IEEE Computer Society, Los Alamitos, pp 227–230
51. Wu J (2008) Contract net protocol for coordination in multi-agent system. In: IITA '08: proceedings of the 2008 second international symposium on intelligent information technology application. IEEE Computer Society, Los Alamitos, pp 1052–1058

52. Xing L, Lan Z (2009) A method based on iterative combinatorial auction mechanism for resource allocation in grid multi-agent systems. In: Proceedings of the 2009 international conference on intelligent human–machine systems and cybernetics. IEEE Computer Society, Los Alamitos, pp 36–39
53. Yifei W, Xiaowei L, Mei S, Junde S (2008) Cooperation radio resource management and adaptive vertical handover in heterogeneous wireless networks. In: International conference on natural computation, vol 5. IEEE Computer Society, Los Alamitos, pp 197–201
54. Ylianttila M, Mäkelä J, Pahlavan K (2005) Analysis of handoff in a location-aware vertical multi-access network. *Comput Netw* 47:185–201
55. Zahran AH, Liang B, Saleh A (2006) Signal threshold adaptation for vertical handoff in heterogeneous wireless networks. *Mob Netw Appl* 11:625–640