WIRELESS COMMUNICATIONS AND MOBILE COMPUTING
Wirel. Commun. Mob. Comput. (2007)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/wcm.539

An agent-based perspective to handover management in 4G networks

Vassilis E. Zafeiris*,† and E. A. Giakoumakis
Department of Informatics, Athens University of Economics and Business, Athens 10434, Greece

Summary

In 3G or earlier generation networks handovers are usually initiated and decided by the base station, on the basis of measurements of RSS or SNR received by the terminal, cell congestion, terminal speed etc. In 4G, due to the diversity of available radio access services, additional factors, for example user profile, application requirements, and terminal device capabilities, need also to be taken into account. We propose an agent-based architecture that determines the timing and target network for handovers in a 4G network setting. The capabilities of the architecture are provided as a value-added service on top of network operators' wireless access infrastructure. Network selection spans both wireless access and core routing services and is performed by user agents executing in the network side. In order to deal with trust issues we study the integration of the regulatory authority in the architecture. A performance study of the architecture's impact on handover latency is provided through a simulation system. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: handover management; 4G networks; software agents; bearer negotiation

1. Introduction

The fourth generation (or beyond 3G-B3G) of mobile communications systems is orientated towards the integration of available wireless access technologies. The envisaged architecture of a 4G network comprises a variety of wireless access (WLAN, Bluetooth etc.) and cellular 3G or 4G networks interfacing with access routers to a common core IP network that serves their interconnection and integration with earlier generation networks (PSTN, ISDN, 2G/2.5G etc.) through appropriate gateway routers [1]. Handover management extends its scope in 4G targeting seamless mobility across cells: (a) of the same or different technology (horizontal vs. vertical handover), (b) operated by the same or different providers (intra-domain vs. inter-domain handover). Thus, additional factors need to be taken into account during handover initiation and decision, such as user's cost tolerance and contractual constraints, applications' requirements and priority, terminal device status (battery status, available radio interfaces etc.) [2]. In this context, the role of handover is being enhanced from maintaining connectivity to optimizing connectivity.

The sources of handover triggering events, as well as of the information required for handover decision, do not lie exclusively on the network or link layers. Instead, given that application and user-related factors will be taken into account, higher layers, such as the application layer, would also have substantial contri-

*Correspondence to: Vassilis E. Zafeiris, Department of Informatics, Athens University of Economics and Business, 76 Patission Str., Athens 10434, Greece.
†E-mail: bzafiris@aueb.gr

bution. Thus, an application layer approach is required for a handover initiation and decision solution, as it is also proposed in Reference [3] where the requirements for an always best connected (ABC) service are set.

From a user viewpoint, a technical solution for handover initiation and decision enables efficient connectivity in a context of multiple wireless access networks. However, from a network operator's perspective such a solution acts as a market enabler that allows clients to either join or leave its network depending on their current utility. A trustworthy implementation of this capability cannot be offered by a single network provider. The reason is that it has no incentives to provide reliable and in-time information regarding available wireless networks and consequently let its customers utilize third-party services. A viable solution should: (i) incorporate various wireless operators, (ii) support market competition through easy integration of new entrants, (iii) adopt a common, unambiguous information schema for interoperability of the exchanged information (e.g., descriptions of network capabilities, so as to enable effective decision making), (iv) build on a commonly accepted model of trust relationships so as to be relied upon by users and network operators.

We propose an agent-based architecture for handover management in 4G with main focus on the initiation and decision phases of the handover mechanism. We define agent types and their deployment in user terminals and access networks, and describe their collaboration in order to determine the timing and target of handovers. We assume 3G cellular networks and IEEE 802.11x WLANs and study the integration of our architecture into their network infrastructure. Finally, an initial study of performance implications of the proposed architecture is provided based on results from a simulation system. The rest of the paper is organized as follows: Section 2 compares our approach with related work and Section 3 analyzes the business environment where the architecture will be integrated. An overview of the architecture along with the merits of an agent-based approach in handover management is presented in Section 4. Section 5 focuses on agent collaboration during handover management. A performance evaluation of the proposed architecture is presented in Section 6 and Section 7 concludes the paper with future work.

2. Related Work

We can identify two main approaches in the design of a 4G handover management solution with respect to the distribution of its components (a) exclusively in end-hosts or (b) both in end-hosts and nodes in the network side. As regards the first approach, a middleware architecture is proposed in Reference [4] for supporting the development of mobility- and context-aware applications. Specifically, an API is provided for receiving mobility-related events and performing handovers among Bluetooth and WLAN access points (APs). However, handover management is handled in an application specific manner while a mechanism for resolving possible conflicts is not provided. In Reference [5] an end-to-end middleware is proposed that provides applications with transparency from changes in connectivity through a channel abstraction. Each channel is associated with a transport layer connection and is constantly monitored and managed by a channel management agent (CMA) executing in each end-host. An advantage of such architectures is their support of handover management without modifying the existing networking infrastructure. However, legacy applications need to be rewritten in order to utilize their features. Moreover, the overhead related to handover management is distributed to the usually resource constrained mobile terminals (MTs).

The proposed architecture can be classified to the approaches that introduce intelligence in the network side for support of handover management in a 4G setting. A broad set of solutions belong to this category and some representatives will be compared against our architecture on a number of design decisions: (a) support for integration of different domains, (b) context information interoperability, and (c) degree of distribution of the handover decision between the MT and the network.

In Reference [6] the concept of network interoperating agent (NIA) is introduced that handles authentication, authorization, and accounting (AAA) and inter-domain mobility management across various wireless providers. NIA is managed by a third-party that maintains SLAs with the network operators. Wireless networks integrate with the NIA through interworking gateways (IG). A similar approach is adopted in our architecture where the role of NIA is played by a multi-access provider (MAP). Moreover, MAP offers information services such as wireless network presence information, access to user and device profiles etc.

Information interoperability regarding the status and provided services of each network and other context information is important for handover initiation and decision. In Reference [7] a context-aware framework for handovers is proposed that introduces repositories for collection of context information and

DOI: 10.1002/wcm

an execution platform for the dynamic deployment and execution of context handling components. PROTON [8] specifies an end-host architecture for the collection, classification, and aggregation of context information. However, the collected context is restricted by the MT's context perceiving capabilities while third-party information is not utilized. A comprehensive framework for exchange and management of context is proposed in Reference [9]. In this approach, shared ontologies play a key role in the representation and exchange of context information. This feature is also inherent in an agent-based approach enhancing thus information interoperability. Moreover, well-established agent communication languages (ACLs) [10] provide a messaging framework for context exchange among system components. A more elaborate discussion on these merits of software agents is included in Section 4.1.

Many approaches for handover decision in a 4G setting (e.g., [11–13]), delegate the execution of the decision logic to the MT. Although this configuration ensures fast response to handover initiation events, it is not always efficient in terms of consumption of the MT's usually restricted resources (e.g., power, CPU etc.). In Reference [6] a two-level handover decision process is proposed involving both the NIA and the MTs. Initially, the MT performs network selection on the basis of various factors while in a second level its decision is adjusted by the NIA that focuses on global load balancing. PROTON [8] uses policies for handover decision that are compiled into automata, in appropriate network nodes, and sent to the MT in order to guide its behavior. We propose personalized handover decision delegated to agents/user representatives that execute in the network side. In order to minimize the latency in its communication with the MT, the user representative migrates and executes in appropriate nodes in the MT's current wireless access network. As access network selection does not guarantee efficient end-to-end service provision we incorporate in the handover decision process a second level selection of the core network that will transit user traffic to the correspondent nodes.

3. Business Actors and Their Relationships

The architecture focuses on inter-domain handovers, as well as on vertical intra-domain handovers. We assume that horizontal handovers in the bounds of a domain are transparently handled by the network's mobility management mechanism. In the case of a single network provider, available radio access networks (RANs) may be interconnected through a common core network. However, we study the more general case where RANs provide access to different core networks, interconnected either directly or through the public internet. Two types of network providers are assumed (a) core network operators that manage zero or more possibly heterogeneous RANs attached to their core network, (b) RAN operators that manage one or more RANs attached to the networks of core operators.

End-users gain access to the services of various wireless providers through a single subscription. This is enabled by a business entity that maintains roaming agreements with the network operators and offers a value-added service on top of their network infrastructure. Its role is analogous to that of an ABC service provider, as defined in Reference [3]. However, it will be referred to as MAP as we focus on a subset of the ABC service capabilities. MAP subscribers are not required to maintain contracts with network providers-partners of the MAP. The MAP is responsible for their authentication and bills them by aggregating their respective charges from the various network providers. In addition, it supports handover management by providing a platform where (a) network operators publish their services and (b) user terminals discover network services and select the most appropriate on the basis of user, device and application-related constraints.

However, MAP cannot guarantee the accuracy of information published through its infrastructure. For instance, a network provider may publish service descriptions that do not correspond to their actual, probably inferior, performance. In order to discourage providers from abusing the infrastructure, the presence of the regulatory authority is deemed important. The regulator is trusted by all actors and enhances the credibility of offered services. Its role will be further analyzed in Section 4.3.4.

4. Architecture Overview

4.1. An Agent-Based Approach

The use of software agents facilitates the deployment, maintenance, and management of the system. The capability of a network node to host the execution of system components depends on the presence of an agent platform. Once platforms are deployed, agents can populate them either by self-motivated mobility

DOI: 10.1002/wcm

or remote instantiation with the use of appropriate management software. In the same way new software versions can be deployed. Management is enabled by management capabilities provided by the platform through a standard interface (FIPA compatible agent platforms) [14].

Interoperability of system components is a critical design issue as they will be implemented by different parties (MAP, regulator, network providers, other application service providers) with possibly different perspectives on the problem domain. A straightforward approach for interoperability is based on the definition of APIs and messaging protocols. Such specifications are implementation specific, focus on the syntax of the messages and may have different interpretation by the various providers. Agent interoperability is based on the exchange of messages expressed in one of the widely accepted ACLs, FIPA ACL or KQML [10]. An ACL message encapsulates the communication payload and describes it in a domain independent way with a predefined set of attributes. The payload is expressed in a content language (e.g., KIF, RDF) with the use of vocabulary from shared ontologies. Consensus, thus, among providers is reached at a higher, conceptual level that ensures unambiguous interaction.

4.2. MAP Network Architecture

The specification of the proposed architecture, as well as a study of its integration in the infrastructure of current wireless networks, will be presented with reference to a 4G heterogeneous network setting. For the sake of simplicity three types of wireless providers will be considered: (a) *3GPP* (*GSM/UMTS*) *operators*, offering packet switched as well as circuit switched services, (b) *WLAN providers*, operating IEEE 802.11x WLANs for internet access in public hot spots, and (c) *I-WLAN providers*.

An I-WLAN provider manages a WLAN RAN that interworks with the core networks of one or more 3GPP systems, henceforth referred to as 3GPP public land mobile networks (PLMNs) or simply PLMNs. A 3GPP-WLAN interworking architecture is being specified by 3GPP [15] with purpose of (i) enabling WLAN public access to subscribers of PLMN operators and (ii) enabling WLAN access to IP-based services of the 3GPP PS domain. An I-WLAN allows a MT to be authenticated by its home PLMN (HPLMN) and use packet switched services of the HPLMN, other visited PLMNs (VPLMNs) or directly connect to the public internet via the I-WLAN network depending on user subscription terms.

Figure 1 depicts a simple 4G setting that will be used as a case study for the description of various operational scenarios of our approach. In this setting, the MAP cooperates with a WLAN, a UMTS (UMTS$_1$) and an I-WLAN operator. The latter provides access to core services of UMTS$_1$, in addition to basic internet access. Moreover, we assume that the coverage areas of all wireless access networks intersect in the user's geographical region. A contract with the MAP allows a user to access services offered by all MAP partners. Each wireless provider and the regulator incorporate

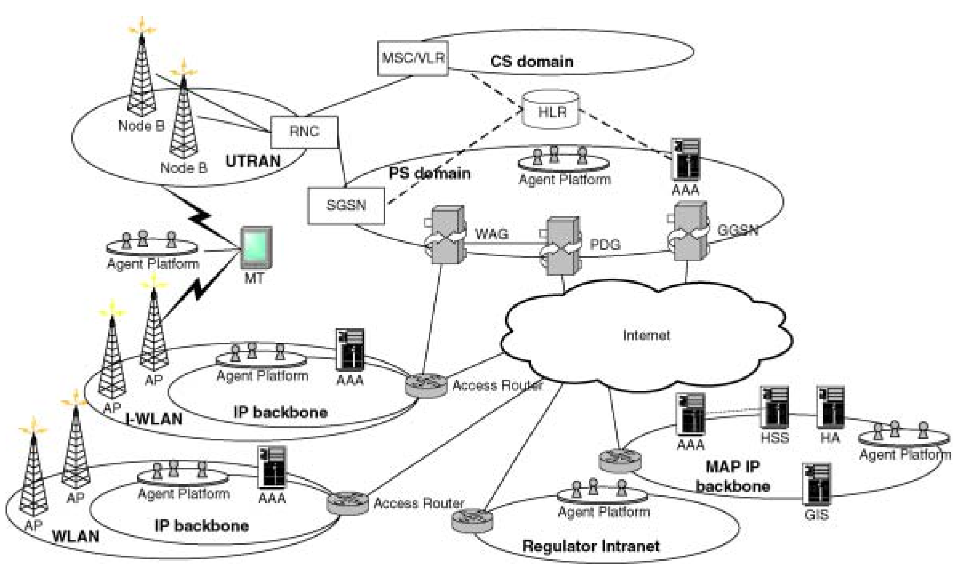


Fig. 1. MAP network architecture.

DOI: 10.1002/wcm

an agent-platform in their core networks in order to participate in the architecture. Especially for operators of networks with broad geographical coverage, the management of more than one instances of an agent-platform may be required for efficiency and load distribution purposes. Each instance can serve the users of a geographical region, for example, a city or a more restricted area.

The MAP's network infrastructure is also outlined in Figure 1. It comprises an IP backbone that interconnects a set of core elements. These core elements, accessible by terminals and wireless networks through the public internet, include: (a) *a geographical database* (*GIS*) for location-based retrieval of information related to the type and coverage of its partner-networks, (b) *a mobile IP home agent* (*HA*) for macromobility support, (c) *an agent platform* hosting agents that wrap services offered by other core elements and make them accessible to agents distributed in user terminals and access networks, (d) *a home subscriber subsystem* (*HSS*) accessible by HSSs of 3GPP networks for authentication and authorization purposes, call routing and user profile retrieval, and (e) *an AAA server* that interfaces with respective AAA servers of network providers using RADIUS or Diameter. It is referenced by them for user authentication and authorization, while it concentrates their charging records for billing purposes. The AAA server interfaces and acts as a proxy to the HSS that is the primary source of authentication, profile, and authorization information.

As regards user terminal equipment, a multi-mode MT is assumed, capable of connecting to both IEEE802.11x and UMTS networks. A basic requirement for a MT is the integration of a UICC smart card with USIM/SIM modules for authentication with UMTS/GSM networks respectively. UICC smart cards are owned by the MAP and include each subscriber's international mobile subscriber identity (IMSI) that is also issued by the MAP.

4.3. Agent Types and Functionality

Figure 2 depicts the architecture's agent types, along with the agent platforms that host their execution. Agents' different shading is indicative of the authority they represent. Platforms comply with the FIPA agent management reference model [14] that specifies three basic logical components: (i) agent management system (AMS) agent, (ii) directory facilitator (DF) agent, and (iii) the message transport service (MTS), the default communication method between agents on different platforms. AMS is a mandatory component

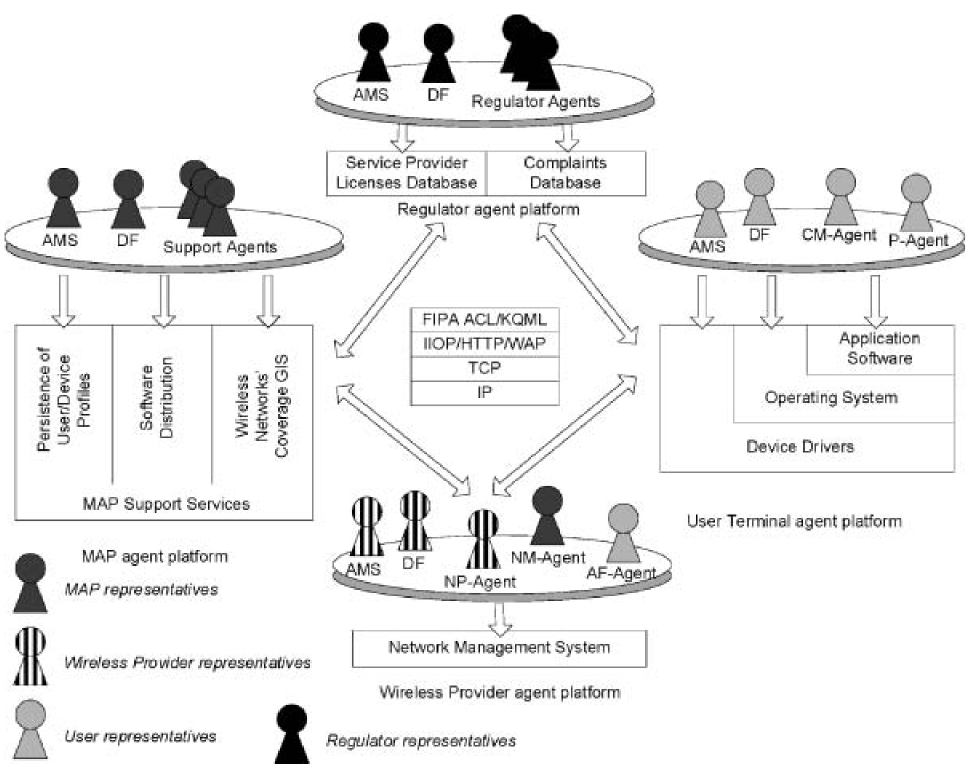


Fig. 2. Agents comprising the architecture.

DOI: 10.1002/wcm

that provides agent registration, life cycle control, and white pages services, while DF is an optional component that provides yellow pages services to the agents of a platform. The ownership of AMS and DF correlates with the administration of the platform. Agent communication is based on ACL messages expressed in either FIPA ACL or KQML [10]. Message transport is carried out over TCP/IP with use of HTTP, IIOP, or WAP etc.

4.3.1. Multi-access provider agents

MAP agents wrap services offered by the MAP's core elements and provide them via an ACL interface to other agent types executing in wireless networks and user terminals. Among the main support services offered by the MAP agents are: (i) software distribution, where latest versions of drivers and software libraries are distributed on demand in order to enhance the utilization of the terminals' network interface(s), (ii) management of user profile information, that is transferred to authorized agents in order to reason and act upon it, (iii) location-based retrieval of network information. The geographical coverage of each RAN as well as information concerning its type and offered services are stored in MAP's GIS. MAP agents interface with the GIS and serve requests on (a) the wireless networks of a certain type that cover a geographical location and (b) the wireless networks that have overlapping coverage and the same type with a given network.

4.3.2. Wireless provider agents

A wireless providers platform executes agents that represent all business actors except for the regulator. The platform includes one Network Provider agent (NP-agent), one Network Monitor agent (NM-agent) and several Access Facilitator agents (AF-agents) representing the wireless provider, the MAP and currently connected users respectively. *NP-agent* provides an ACL interface for controlled access to network management information of the current network. The information exposed by the NP-agent includes descriptions of information transfer services provided by its network characterized by various attributes such as QoS, cost etc. *NM-agent* aggregates information regarding access networks that intersect with the current network's coverage area. This information is retrieved through subscription to respective NP-agents. Each NM-agent serves numerous *AF-agents*, corresponding to the current network's users. AF-agent is a user proxy responsible for handover initiation and decision and subscribes to NM-agent for receiving information on networks with presence in the MT's current location. It incorporates user and terminal profile information required for its decision-making. In order to minimize network latency in its communication with the terminal's agents, AF-agent migrates and executes each time in the platform that corresponds to the current access network.

4.3.3. Terminal device agents

Each terminal device utilizing the MAP services, hosts two agent types that are user representatives, Profile-agent (P-agent) and Connection Manager-agent (CM-agent). *P-agent*'s role is to communicate the perceived QoS, user preferences and application requirements to AF-agent in order to make informed decisions on the target and timing of handovers. *CM-agent* is responsible for successful execution of handovers initiated by the user's corresponding AF-agent. Handover execution is preceded by a procedure that determines the terminal's capability (in terms of software requirements, RSS) of accessing the selected network. Note that the terminal's hardware configuration is taken into account during handover decision by AF-agent that informs CM-agent of the appropriate driver versions and protocol implementations that the terminal should support in order to connect effectively to the specified network. CM-agent checks the terminal's software configuration and reports any deficiencies to a MAP agent in order to download updated versions.

4.3.4. Regulator agents

The software agents that execute on the regulator's platform wrap appropriate databases, and either update or make their contents available to other agents. Two types of databases are considered: (a) a database DB_l with service licenses granted to the various network operators, (b) a customer complaints database DB_c. On the basis of DB_l contents, regulator agents provide information regarding network providers and the services they are licensed to offer. Such information is requested by NM-agents whenever a new type of service is retrieved from a NP-agent. Moreover regulator agents, update DB_c on the basis of notifications related to misleading service descriptions published by providers. Such notifications are sent by P-agents whenever the QoS of a service declines

DOI: 10.1002/wcm

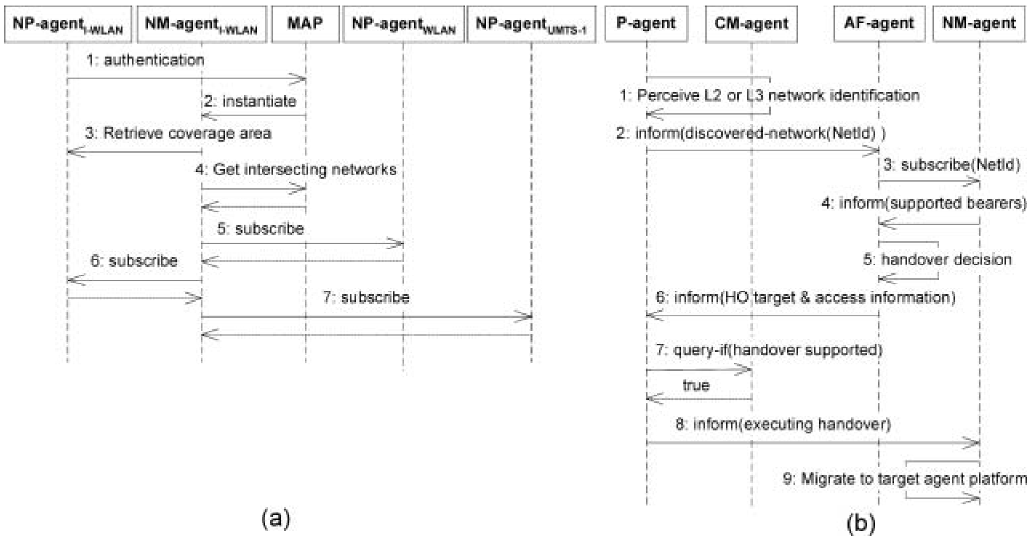


Fig. 3. (a) Bootstrapping of network provider's platform, (b) handover initiation and decision.

significantly from that advertised. These notifications are analyzed by the regulator and, if necessary, further investigation is conducted. DB_c is also updated with notifications regarding locations with high radio interference. P-agents send such information whenever high interference is perceived by the MT. The regulator can check areas with high concentration of notifications for operation of unlicensed antennae.

5. Handover Mechanism

5.1. Network Provider Platform Bootstrap

Bootstrapping of a network provider's platform involves authentication of the platform's NP-agent with the MAP and download of the latest version of the NM-agent software. Figure 3a presents the message exchange during I-WLAN's agent platform initialization. After its instantiation (2), NM-agent requests from the local NP-agent geographical information regarding the coverage area that this platform serves (3). On the basis of such information NM-agent retrieves from the MAP the addresses of NP-agent$_{\mathrm{WLAN}}$ and NP-agent$_{\mathrm{UMTS-1}}$ that correspond to networks that intersect with the given coverage area (4). Finally, NM-agent retrieves bearer service descriptions from all relevant NP-agents through subscription (5, 6, 7). The adoption of a subscription model allows only modified service descriptions to be propagated to NM-agents, thus minimizing network traffic that could be generated by periodically polling NP-agents for their offered services.

5.2. Handover Initiation

Handover initiation is an iterative task that triggers handovers whenever the terminal's current connection(s) do not meet the applications' requirements in terms of bandwidth, QoS, security etc. P-agent perceives and makes available to AF-agent a series of handover initiation events such as: (i) link quality degradation on an active radio interface, (ii) change to connectivity requirements due to the start or termination of an application, (iii) change to the status or configuration of the terminal device, for example, low battery level, availability or unavailability of a radio interface, (iv) special core network capabilities required by an application, for instance sending a MMS message requires access to the core of a 2.5/3G network, (v) discovery of a new access network. AF-agent also perceives handover initiation events from NM-agent whenever the status of a subscribed network (e.g., congestion level) changes.

Agent interaction in a scenario of new wireless network discovery is presented in Figure 3b. We assume that the MT is initially connected to the UMTS network. As it approaches a hot spot covered by the WLAN provider, its WLAN interface starts receiving IEEE802.11 beacon frames from the WLAN's APs. Each beacon frame incorporates a SSID information element that identifies the WLAN provider. This information is extracted and made available to P-agent through an appropriate event (1). However, besides network identification, beacon frames do not include other information attributes that may assist the MT's handover decision. Such information could

DOI: 10.1002/wcm

be retrieved after associating with the AP, a relative time consuming task that may prove useless, if finally the WLAN is not selected. In our approach discovery of bearer service information is delegated to the AF-agent. P-agent notifies AF-agent through its current UMTS connection on the perceived event with an initiate handover message that includes the discovered network's SSID (2).

Information on the types of bearer services offered by each wireless network is retrieved by AF-agent through subscription to the local NM-agent (3, 4). On subscription initiation, AF-agent provides to NM-agent an identification of the network that is interested on receiving information about. AF-agent's subscriptions span only wireless networks that are available in the MT's current location. These networks are discovered by the MT terminal on the basis of L2/L3 information received by its radio interfaces. As an alternative, MT's current location, retrieved by a GPS receiver, could be utilized for network discovery. AF-agent *decides* on the most appropriate target for handover (5) and forwards to P-agent all relevant information for handover execution (network type and registration information, protocol versions, authentication type etc.) (6). P-agent requests the execution of the handover from CM-agent (7) that assesses the terminal's capability of connecting to the proposed network. In the positive case P-agent notifies AF-agent (8) that migrates to the target network's agent platform in order to serve its principal with the minimum network latency (9).

A more detailed description of AF-agent's migration procedure is depicted in Figure 4. Continuing the above scenario, AF-agent decides to migrate to the agent platform of WLAN, that is the MT's target for handover execution (1). Before migration, AF-agent unsubscribes from NM-agent$_{UMTS-1}$ (2) and informs P-agent of the new platform that will host its execution (3). Agent migration is requested by the current platform's AMS (AMS$_{UMTS-1}$) (4) that transfers agent state and code to AMS$_{WLAN}$ (5). The latter restores AF-agent's execution (6) and informs the former on the migration success (7). After resuming its execution on the new platform, AF-agent waits a contact from P-agent in order to continue its normal operation (8). P-agent communicates with AF-agent once handover execution to the new network is complete. A key exchange takes place among agents in order to prove their identity (9) and finally AF-agent subscribes to NM$_{WLAN}$ (10).

5.3. Handover Decision

In our case study scenario we assumed a multi-mode MT located in an area served by WLAN, UMTS, and I-WLAN network operators. In this setting, handover decision extends its scope to the selection of both wireless access and core networks. A basic requirement for optimum selection of access and core networks is a consistent and commonly adopted data model for describing data transfer services. Our approach for the description of the various services is based on the categorization and attributes used in Reference [16] for the characterization of telecommunication services in a 3GPP PLMN. Specifically, telecommunication services are classified into *bearer services* that provide information transfer between APs of a network, and *teleservices* that provide a complete capability. The service descriptions published by network providers,

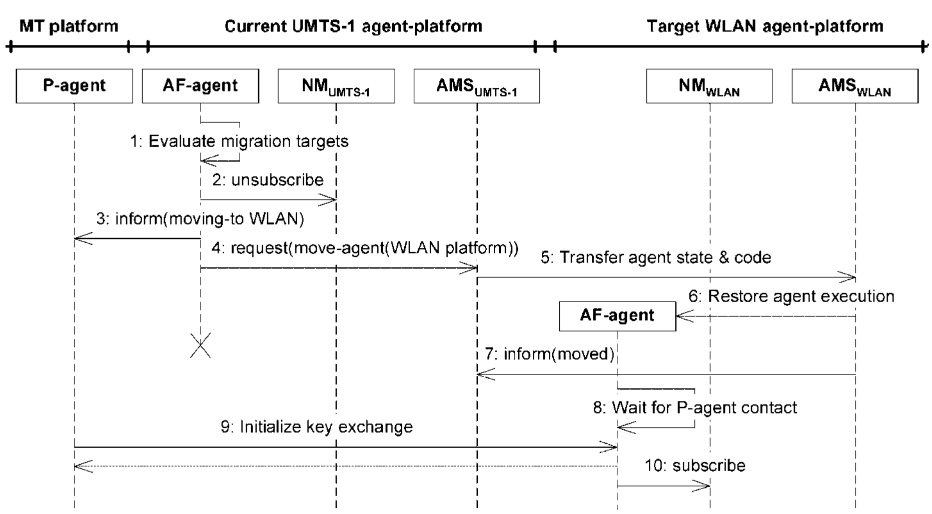


Fig. 4. AF-agent migration from UMTS-1 to WLAN agent platform.

DOI: 10.1002/wcm

Table I. Bearer service description.

Attribute name	Values
Information Transfer attributes	
Connection mode	Connection oriented, connectionless
Traffic type	Constant or variable bit rate
QoS	Instance of qos class
Communication configuration	Point to point (PTP), Point to multipoint (PTM), broadcast
Symmetry	Unidirectional, bidirectional symmetric, bidirectional asymmetric
Access attributes	
Access interface	Instance of protocol endpoint class
Interworking attributes	
Type of terminating network	PSTN, ISDN, PSPDN, PDN*, PLMN, direct internet access
Accessible core network types	PSTN, ISDN, PSPDN, PDN, PLMN
Core access interface	Instances of Protocol Endpoint class
General attributes	
Cost	Instance of Cost Structure class

*Packet Switched Public Data Network, Public Data Network.

through NP-agents, belong to the bearer service category. As handover execution, in most approaches, is handled in the data link (L2) or network layer (L3), handover decision involves selection of L2 or L3 bearer services offered by different networks. Bearer services are described in Reference [16] through a set of attributes (see Table I for a subset of them).

The structure of a bearer service description is depicted in the class diagram of Figure 5. The protocol that is used by a bearer service for information transfer as well as a description of its endpoint and its access requirements (e.g., authentication method) are provided by the 'access interface' service attribute. Its range is set to instances of the protocol endpoint (PE) class that has the same semantics as the respective class defined in the CIM (Common Information Model) Network Model specified by DMTF [17]. The PE class identifies the address or location where the bearer service is available and incorporates service configuration information (e.g., supported packet size, negotiable QoS attributes, authentication methods etc.). Moreover, the class characterizes the OSI stack layer that the service belongs to. For instance, the LAN Endpoint and IP PE classes, defined in Reference [17], extend PE and describe L2 and L3 bearer service endpoints respectively.

For our case study scenario, a set of additional protocol endpoint subclasses need to be introduced for the description of available bearer services. These classes, depicted in Figure 5, are: (a) I-WLAN Endpoint that extends the WLAN Endpoint [17] by appending the 'EAP Method' attribute that specifies the

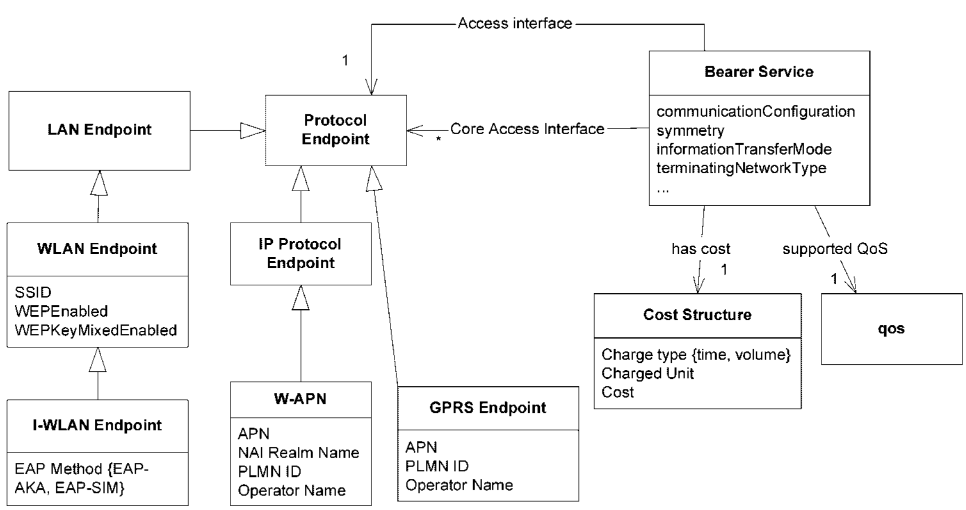


Fig. 5. Class model of bearer service description.

DOI: 10.1002/wcm

protocols that are supported for 3GPP authentication, (b) W-APN (Wireless Access Point Name) that identifies a L3 AP to the PS domain of a 3GPP network interworking with a WLAN (note that instances of W-APN class characterize the 'access interface' of 3GPP core bearer services), and (c) GPRS Endpoint that identifies a packet data network in a PLMN.

The types of core networks that are accessible through a given RAN are specified in the 'accessible core network types' attribute, while the endpoints of their respective core bearer services are given in the 'core access interface' attribute. The latter is a multi-value attribute and ranges over instances of the PE class. The bearer service descriptions that correspond to the aforementioned core access interfaces should also be provided by the NP-agent of the wireless network that interworks with them and made available to NM-agents. Finally, the cost of using a certain bearer service is specified in the 'cost' attribute that ranges over instances of the 'cost structure' class.

On the basis of bearer descriptions retrieved from NM-agent, AF-agent creates the best combination for user access. At first a selection of access bearers (AB) (i.e., bearer services with an L2 wireless PE) takes place that are usable by the MT's radio interfaces. Next, core bearers (CB) are retrieved that are accessible through the selected AB, that is, bearers that their 'access interface' matches the 'core access interface' of at least one of the selected AB. As a result a set of pairs (AB_i, CB_i) is created. The value of 'Type of terminating network' attribute is then checked in order to ensure that the type of termination required by MT applications is provided by selected bearers. This filtering does not apply to AB that interwork with at least one core bearer. The combined cost and QoS of each pair (AB_i, CB_i) is then calculated and used as input to one of the methods proposed in References [12] or [13].

5.4. Mobile Terminal's Initial Access and Authentication

During initial log on of a disconnected MT, network selection aided by AF-agent, is not possible. Its instantiation takes place after successful authentication through a network provider supporting the MAP services. In this case network selection is based on a predefined priority list of wide coverage providers. After user authentication, P-agent and CM-agent are instantiated on the MT's agent platform while an AF-agent instance is created in the MAP agent platform and moves to the platform of the user's current network. The terminal authenticates with the MAP at initial log on and prior to each inter-domain handover. Authentication is carried out by the terminal's UICC that incorporates appropriate software modules such as USIM, for authentication with UMTS over UTRAN, and EAP-AKA that allows the execution of the UMTS AKA protocol over a WLAN access network [18]. Authentication information is forwarded through the current wireless provider's AAA or HSS to the corresponding elements of the MAP that is responsible for user authentication and authorization.

5.5. Handover Execution

Our approach is transparent to handover execution protocols and various mechanisms, either network, transport, or application layer may be employed in order to ensure flow continuity across different radio access technologies or domains. As regarding the incorporation of Mobile IPv6 in the architecture, MIPv6 HA is situated, as described in Section 4.2, in the MAP's domain. MTs authenticated with the MAP perform binding updates to the HA with their current care-of-addresses in order to be reachable by correspondent nodes through their public home address. Routing optimization should be applied where possible (it depends on the capabilities of the correspondent node) so as to avoid transforming the MAP domain to a traffic bottleneck.

Other aspects of MIPv6, such as the multicast of router advertisements for mobility detection can be incorporated in the network discovery mechanism of the proposed architecture. Router advertisements are used in IPv6 and MIPv6 for stating the availability of a router. A MT receives router advertisements and forwards them to AF-agent in order to subscribe to NM-agent to network descriptions that correspond to the discovered access router. The new network descriptions are then utilized by AF-agent in its decision process. On the other hand, when router advertisements of the current network are no longer received by the MT, AF-agent is notified in order to unsubscribe from the respective network descriptions.

6. Performance Evaluation

In Reference [19] a partitioning of the handover latency is proposed, where the total latency T_{HO}, is broken into three main components, namely detection period T_d, address configuration interval T_c and network registration time T_r, $T_{HO} = T_d + T_c + T_r$. As the

DOI: 10.1002/wcm

Table II. Simulation parameters.

Pythagor simulation parameters		Network load generation parameters	
AP data rate	11 Mbps	Call duration	Exponential $\mu = 0.0083\,s^{-1}$
VoIP packet length	200B	Call arrival	Poisson $\lambda = 0.\ 245\,s^{-1}$
VoIP packet rate*	25 pkt/s	AP capacity	C = 37 VoIP sessions
VoIP bit rate*	80 Kbps	Call blocking prob.	30%

*One way packet rate. A total of 80 Kbps is required for a single call, 40 Kbps in each direction.

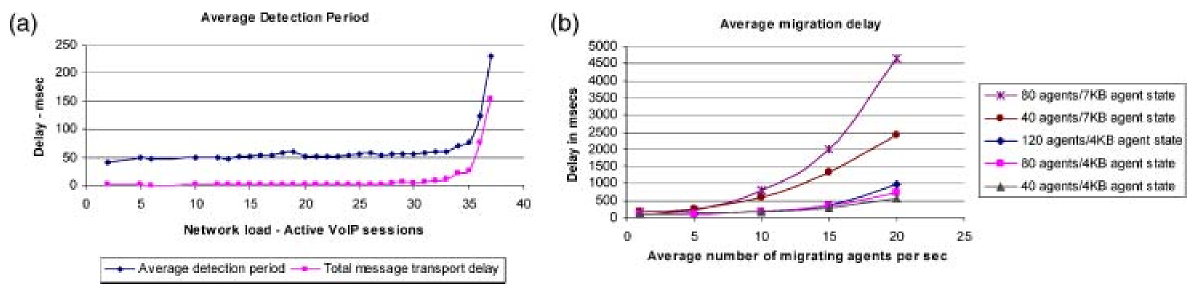


Fig. 6. (a) Average detection period on various network loads and (b) average migration delay.

main focus of our approach is on the selection of the timing and target for handover execution, the latency introduced to the handover procedure due its operation corresponds to the detection period component of the above partitioning.

A simulation system has been implemented in order to estimate T_d and study the performance of agent migration. System implementation is based on JADE 3.4,‡ a widely used framework for developing agent-based applications. The simulation setting that has been deployed in order to estimate T_d comprises two agent platforms P_1 and P_2 that serve the users of two IEEE 802.11g AP, AP_1 and AP_2, respectively. Each platform is executed on a Pentium 4 3.2GHz/1GB RAM workstation and hosts the set of agents described in Section 4.3.2. As concerning user representative agents, an instance of AF-agent is executed for each MT that is associated with an AP regardless of having or not an active data session with it. A third agent platform P_0 hosts P-agents that correspond one-to-one to AF-agents executing in P_1 and P_2. Each P-agent generates VoIP call establishment requests *REQ* that trigger handoff decision to their corresponding AF-agents. AF-agents incorporate an implementation of the AHP-GRA network selection algorithm [12] that is executed upon each request. The result *RES* is sent back to the P-agent and the round-trip time corresponds to T_d.

The VoIP service is selected as a type of traffic in order to study performance results in a setting with strict QoS requirements. The call duration and inter-arrival times are exponentially distributed. Call servicing and thus the network load is simulated by NP-agents. The average packet delay under various network loads is estimated by a WLAN simulator, Pythagor.§ Pythagor is also used for an estimate of the AP capacity, in terms of VoIP sessions. As the VoIP service requires delay values to be less than 150 ms in each direction, capacity threshold is set to a number of sessions that satisfies this requirement. For the simulation setting included in Table II the AP capacity is 37 sessions.

The average T_d has been estimated for 200 AF-agents executing in each of P_1, P_2. A high network utilization is assumed with call blocking probability 30%, as it results from Erlang B loss formula given the values of λ, μ, and C that corresponds to the number of available servers:‖ Figure 6a presents the average T_d as it evolves with the number of active sessions on the AP. The message transport delay (T_m) corresponds to the latency introduced by the wireless access network

‡ Java Agent Development framework, 2007 URL: http://jade.cselt.it

§ Pythagor WLAN Simulator, 2007. URL: http://www.icsd.aegean.gr/telecom/Pythagor/pyth_main.htm

‖ The system can be modelled as a M/M/m/m queuing system.

DOI: 10.1002/wcm

for the transport of both *REQ* and *RES*. The average message size for REQ and RES is 300B and 1700B, respectively. The diagram shows that T_d is equal to T_m plus an overhead of about 50 ms due to agent collaboration. Thus, T_d remains in reasonable levels, considering that in case of WLAN/GPRS vertical handovers it has values in the order of 100 ms [19]. In our future work, T_d will be studied under larger scale deployments.

As JADE 3.4 does not support inter-platform mobility, a third-party service has been employed for the implementation of agent migration in the simulation system. The inter-platform mobility service (IPMS)[¶] integrates with the JADE agent platform and utilizes the platform's MTS for implementing agent migration between different JADE platforms. In IPMS, the platform's AMS packages the migrating agent's code and state in an ACL message—ACL_m—and sends it to its peer AMS in the remote platform that restores the agent's execution.

A simulation setting comprising two JADE platforms has been deployed in order to study the performance of agent migration. Each platform executes a number N of AF-agents that migrate at exponentially distributed intervals, thus generating a Poisson distributed series of migrations with a total rate λ. Experiments have been performed for various values of λ and N and the results are presented in Figure 6b. A third determinant of migration delay is the size of the migrating agent's code and state. Due to the relatively large size of AF-agent's code (100 KB) a modification was introduced to the IPMS source code in order to disable agent code packaging in ACL_m. Our assumption is that agent code will be proactively transferred and cached to platforms of 'neighboring' cells before handover execution. The graph in Figure 6b shows a significant increase in migration latency (especially for high migration rates) as the agent state raises from 4 to 7 KB. A good design of AF-agents that limits the size of agent state, as well as the use of high processing power servers for hosting the agent platforms can moderate the migration delay. However, it must be highlighted that migration delay does not contribute to handover latency, as agent migration takes place in parallel with handover execution. The restriction here is that its value should be lower than handover latency so that AF-agent will be available to the MT when the connection to the new network is established.

[¶] Inter-Platform Mobility Project, 2007. URL: https://tao.uab.cat/ipmp/node/16

7. Conclusions and Future Work

We proposed an agent-based architecture for handover management in 4G. The software agents comprising the architecture represent the users, wireless operators, a Multi-Access Provider (MAP) and the regulatory authority. MAP is a third-party maintaining roaming agreements with wireless operators and enabling user utilization of their services through a single subscription. Moreover, MAP serves AAA and billing purposes while supporting handover management. The regulator enhances user trust by monitoring the behavior of the operators and intervening when required. In our approach, handover initiation and decision are delegated to software agents-user representatives. Handovers are initiated by user agents that execute either in the terminal or the network side, depending on the source of handover triggering events. On the other hand, handover decision takes place in the network side for saving terminal's usually limited power and computational resources. We propose the selection of both access and core bearer services during handover decision for better adaptation to user and application requirements. For this purpose, a data model of bearer service is presented and a procedure for bearer selection. Performance evaluation of the architecture through a simulation system has been performed focusing on its impact on handover latency. The results are promising for the feasibility of our approach.

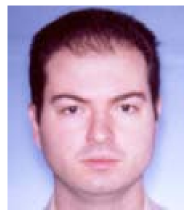
Our future work will focus on architecture enhancements and evaluation of its efficiency on the basis of a system prototype embedded in a real wireless networking environment. Joint resource and handover management performed by agents on network providers' platforms for global load balancing will also be studied. Moreover, we will focus on more complex configurations of the user terminal equipment, as is the case of personal area networks (PANs).

References

1. Zahariadis T. Trends in the path to 4G. *IEE Communications Engineer Magzine* 2003; **1**(1): 12–15.
2. Siddiqui F, Zeadally S. Mobility management across hybrid wireless networks: trends and challenges. *Computer Communications* 2006; **29**(9): 1363–1385, DOI: 10.1016/j.comcom.2005.09.003
3. Gustafsson E, Jonsson A. Always best connected. *IEEE Wireless Communications* 2003; **10**(1): 49–55, DOI: 10.1109/MWC.2003.1182111
4. Bellavista P, Cinque M, Cotroneo D, Foschini L. Integrated support for handoff management and context awareness in heterogeneous wireless networks. In *Proceedings of the 3rd int. workshop on Middleware for Pervasive and Ad-hoc Computing* 2005; pp. 1–8, DOI: 10.1145/1101480.1101495

5. Sun JZ, Riekki J, Jurmu M, Sauvola J. Adaptive connectivity management middleware for heterogeneous wireless networks. *IEEE Wireless Communications* 2005; **12**(6): 18–25, DOI: 10.1109/MWC.2005.1561941
6. Akyildiz IF, Mohanty S, Xie J. A ubiquitous mobile communication architecture for next-generation heterogeneous wireless systems. *IEEE Communications Magzine* 2005; **43**(6): 29–36, DOI: 10.1109/MCOM.2005.1452832
7. Wei Q, Farkas K, Prehofer C, Mendes P, Plattner B. Context-aware handover using active network technology. *Computer Networks* 2006; **50**(15): 2855–2872, DOI: 10.1016/j.comnet.2005.11.002
8. Vidales P, Baliosian J, Serrat J, Mapp G, Stajano F, Hopper A. Autonomic system for mobility support in 4g networks. *IEEE Journal on Selected Areas in Communications* 2005; **23**(12): 2288–2304, DOI: 10.1109/JSAC.2005.857198
9. Ambient Networks Consortium. Ambient networks contextware-second paper on context-aware networks. Deliverable IST-2002-507134-AN/D6-3, 2005.
10. Labrou Y, Finin T, Peng Y. Agent communication languages: the current landscape. *IEEE Intelligent Systems* 1999; **14**(2): 45–52, DOI: 10.1109/5254.757631
11. Chan PML, Sheriff RE, Hu YF, Conforto P, Tocci C. Mobility management incorporating fuzzy logic for a heterogeneous IP environment. *IEEE Communications Magzine* 2001; **39**(12): 42–51, DOI: 10.1109/35.968811
12. Song Q, Jamalipour A. An adaptive quality-of-service network selection mechanism for heterogeneous mobile networks. *Wireless Communications and Mobile Computing* 2005; **5**(6): 697–708, DOI: 10.1002/wcm.330
13. Xing B, Venkatasubramanian N. Multi-constraint dynamic access selection in always best connected networks. In *Proceedings of 2nd Int. Conf. on Mobile and Ubiquitous Systems* 2005; pp. 56–64, DOI: 10.1109/MOBIQUITOUS.2005.39
14. FIPA. FIPA Specifications 2006; www.fipa.org/specifications/index.html
15. 3GPP. 3GPP system to wireless local area network (WLAN) interworking; system description (Release 6). Specification TS 23.234. v.6.2.0, 2004.
16. 3GPP. Principles of circuit telecommunication services supported by a public land mobile network (Release 6). Specification TS 22.001. v.6.0.0 2005.
17. DMTF. Common information model schema. Standard Specification Version 2.11 2005.
18. Koien GM, Haslestad T. Security aspects of 3G-WLAN interworking. *IEEE Communications Magzine* 2003; **41**(11): 82–88, DOI: 10.1109/MCOM.2003.1244927
19. Chakravorty R, Vidales P, Subramanian K, Pratt I, Crowfort J. Performance issues with vertical handovers—experiences from GPRS cellular and WLAN hot-spots integration. In *Proceedings of 2nd IEEE Pervasive Computing and Communications Conference* 2004; pp. 155–164, DOI: 10.1109/PERCOM.2004.1276854

8. Authors' Biographies



Vassilis E. Zafeiris is Ph.D. candidate at the Athens University of Economics and Business. He received his Diploma in Electrical Engineering from the National Technical University of Athens (NTUA) in 2001 and a M.Sc. in Information Systems from the Athens University of Economics and Business in 2003. His research interests include software architectures and interoperability, service-oriented computing, software agents, and their application in mobile networking.

Emmanouel A. Giakoumakis is Associate Professor of Computer Science at the Athens University of Economics and Business. He received his Diploma in Electrical Engineering and Ph.D. in Computer Engineering from the National Technical University of Athens (NTUA) in 1981 and 1988, respectively. He has served as President of the Hellenic Regulatory Authority for Telecommunications (2000–2005). His research interests include software architectures and interoperability, regulation and competition in electronic services, bioinformatics, and software engineering.