NGIneUS: Intelligent User Shadows for Next Generation Wireless Services

Gergely V. Záruba, Wei Wu, Mohan J. Kumar, Sajal K. Das
zaruba, wuwei, kumar, das@cse.uta.edu

Technical Report CSE-2002-1
NGIneUS: Intelligent User Shadows for Next Generation Wireless Services

Gergely V. Záruba, Wei Wu, Mohan J. Kumar, Sajal K. Das

Center for Research in Wireless Mobility and Networking (CReWMaN)
Department of Computer Science and Engineering
The University of Texas at Arlington, 76019-0015
{zaruba, wuwei, kumar, das}@cse.uta.edu

Abstract

This paper proposes NGIneUS (pronounced: ingenious), a novel framework for seamless integration of multi-tier wireless networks, i.e., 4th generation (4G) or in general next generation (NG) wireless systems. NGIneUS introduces software agents called user shadows that augment wireline networks and cater to the needs of the mobile user in overlay wireless access networks. With the help of the user shadows, most signaling interaction between mobile devices, corresponding hosts and wireline networks is restricted to the wired links, reducing the burden on the limited wireless bandwidth. Furthermore, the heterogeneity of the access networks is made transparent to the users, as the user shadows take charge of the functionalities of multi-tier stream admission control and scheduling while providing with seamless handoffs. Additionally, the novel concept of software IP mobility is introduced, where software entities are enabled to communicate using their own IP stacks, enabling IP mobility of user shadows. Software IP mobility enhances the current IP mobility approaches and can help in providing with seamless mobility. User profiles to be managed and included by the user shadows are outlined and discussion on their importance on NGIneUS is provided.

1. Introduction

Fueled by the advancement of wireless technologies and the emergence of multimedia data services, cellular wireless networks have evolved to their third generation (3G) in just two decades. The first generation mobile wireless systems were based on analog transmission technologies, deployed around the 1980s, with voice transmissions as the only available service. During the 1990s, second generation (2G) digital wireless systems replaced analog wireless
systems in most countries. Digital communication not only results in better spectrum utilization and superior quality of voice but it also enables users to take advantage of advanced services, such as short messaging services (SMS) and digital modem services. As a transition phase towards 3G wireless systems, 2.5G wireless networks (such as General Packet Radio Service - GPRS) appeared to enhance the 2G wireless networks with best-effort packet switched data services. However, with 2.5G wireless systems, data rates are limited to only a couple of tens of kbps. One of the projected goals of 3G systems was to provide with high data rate services to enable the delivery of multimedia information to mobile devices (384kbps to 2Mbps). It is predicted that comprehensive 3G wireless networks will only be available around 2005 due to the costly deployment/upgrade of system equipment; while other predictions foresee a “generation jump” directly to 4G networks [1,2,3], and thus use the term NG (Next Generation). The theories on the generation jump are based not only on the current slowdown in economy being especially hard on telecommunications, but also on 3G’s latency on the market – not being able to provide with the desired levels of service. 4G wireless systems will have to provide users with 100-200 Mbps while enabling users to dynamically switch between various wireless access technologies to the IP-based core network. Users are expected to use different interfaces transparently and simultaneously to access the network. Therefore, system integration is the most important enabler to the evolutionary step from 3G to 4G wireless systems.

This paper presents NGIneUS (Intelligent User Shadows for Next Generation Wireless Services – pronounced: ingenious), a framework to fulfill the task of service integration into a uniform perception wireless access system. The novel concept of NGIneUS aims to solve the system integration problem by instituting mobile software entities, called user shadows, in the wired part of the network. Shadows mimic users for the requested content services, while users’ equipment function as remote terminals to the shadows. NGIneUS makes the heterogeneous wireless networks look like a single, transparent, and high-output network to the users, without
requiring them to negotiate with the networks and manage their device’s network attachment point.

For the sake of completeness, in the next section, we provide an overview of some parallel research projects on 4G wireless services.

1.1. Research Related to 4G Services

Although most service providers are still expanding and improving their 2.5G networks, research towards 4G systems is already making significant progress. In this chapter the reader will be introduced to some of the major ongoing 4G research efforts.

The Focus project of WINLAB at Rutgers University [4, 5] investigates the fundamentals of 4G mobile network architectures and protocols. In particular, they propose an open-architecture, programmable mobile networks approach for gradual evolution of service features via ad-hoc peer-level collaboration of wireless network entities. The advantage of the open-architecture is the reduced need for complex standards that anticipate all future needs. Focus envisions 4G networks to relay on an IP-based core network for global routing along with more customized local-area radio access networks that support features such as dynamic handoff and ad-hoc routing. This project is currently in the early design and experimental mobile network testbed establishment phase. An experimental open-architecture 4G mobile network testbed consisting of open API wireless terminals, forwarding nodes, access points, switches and routers is being set up to evaluate the proposed techniques. Topics under consideration include: 3G/WLAN interworking, open-architecture wireless networks, data caching and content delivery for mobile users, protocols for self organization, routing in hierarchical ad-hoc networks and unlicensed band spectrum etiquette.

The 4G research project of BWN-Lab at the Georgia Institute of Technology [6, 7] focuses on mobility management in heterogeneous 4G networks, particularly; network-layer mobility is investigated as the key issue. Network-layer mobility schemes are proposed to overcome the
drawbacks of Mobile IP in supporting real-time location management and fast, seamless handoff. They propose a distributed and dynamic regional location management scheme to minimize the location update and packet delivery cost of Mobile IP. In this scheme, the regional network boundary is dynamically adjusted based on the mobility pattern and traffic load for each mobile node. An analytical model was developed to capture the mobility and packet arrival pattern of individual mobile nodes.

The 4GW project [8, 9] of the Personal Computing and Communications group at the Lund Institute of Technology aims to propose an overall 4G wireless infrastructure with respect to technological as well as economical and regulatory perspectives. The research topics were identified using a scenario-based approach; scenarios are not used to predict the future, but to investigate which future developments towards 4G are possible. Main problem areas identified include air interface and infrastructure design, economical feasibility, services, terminal and base-station design, and mobility management.

Today, research on 4G wireless systems is still in its infancy. Research is mainly focusing on identifying problems and research directions. It is widely agreed that system integration of various wireless access networks is the major research topic leading to NG wireless systems.

1.2. A Possible 4G Usage Scenario

Before presenting the NGIneUS framework, let us motivate the reader with a possible 4G scenario. Julie is a corporate employee, who likes to watch the news on the TV in the morning. When she rushes to work, the streaming video news is delivered to her mobile device; the data may be pre-processed using methods, such as transcoding, to be compatible to the processing and display capability of her mobile device. As Julie rides the subway-train to work, she can get access from different wireless access networks (e.g., UMTS, GPRS, CDMA) along her route; her multi-mode mobile device has to choose the appropriate wireless interface for communications depending on her profiles, e.g., requirement of image resolution, memory size of mobile device or...
subscription plan. The interface switching remains transparent to Julie, she is unaware of which wireless network provides her with the video stream. During her ride to work, she receives a video-phone call from a customer. She prefers high-quality voice when talking to customers, so the UMTS air interface is chosen to transmit the voice although it costs more, while the image of the customer is transmitted through cheaper GPRS network with lower quality. The same call is still ongoing while Julie reaches her work, but to save cost, the video stream is handed over to her corporate wireless LAN. Another day, Julie realizes in the morning that she has forgotten to charge the battery of her mobile device as this turns off while she is watching the TV news. However, she is expecting the access system to buffer the TV news to be delivered to the mobile device after Julie replaces her battery. With today’s technology this scenario is rather futuristic, emphasizing the need for novel frameworks to achieve seamless system integration leading to 4G networks.

The rest of the paper is organized as follows. In Section 2, the NGIneUS framework is presented: first, the concept of user shadows and software IP mobility are introduced. Then, three key components of the framework are described, namely: mobility management, stream admission control- and scheduling, and user profile management. Section 3 concludes the paper and provides a brief overview of the future-work challenges.

2. NGIneUS Framework

The motivation behind our work leading to NGIneUS is the need for seamless integration of various wireless access networks for NG wireless systems. NGIneUS uses the fact that processing power and wireline bandwidth are easily expandable and thus their extensive use can be justified in order to efficiently control the scarcest resource, namely the wireless bandwidth.

In our approach, the 4G wireless system is a heterogeneous network consisting of different access networks, which may overlap with one another (i.e., wireless overlay networks). In this
environment, a mobile is equipped with a mobile device containing multiple wireless interfaces or a multi-mode interface to connect to wireless access networks anytime and anywhere. We envision software entities, called user shadows, controlling and negotiating with the network on behalf of the users, while using these shadows to constrain the signaling overhead to the wired network as much as possible. There is a one-to-one mapping between user shadows and users, i.e., an individual owns one user shadow, and a user shadow belongs only to one user. User shadows appear to the content services as if they were the users themselves, while providing with a remote terminal interface to the user through one or more of the available wireless interfaces. User shadows run and migrate between dedicated environments, namely NGIneUS servers (NGS). An architectural view of NGIneUS is provided in Figure 1.

![Figure 1. Architecture layout of NGIneUS](image)

NGSs are deployed in each wireless domain or autonomous wireless domain (AWD) to ensure low latency data exchange between the AWD controller/router and the NGS. For example,
in a regular cellular architecture, an NGS could be mapped to each mobile service switching center (MSC) controlling the corresponding base station controllers (BSC) of that AWD; while in a corporate wireless LAN environment, the NGS can be a computing server attached to the backbone interconnecting the access points. The NGS and the AWD routers or base stations will have a common query/command interface using which the NGS can query the status of AWD, receive asynchronous status information and commands, or relay crucial information to the AWD controller to help perform tasks like handoff or stream (call) admission control.

User shadows will receive and terminate the streams of the services to and from the mobile terminals; users do not directly receive and transmit packets to the peer of the streams but will communicate only with their respective shadows residing at the appropriate NGS. The users’ device thus works as a remote terminal to the shadows in the wired networks, and can control the behavior and display status of its corresponding shadow. With the use of the NGS, the shadows representing the users terminate all network layer connections or streams (such as TCP and UDP), and provide the user’s equipment with the filtered information via wireless connections. Thus the user shadows running at the NGS provide transparency to the “wired world”, appearing to the peer applications as if they were the users themselves, i.e., there is no difference between static or mobile entities for peer applications and corresponding hosts.

2.1. User Shadows

In the NGIneUS framework, each mobile user is associated with a mobile software entity called user shadow, that follows the user’s path in the wired infrastructure and controls network access on behalf of the user, hence the name. Shadows reside inside and migrate between designated NGSs co-located with each segment of the wireless access points. User shadows attempt to enhance infrastructure proactivity and user transparency, to make the heterogeneous wireless networks look like a single, transparent and high-output network to the user without requiring her to negotiate and manage her device-network attachment point.
User shadows move with the users from one location (and its respective NGSs) to another. As the user moves in the physical world her shadow will migrate in the wired network, trying to optimize the distance between itself and its owner. An interface manager residing on the users’ equipment monitors the activities of the multiple wireless interfaces of the mobile device. It is the task of the user shadows to determine whether to migrate to another NGS closer to the user based on the information collected from the interface manager as well as the user profiles (as explained later). The availability of a new access domain can be determined either by the users’ device detecting radio signals from a new access domain, or by the shadow querying the information database of the current NGS about other overlay networks in the area or the location the user resides at. Shadow migration will happen either if the user has moved out from the area served by the NGS or if the user’s profile patterns indicate another, better suited NGS (access network) in the area. For example, if the user leaves her home for work, the shadow can follow her using NGSs in the UMTS network along her path to work. Upon the user’s arrival at her workplace, the shadow is expected to move to the NGS of the corporate LAN environment. Shadow migration decision can greatly benefit from prediction based on user mobility pattern profiles. Figure 2 shows the migration procedure of the user shadow from one access domain to another.

Migration is not the only way for user shadows to move closer to the users. The fact, that there is a one-to-one mapping between users and shadows implies that a shadow can only be present at the NGS of one of the wireless interfaces. Yet, the user has to be enabled to use the other available access networks as well. Thus, user shadows can instruct the NGSs of those access networks to spawn a process, which we refer to as the **slave user shadow**. The slave user shadow’s task is to receive/transmit packets to/from the user through the wireless interface relaying those packets to the shadow. Another task of the slave is to negotiate access parameters with the access network controller on behalf of the user (and the shadow). Since slave shadows are access network specific, no data or code migration is initiated by the user shadow; it simply
instructs the NGS to create a new process. User shadows and their slaves communicate through a common protocol, called *inter-shadow messaging protocol* (ISMP).

![Diagram of network connections](image.png)

**Figure 2.** Migration of user shadows

Another major difference between user shadows and their slaves is that the former have their own IP address while the slaves use the IP address of the corresponding NGS and a port number. While user shadows never cease to exist though they migrate from one NGS to another, slave shadows can as easily be terminated by their masters as they were created; slave user shadows do not carry any important information about the user, her behavior or the properties of her connections. Using this approach one software entity, i.e., the master user shadow can control the information flow arriving and departing simultaneously through various wireless interfaces. Although the message exchange between shadow master and slaves introduces signaling overhead, such messages are within the wired networks, thereby not affecting the bandwidth constrained wireless links. The information collected from the slave shadow plays an important role in the mobility management, stream admission control, scheduling, and profile management.

Figure 3 depicts the slave user shadow spawning procedure where the user is originally connected to only a macro-cell provider thus 100% of her traffic is routed on the only available
interface. When the user moves into a hot-spot area, a slave shadow is spawn in the NGS of the hot-spot; 40% of her traffic in the example is routed over the hot-spot’s interface.

![Diagram of Master vs. Slave User Shadows](image)

**Figure 3.** Master vs. slave user shadows

### 2.2. Software IP Mobility

Current IP mobility standards, models and proposed drafts deal with the mobility of hardware devices, thus associating an IP address with the wireless interfaces of devices as these physically move around in the wireless coverage area. In the NGIneUS framework, mobile software entities are allowed to be associated with IP addresses, i.e., applications are allowed to address and communicate with mobile software code using a simple IP address. This is essential to the goals of NGIneUS, since the user shadows are mobile software entities and yet they will have to function as terminating/source points for IP packets and overlying transport layer data.
streams. In NGIneUS, it is the user shadow that carries the user specific IP address and terminates all (network and) transport layer streams. Since user shadows reside inside NGSs, routing functions have to be included in all NGSs; routing must be done between the IP layer associated with the physical interface of the NGS and the IP stacks of each of the software interfaces created by the individual user shadows. Figure 4 shows the protocol stack in NGS.

![NGS Protocol Stack Diagram](image)

**Figure 4.** NGS protocol stack.

User shadows create their own virtual (software) interface; thus each user shadow has its own TCP/IP protocol stack and routes all the packets to the default IP of the NGS IP layer. Similar to the IP address assignment to a physical mobile device using Mobile IP, a user shadow can also have its home IP address and care-of-address in the foreign networks. Alternatively, the shadow IP address can be acquired from any DHCP server with dynamic update to the DNS server. In other words, user shadows can be treated just like “mobile devices” and as such devices, they need IP mobility handling; since user shadows are software agents, we refer to their mobility as software IP mobility.

With the unique IP address associated with shadows, they can always be reached even after migration among the NGSs. Since a mobile user is represented by its shadow to the outside
world, a binding mechanism is needed between the user and her shadow. To update the binding with the shadow, the mobile user sends refreshing messages to the NGS in the same domain periodically or on-demand. In turn, the NGS relays the message to the corresponding user shadow by checking the MAC address of the mobile device.

### 2.3. Mobility Management

The user shadows provide a framework to perform the tasks of handoff decisions whether these handoffs are horizontal or vertical. *Horizontal handoffs* occur when the servicing of mobile terminals are handed off between base stations of the same wireless access technology. On the other hand, *vertical handoffs* take place if the wireless attachment point of mobile terminal is changing the type of the wireless interface to another technology (e.g., switching an ongoing voice communication stream from UMTS to HiperLAN/2).

In the NGIneUS framework, a horizontal handoff always implies the user shadow migrating from one NGS to another, while vertical handoffs are not hard handoffs in the sense that all interfaces can be used simultaneously (vertical handoffs may or may not require spawning a new slave shadow). Prediction techniques and user mobility profiles could enable the shadow to start the communication, or delegate itself close to the access points that are likely to be visited by the user in the near future, thus being able to provide her with the same QoS during the process of changing the wireless attachment point. Figure 5 shows an example of a mobile user performing horizontal and vertical handoffs in a heterogeneous network. The dashed lines represent periods when several wireless interfaces are exploited simultaneously and perform different kind of handoffs as the user moves.
Figure 5. Horizontal and vertical handoff in heterogeneous networks.

Mobility management’s other aspects such as paging (finding a mobile user in the wireless network), registration (enabling mobile users to register with access points to relax requirements on paging) and possible mobility tracking receive an additional level of complexity by introducing heterogeneity to the wireless system. If all systems are considered as individual infrastructures, then each of them requires the above functions individually. In a combined – integrated wireless access system such as a 4G system, these mobility related functions might be integrated thus reducing the wireless bandwidth used for the related signaling traffic required. Furthermore, because of the non-overlapping cell boundaries of different technologies, the precision of the user’s location and hence the precision on a possible location/mobility prediction can be increased which would benefit the systems as a whole. NGIneUS can provide with the required infrastructure to take advantage of the integration of the interfaces to create a better mobility management framework by enabling the interaction of the controllers of different
technologies on a case-by-case basis to the users. Techniques that minimize paging/update
message overheads such as [10] can be extended to handle hierarchical structures and are
expected to work well with user shadows.

2.4. Stream Admission Control and Scheduling

Stream (or service) admission control is the generalized version of the term “call admission
control” and describes the process in charge of determining which services (calls) of users can be
allowed in the system without overloading the system. While finding efficient and proper call
admission control policies is and has been a problem frequently investigated by researchers, most
work concentrates on the case where a single wireless technology (or several technologies served
by the same base stations – interwoven into each other) provides all the service [11, 12]. The
availability of several wireless interfaces and the need to exploit them at the same time creates an
additional dimension to the stream admission control process, since a new service may be
admitted by several different interfaces (possibly by “splitting up the traffic” and thus combining
the resources available from all wireless interfaces).

To serve quasi static users in heterogeneous wireless environments, the user shadow can
make a decision on which of the available transports/interfaces would best suit the user for the
current service she is using or requesting (note that changing the wireless interface or the degree
of involvement of wireless interfaces may require vertical handoffs). The shadow coordinates
with all the individual wireless access networks, by employing slave shadow representatives to
determine which of them would best serve the stream requested by the user. This determination
requires a continuous optimization process operating on information acquired from the access
points and the user.
2.5. **User Profile Management**

Collecting information on user behavior and preferences can help determine the system performance more accurately while enabling a more precise prediction on what a user is going to do and how she is going to behave in the near or distant future. Integration of user profiles at different layers of the hierarchy not only enables global prediction of user behavior for better network efficiency but also ensures that users are satisfied with services (retaining the user as the subscriber). Profiles to be maintained by the user shadows about their respective users in the NGIneUS framework include the following:

- **Mobility and location profiles** help track mobility. They contain information on the mobility behavior of users, which can be used to predict future cell crossing and thus can be of great benefit in providing with seamless handoffs. They are also extremely helpful in determining the current location of users to reduce wireless signaling overhead associated with the paging and registration processes.

- **Device profiles** contain information about what kind of devices the user is using, and which one of them she is using at any time. These profiles are necessary to enable users to connect to the same services with little impact on what devices are being used. Employing device profiles enables device heterogeneity in NG networks, where data flowing to devices is pre-filtered via the shadows according to the devices’ capabilities, reducing the amount of data that gets thrown away at the device and increasing wireless bandwidth efficiency. Device profiles for common devices may be downloaded from designated directories but may also be created and modified by the user. Compared to mobility profiles, device profiles are quasi-static, so they do not need constant updates and have less dependence on the user behavior.

- **Service profiles** can be used to collect information on the services and their QoS (Quality of Service) properties that individual users are requesting by possibly maintaining
frequency (arrival rate), duration (holding times), and time-of-day information on these requests. The collection and availability of such information enables the system in making better admission and handoff decisions.

- **Satisfaction profiles** maintain information about how satisfied the user may be with the services received from the 4G system hence they assist in regulating policies to ensure higher satisfaction, which may help reduce and prevent users from leaving the service provider by terminating the contract. Information collected may contain such data as the frequency of traffic contract violations, frequency and probability of call blocking, etc.

- **Subscription profiles** maintain information about the user’s willingness to pay for the requested QoS to enable a more proper and individually fit wireless interface selection. For example, such profiles may prevent a user receiving a stream from a low (or no) cost interface to be handed off to a high cost interface.

- **Commercial profiles**, maintain information about the user’s willingness to receive unsolicited information based on location, that is, whether the user is willing to receive such information for a reduced access rate or free local service. These profiles can enable the users to take advantage of free commercial services as long as they are accepting the advertisements from the operators of those services.

The user shadows of NGIneUS not only facilitate the collection of profile information as described above, but also can incorporate the information collected in various decision processes. Among other advantages, user shadows can move with the users with fewer overheads while not being involved in any active communication. Shadows can more efficiently take advantage of the user profiles, can better position themselves and initiate resource reservation ahead of time by predicting the next cells that will need to be covered. They can also pre-filter data according to the device profiles due to their “being the user” property seen by the requested services.
3. Conclusions and Future Work

In this paper, we presented NGIneUS as a framework for seamless integration of multi-tier, overlay wireless networks, such as 4G (or NG in general) wireless systems. A new software entity called user shadow that takes care of all the needs of the mobile user of the overlay wireless access networks has been introduced to the wireline network. With the help of the user shadows, most signaling interaction between mobile devices and wireline networks is restricted to the wired links, which can significantly reduce the usage of the limited wireless bandwidth. Furthermore, the heterogeneity of the access networks is made transparent to the users by the functionality of the user shadows, such as stream admission control and scheduling. Additionally, the novel concept of software IP mobility has been introduced, where software entities are enabled to communicate using their own IP stacks thus enabling IP mobility of user shadows. Software IP mobility enhances the current IP mobility approaches and can help provide with seamless mobility. Various kinds of user profiles managed and included by the user shadows, have been outlined and discussion on their importance on NGIneUS has been provided. All the decisions made by the user shadows are based on the technology perspective as well as user profiles, which is the one of the unique features of 4G systems compared to the previous generation mobile wireless systems.

Let us revisit the scenario presented in Section 1.2 using NGIneUS as the framework for 4G system integration. When Julie leaves her home, the streaming video news is delivered to her user shadow, where it is pre-processed according to her device profile. The user shadow migrates between the NGSs along her way to the office optimizing its location based on the ongoing services. Depending on Julie’s user service- and subscription profiles, the voice and video of the video-phone call will be delivered automatically via the appropriate air interfaces. When she arrives at her company, her shadow migrates to the corporate WLAN, and re-optimizes the traffic between the available interfaces. While she is involved in the video conversation with her
customers, a slave user shadow is spawned in the GPRS network managed by the master user shadow in the UMTS network. During the time the device is turned-off due to the battery discharge, the user shadow continues to receive the TV news streams, buffers them at the NGS, and when the device comes back alive the shadow delivers the buffered data.

Currently we are in the process of implementing a software IP mobility testbed, to support IP communication and mobility of software entities. We are also designing and evaluating protocols and algorithms for stream admission control and scheduling, particularly, methods of splitting and recombining the traffic to and from different wireless interfaces. Handoff management among heterogeneous systems is another challenging problem. The handoff decision will be made by the user shadows based on both the technical consideration such as signal to noise ratio (SNR) and information of user profiles. Therefore, user profile management is also to be investigated in the profile collection, maintenance and appropriate reaction, prediction and decision on the collected data. Since user shadows represent the mobile users in the networks, the security problem like the authentication, authorization and accounting (AAA) association among users, user shadows and NGSs is also an important challenge.

References


