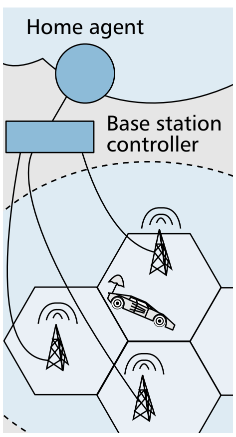
VERTICAL HANDOFFS IN FOURTH-GENERATION MULTINETWORK ENVIRONMENTS

JANISE MCNAIR AND FANG ZHU, UNIVERSITY OF FLORIDA



Traditional operations for handoff detection policies, decision metrics, and radio link transfer are not able to adapt to dynamic handoff criteria or react to user inputs and changing network availabilities. Thus, new techniques are needed to manage user mobility between different types of networks.

ABSTRACT

Revolutionary drivers for 4G include a push toward universal wireless access and ubiquitous computing through seamless personal and terminal mobility. One of the major challenges for seamless mobility is the creation of a vertical handoff protocol: a handoff protocol for users that move between different types of networks. Traditional operations for handoff detection policies, decision metrics, and radio link transfer are not able to adapt to dynamic handoff criteria or react to user inputs and changing network availabilities. Nor are they able to deliver context-aware services or ensure network interoperability. Thus, new techniques are needed to manage user mobility between different types of networks. This article presents a tutorial on the design and performance issues for vertical handoff in an envisioned multinetwork fourth-generation environment. Various network architectures and technologies for 3G and beyond are described, including wireless LANs, cellular, satellite, and Mobile IP. Then the problem of vertical handoff is defined in the context of such a diverse network environment. Finally, research efforts to resolve the open problems are explored, including new techniques for dynamic handoff decision and detection algorithms and context-aware radio link transfer.

INTRODUCTION

At the turn of the 21st century, the widespread success of wireless and mobile communications has resulted in the creation of a large variety of wireless technologies, including second- and third-generation (2G and 3G) cellular, satellite, Wi-Fi, and Bluetooth. Each technology is tailored to reach a particular market, or a particular type of user with a specific service need. The advantage to these diverse networks is that they offer many choices for increasing bandwidth, accessing the Internet, and increasing the coverage area for the average user. However, expanding services through the use and coordination of diverse networks creates the challenge of developing novel interoperable network protocols to manage user mobility between different types of systems — a level of interoperability currently not available in 3G wireless systems [1].

The fourth generation (4G) of wireless communications refers to the next evolutionary step after standardization of the 3G infrastructure and the next revolutionary step for wireless telecommunications in general [2]. The evolutionary goals of 4G beyond 3G include building on packet-based code-division multiple access (CDMA) networks under such systems as the Universal Mobile Telecommunications System (UMTS). The revolutionary goals are visionary in nature, requiring an evaluation of the technological, societal, and market developments over the next 10 years. Some goals may be forecast by emerging issues, such as spectrum efficiency, dynamic bandwidth allocation, security, quality of service (QoS), and transceiver technology, while other goals may arise from factors that dictate entirely new approaches and novel infrastructure solutions.

Revolutionary drivers for 4G include a push toward universal wireless access and ubiquitous computing through seamless personal and terminal mobility [1, 3]. Universal wireless access refers to the ability of a user to connect anywhere at any time from any network. The change in connection may be initiated by the user or may be initiated by the network, transparent to the user. For example, a user may choose to access a wireless LAN (WLAN) to send a large data file, but may choose the cellular network to carry on a voice call. On the other hand, a network may decide to hand off a stationary data user to a WLAN in order to increase bandwidth availability for mobile users in a 3G cellular network. Ubiquitous computing refers to the ability to move seamlessly within a network while receiving intelligent, context-aware services. Personal mobility allows a user to receive services at any terminal device, while terminal mobility allows the device to receive services even as it moves between network access points. To achieve seamless mobility, network management operations must be conducted without causing degradation of services, and without need for user intervention.

The movement of a user within or among different types of networks can be referred to as intersystem or vertical mobility. One of the major challenges for seamless vertical mobility is vertical handoff, where handoff (or handover) is the process of maintaining a mobile user's active connections as it changes its point of attachment. Traditionally, handoff research has been based on an evaluation of the signal strength received at the mobile node, followed by a change in access point, if needed, and an updated routing path for the user connection. However, with a vision of a diverse multinetwork environment, and considering the goals of transparent universal access, ubiquitous computing, and seamless mobility, traditional signal strength comparisons are not sufficient to make a handoff decision, as they do not take into account the current context or the various attachment options for the mobile user. Another issue in vertical handoff is the timely and reliable transfer of a mobile user's connection(s). While traditional link transfer techniques can achieve fast handoffs, there is now a need to consider the context of the link transfer, including security associations, QoS guarantees, and any special processing operations. Thus, the vision of 4G requires investigation of a more adaptive and intelligent network approach to vertical handoff.

This article presents a tutorial on the design and performance issues for vertical handoff management in an envisioned 4G multinetwork environment. We describe various network architectures and technologies currently evolving beyond 3G, including WLANs, cellular, and Mobile IP. We explore the problem of vertical handoff design in the context of the envisioned environment. Finally, we describe the open research problems for achieving an interoperable and transparent handoff decision and detection algorithm, and a context-aware radio link transfer, respectively. We then conclude the article.

Architectures and Enabling Technologies

The evolutionary architecture beyond 3G builds on a hierarchical cellular system for wireless wide area services, and a mobile satellite network to provide GPS location services, high-bandwidth pipes, and the ability to reach customers in rural areas [1]. However, as mentioned previously, the widespread success of wireless communications has resulted in the addition of an even greater variety of wireless networks that must coexist. Some of the various types of networks include the following:

- Wireless personal area networks (WPANs) and enabling technologies, such as Bluetooth, that provide range-limited ad hoc wireless service to users for access to a variety of personalized items
- WLANs, such as 802.11, that provide Ethernet access to wireless users without the costly infrastructure of 3G
- Wireless wide area networks (WWANs), such as UMTS, that provide global cellular service to mobile users
- High aeronautical altitude platforms (HAAPs), such as unmanned air vehicles (UAVs), that use aircraft to provide flexible wireless access without the costly infrastructure of a satellite network

The evolution of 3G packet-based communication has brought increased significance to wireless and mobile Internet access, and a vision of a future based on all-IP networking [3]. In support of the all-IP vision, the Third Generation Partnership Projects (3GPP and 3GPP2), which represent the standards of the global wireless industry, have begun to develop all-IP versions of their respective 3G wireless architectures. In [4], a study is performed to compare five different prospective architectures for implementing a vertical handoff between networks based on the WLAN standard, IEEE 802.11, and networks based on the 3G cellular data standard, General Packet Radio Service (GPRS). The two architectures found to be the most efficient, without requiring a master/slave relationship between the different networks, were the mobility gateway/proxy-based architecture, which consists of a proxy implementation between a GPRS network and a WLAN, and an architecture based on Mobile Internet Protocol (Mobile IP).

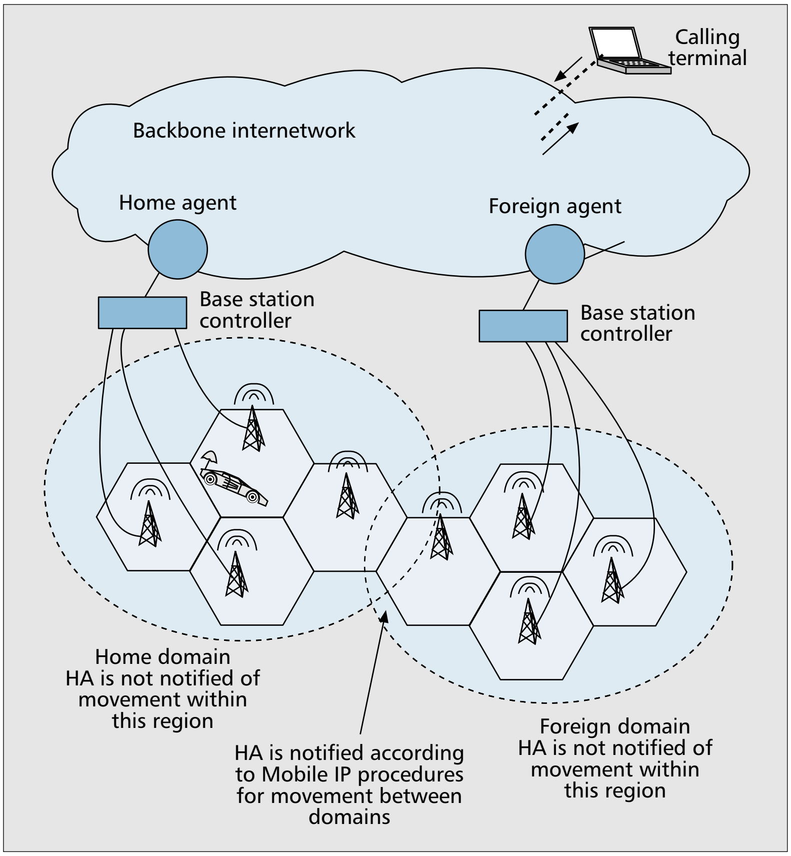
Evolution of Mobile IP

In the mid-1990s, Mobile IP was standardized by the Internet Engineering Task Force (IETF) to allow mobile nodes to change their point of attachment to the Internet while still being able to maintain a connection to the network [5]. Under Mobile IP, a mobile node that is currently residing in its home subnetwork is served by a home agent that forwards all incoming packets to the mobile node at its home IP address. When the mobile node moves away from its home subnetwork to a new location, the node must contact a foreign agent at the new subnetwork to obtain a new IP address, called a *care-of address*. A binding update must then be performed to notify the home agent about the mobile node's new care-of address. The home agent then forwards all incoming packets to the mobile node using a process referred to as *tunneling*: the home agent encapsulates the incoming packets for the mobile node and forwards them to the foreign agent, which in turn decapsulates them and delivers them to the mobile node. Meanwhile, the mobile node can continue to transmit packets directly to the correspondent node.

For Mobile IP, the binding updates and care-of address exchanges that establish the mobile node at each new location cause an increased signaling load as well as delays that may be detrimental to the service being received at the mobile node. In some cases, these delays may not be necessary, since other techniques can be used to resolve packet forwarding to the roaming mobile node. One such method is referred to as *micromobility*.

For Mobile IP, the binding updates and care-of-address exchanges that establish the mobile node at each new location cause an increased signaling load as well as delays that may be detrimental to the service being received at the mobile node.

Micromobility Protocols — Micromobility protocols reduce the need to change the foreign agent of the mobile node for intradomain mobility. As illustrated in Fig. 1, a hierarchy of base stations is organized under one foreign agent, so when the mobile node changes base stations within the same domain, location information is propagated only through the domain's local routers, transparent to the home agent. Incoming packets arriving at the domain can be efficiently and quickly forwarded to the mobile node's current location without the need for another binding update or care-of address. Note, however, that



■ Figure 1. *Mobile IP and micro-mobility.*

when a mobile node moves into a new domain, a traditional Mobile IP handoff is necessary.

Micromobility protocols discussed in the research literature include *Cellular IP* [6] and HAWAII [7]. In Cellular IP, routing decisions in the local domain are conducted using a local gateway and base station that cache the forwarding path of each packet that arrives from or goes to the mobile node, while HAWAII uses a crossover router to manage handoff between two base stations within the same domain.

Mobile IP Version 6 — Other improvements have been proposed and adopted for Mobile IP under the title of Mobile IP version 6 [5]. For example, Mobile IPv6 eliminates triangular routing and enables the correspondent node to reroute packets on a direct path to the mobile node. This process is referred to as *route optimization*, which is not always available in Mobile IPv4. In addition, Mobile IPv4 also suffers from a lack of security constructs for authorization, authentication, and accounting, as well as for source routing. Mobile IPv6 includes embedded binding updates and care-of address configuration for the execution of location updates and processing the change in the mobile node's address. The newer version also includes authentication header processing to provide validation of mobile nodes. Finally, IPv6 has a fourfold increase in IP address space, which may be useful for developing new mobile node addressing schemes.

Regardless of the type of network, issues such as addressing, route optimization, and authentication are part of the challenging overall problem of creating an efficient handoff mechanism that satisfies the seamless mobility needs of the user population while enabling advanced processing and optimization operations at the network. The resolution of these problems in a multinetwork environment is the goal of vertical handoff research, described next.

Vertical Handoffs in 4G Networks

Any handoff operation is a three-stage process that includes *handoff decision*, *radio link transfer*, and *channel assignment* [1]. Traditionally, handoff decision is performed based on a perception of channel quality reflected by the received signal strength and other measurements, and the availability of resources in the new cell. The base station usually measures the quality of the radio link channels being used by mobile nodes in its service area. This is done periodically so that degradations in signal strength below a prescribed threshold can be detected and handoff to another radio channel or cell can be initiated. Under network-controlled handoff (NCHO) or mobile-assisted handoff (MAHO), the network makes the decision for handoff, while under mobile-controlled handoff (MCHO), the mobile node must take its own signal strength measurements and make the handoff decision on its own. While performing handoff, the mobile node's connection may be created at the target base station before the old base station connection is released. This is referred to as a *make before break* handoff. On the other hand, the new connection may be set up after the old connection has been torn down, which is referred to as a *break before make* handoff. In either case, the mobile node executes a *hard handoff*, which means that the mobile node can only communicate on a channel with one base station at time. In 3G CDMA networks, a mobile node is able to communicate on more than one coded channel, which enables it to communicate with more than one base station. Thus, CDMA networks allow a *soft handoff*, where the mobile node can listen to a set of candidate base stations at the same time before choosing one for its point of attachment.

The second part of the three-stage handoff is radio link transfer. Radio link transfer refers to the responsibility of the network to form new links to the call at its new point of attachment. Handoffs to another radio channel within the same cell, referred to as *intracell handoff*, require no new link transfer operations. However, handoff to another cell as a result of mobile node movement to a new base station is referred to as *intercell handoff* and requires handoff *rerouting* operations to link the mobile's existing communication path to the new cell. The third handoff stage, channel assignment, consists of the allocation of resources to the handoff call at the new point of attachment. (Note that channel assignment for handoff calls is also part of the problem of resource management and call admission control for wireless networks, and thus has not been included in this investigation of vertical handoff issues.)

The traditional handoff process described above is insufficient for the challenges of the 4G system for the following reasons:

Criteria. The use of the signal strength criterion for traditional handoffs limits the ability of the network to initiate a handoff for control reasons (congestion relief, change in data traffic, etc.).

User selection. Traditional handoff does not allow user selection of networks, and assumes that there is only one choice for access technology. In a heterogeneous environment, user choice is a desirable amenity.

Context. Whereas traditional handoff link transfer concerns the delivery of packets to the new point of attachment, there is now a need for the delivery of the context of the information flow between the mobile node and the network as well. Context may include security associations, QoS guarantees, authentication headers, and so on.

Interoperability. Traditional handoff protocols are developed for homogeneous systems that rely on a common signaling protocol, routing technique, and mobility management standard. In heterogeneous environments, mobile nodes and network routers must be able to interoperate with different networks, and with the corresponding protocols and standards.

In the next two sections, we define the open problems for implementing vertical handoffs and present the state of the art in current research in these areas.

Vertical Handoff Decision Metric and Policy Design

Handoff metrics are the qualities that are measured to give an indication of whether or not a handoff is needed. As stated previously, in traditional handoffs only signal strength and channel availability are considered. In the envisioned 4G system, the following new metrics have been proposed for use in conjunction with signal strength measurements:

Service type. Different types of services require various combinations of reliability, latency, and data rate.

Monetary cost. Cost is always a major consideration to users, as different networks may employ different billing strategies that may affect the user's choice of handoff.

Network conditions. Network-related parameters such as traffic, available bandwidth, network latency, and congestion (packet loss) may need to be considered for effective network usage. Use of network information in the choice to hand off can also be useful for load balancing across different networks, possibly relieving congestion in certain systems.

System performance. To guarantee the system performance, a variety of parameters can be employed in the handoff decision, such as the channel propagation characteristics, path loss, interchannel interference, signal-to-noise ratio (SNR), and bit error rate (BER). In addition, battery power may be another crucial factor for certain users. When the battery level is low, the user may choose to switch to a network with lower power requirements, such as an ad hoc Bluetooth network.

Mobile node conditions. Mobile node conditions include dynamic factors such as velocity, moving pattern, moving histories, and location information.

User preferences. User preferences can be used to cater to special requests for one type of system over another.

The use of new metrics will increase the complexity of the handoff process, making the handoff decision more and more ambiguous, and the development of a cost function to simultaneously evaluate various metrics becomes crucial to the success of a 4G handoff decision. In [9], a cost function is developed for multiservice networks that considers several of the factors outlined above in a two-dimensional handoff cost function. In one dimension, the function reflects the types of services requested by the user, while in the second dimension, it represents the cost to the network according to specific parameters, such as bandwidth, power consumption, and monetary cost. The general form of the cost function, *f*n, is

$$f^{n} = \sum_{s}\sum_{i} w_{s,i} \cdot p^{n}_{s,i} \tag{1}$$

where $p^{n}_{s,i}$ represents the cost in the *i*th parameter to carry out service *s* on network *n*, and $w_{s;i}$ represents the weight assigned to using the *i*th parameter to perform services, where the weight assigned may be related to a level of importance the user assigns to a particular service. If a user wishes to make handoff choices based on bandwidth and monetary cost for data service, the cost function can be calculated:

$$f^{n} = w_{b} \cdot \ln(\frac{1}{B_{n}}) + w_{c} \cdot \ln(C_{n}), \tag{2}$$

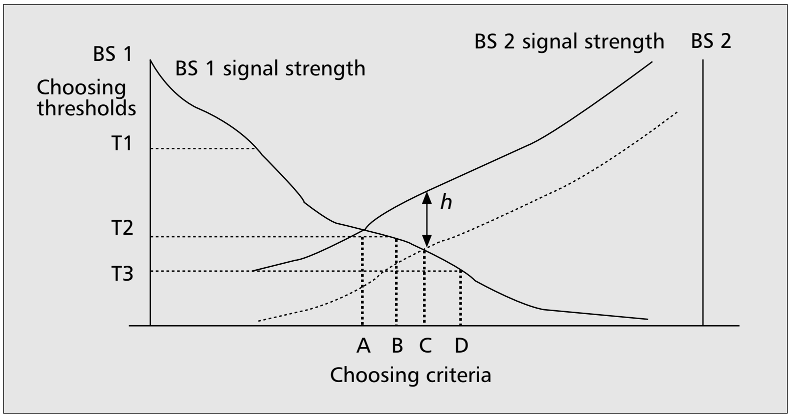
where B_n represents the cost in bandwidth for network *n* to support the handoff call, and C_n represents the monetary cost to support the handoff call. The weights assigned to each parameter are w_b for the bandwidth and w_c for the monetary cost, such that $\Sigma_i w_i = 1$ [9].

As new metrics and cost functions are developed, the implementation of a handoff policy becomes more important. Increased network and user interaction will increase the handoff latency. Thus, intelligent techniques must be developed to evaluate the effectiveness of the new algorithms, balanced against user satisfaction and network efficiency. Next, the current research in handoff policy design is explored.

Handoff Policy Design

Whereas handoff decision metrics help to determine where to hand off (i.e., which network should be chosen), the handoff policy represents the influence of the network on when the handoff occurs. The traditional handoff policy is illustrated in Fig. 2. Each vertical axis represents the signal strength received at the mobile node from each base station, represented as base station (BS) 1 and BS 2. The horizontal axis can represent either the time to hand off or the distance traveled from BS 1 to BS 2. The intersection point between the two curves represents an equal received signal strength from the two BSs, while the points labeled *A*, *B*, *C*, and *D* show the traditional handoff policies for cellular networks. The threshold technique requires the received signal strength from BS 1 to pass below the prescribed level before handoff can occur. In the

Traditional handoff does not allow for user selection of networks, and assumes that there is only one choice for access technology. In a heterogeneous environment, user choice is a desirable amenity.



■ Figure 2. *Traditional handoff policies.*

figure, the corresponding thresholds and handoff points are ($T1$, A), ($T2$, B), and ($T3$, D). A more robust technique uses both a threshold value and consideration of the difference in signal strength between BS 1 and BS 2, known as the *hysteresis*, h. The hysteresis technique is used to prevent ping-pong handoffs (i.e., unnecessarily repeated handoffs between BSs.) A similar handoff policy structure to Fig. 2 can also be applied to soft handoffs, wherein a signal strength above a certain threshold makes a BS a candidate for handoff, and a handoff decision is made according to all of the available candidates. The 4G multinetwork environment requires new handoff policies that reflect the updated criteria but continue to prevent detrimental effects such as ping-pong handoffs.

In [8] the authors consider handoffs between a WLAN and a cellular network. The handoff policy is different, depending on the nature of the user traffic. For example, since the data rate provided by WLAN (1–10 Mb/s) is much larger than that provided by the cellular network (9.2–200 kb/s), the handoff policy for non-real-time services is to attempt to use the services of the WLAN as long as possible. Thus, the preferred handoff point from the cellular network to the WLAN is the first time the signal strength in the WLAN reaches an acceptable level, while the handoff point from the WLAN to the cellular network is the last time the signal strength falls below the acceptable level. (The authors suggest several prediction schemes to determine when the signal strength has fallen low enough to switch.) On the other hand, for real-time applications, the preferred handoff point from the WLAN to the cellular network is the first time the signal strength degrades, while handoff from the cellular network to the WLAN is again the last time the signal strength reaches the acceptable level.

In [9], the authors develop an expression to prevent instabilities for a handoff point based on user and network interaction. A waiting period is added before handoff can occur, based on the handoff latency and a comparison of the cost metric for the current and target networks. The proposed handoff waiting period is calculated as

$$T_s = l_{\text{handoff}} + \frac{l_{\text{handoff}}}{e^{f_{\text{better}} - f_{\text{current}}} - 1}, \tag{3}$$

where l_{handoff} is the handoff latency, f_{better} is the cost function evaluated for the target network, and f_{current} is the cost function evaluated for the current network.

In the Wireless & Mobile Systems Laboratory, a dynamic vertical handoff algorithm has been developed to define and specify the dynamic values of a handoff cost function that are crucial to vertical handoff [10]. The algorithm also incorporates a network elimination feature to reduce the delay and processing required in the evaluation of the cost function. Finally, a multinetwork optimization protocol is performed to improve throughput for mobile nodes with multiple active sessions.

Dynamic Vertical Handoffs with Multinetwork Optimizations

The dynamic vertical handoff algorithm separates cost function factors into three different categories: QoS factors, weighting factors, and network elimination factors. The QoS factors, Q, represent the user- and network-specific constraints listed at the beginning of a later section, such as bandwidth requirements, power consumption, and monetary cost. The weighting factors, w, represent the importance of the particular constraint to the user or network. For example, battery life would be a weighting factor for power consumption, such that when the battery life is low, the power consumption required to hand off to a new network takes on a greater importance to the user. The final category is the network elimination factor, E, which reflects the ability of a network to guarantee certain constraints. For example, a chosen network may not be able to guarantee a certain minimum delay for real-time services.

Scenario 1: Cost Function with Network Elimination — The cost function values are calculated for each network available in the vicinity of the user. The network with lowest value becomes the handoff target. Specifically, the selection of the optimal network, *n_opt*, is based on

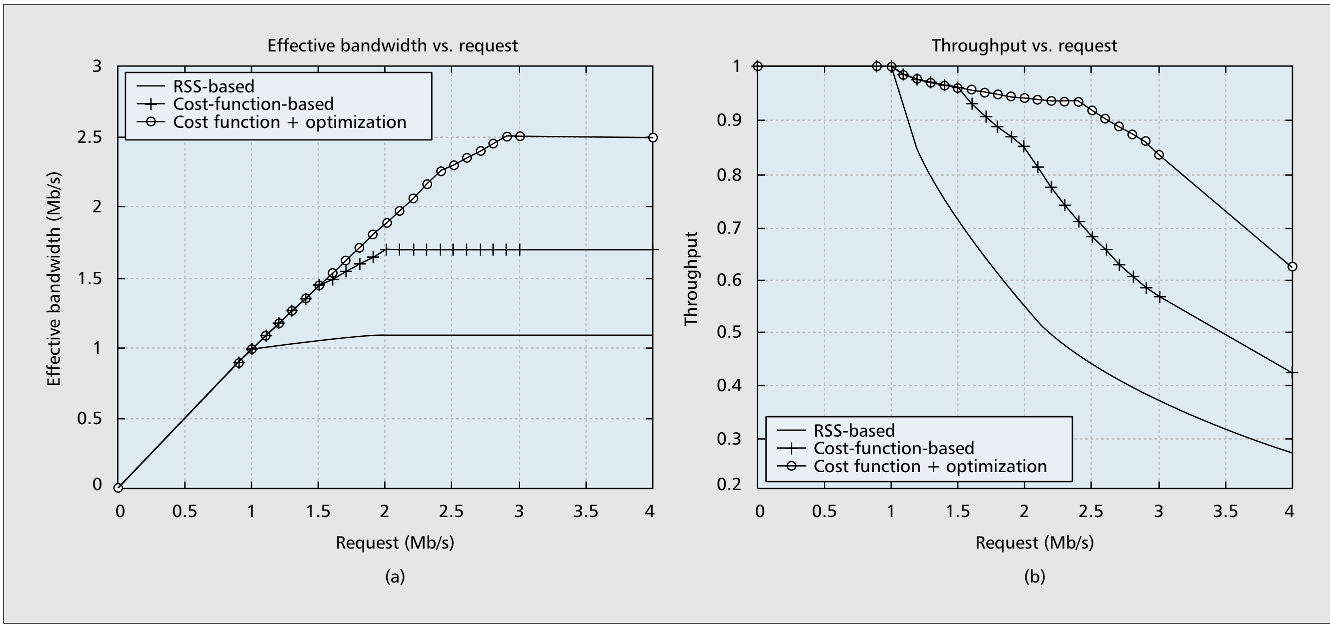
$$n\_opt = \underset{n}{argmin}\,(f^n), \tag{4}$$

where f^n is the handoff cost function for network n, and is calculated as

$$f^n = \sum_s \Big(\prod_i E^n_{s;i}\Big) \sum_j f_{s;j}(w_{s;j}) N(Q^n_{s;j}), \tag{5}$$

where $N(Q^n_{s,j})$ is the normalized QoS parameter, $Q^n_{s,j}$, representing the cost in the jth parameter to carry out service s on network n, $f_{s;j}(w_{s;j})$ is the jth weighting function for service s and $E^n_{s,i}$ is the ith network elimination factor of service s. The network elimination factor is represented by values of one or infinity, to reflect whether current network conditions are suitable for the mobile node's requested services. The multiplication over i excludes networks that are not qualified for service s, while the summation over s considers all services carried by one user. Finally, the summation over j calculates the total cost to network n.

Multi-network Optimization — A multinetwork optimization protocol is added to the dynamic vertical handoff protocol to achieve more efficient



■ Figure 3. *Cost function performance comparisons for a bandwidth-based QoS criterion.*

use of the available resources. In this case, an optimal network and service pair, n_s_opt, is calculated as

$$n\_s\_opt = \underset{\alpha,\beta^n}{argmin} \Big(\sum_{\substack{n\in\alpha,\\ s\in\beta^n}} \prod_i E^n_{s,i}\Big) \sum_j f_{s,i}(w_{s,i}) N(Q^n_{s,j})\Big), \quad (6)$$

where α is any subset of eligible networks in the vicinity of the user, and β^n is a subset of the services offered in network n.

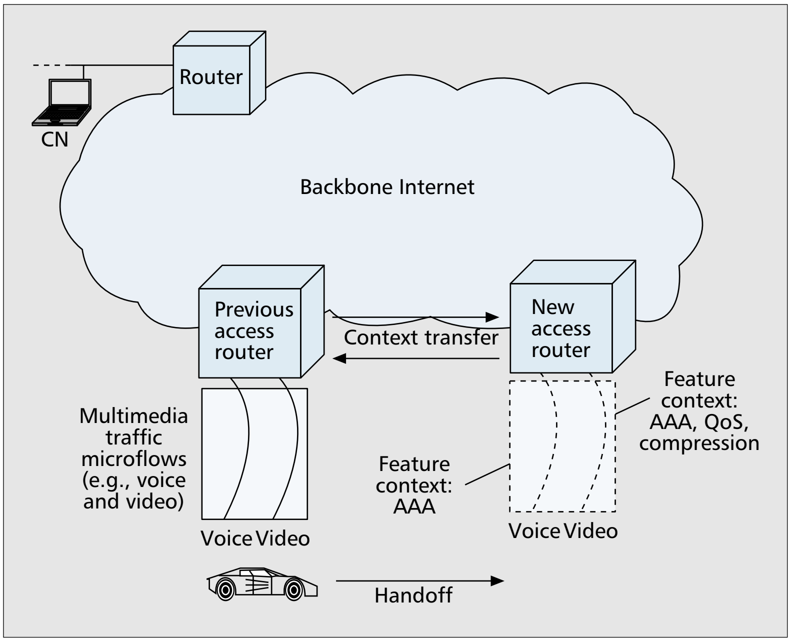
An initial performance analysis was performed to demonstrate the improvements in using a dynamic vertical handoff cost function to maximize a QoS constraint of user bandwidth. In the analysis, two scenarios were considered. In the first scenario, all flows at a single mobile node are handed off together. In the second scenario, the flows at a single mobile node can be handed off individually. The simulation assumed a three-cell overlay network, where the three cells represented three networks of different data rates (1 Mb/s, 1.5 Mb/s, and 2 Mb/s) that were available to each mobile node. A mobile node may request up to a certain amount of constant bit rate (CBR) service, as well as additional varying available bit rate (ABR) service. The results shown in Fig. 3 reflect the performance of the *traditional RSS handoff*, based on the strongest received signal strength, the *dynamic cost function-based handoff*, and the *dynamic vertical handoff cost function with optimization*. For CBR service requests of 0.9 Mb/s and ABR service requests that are variable, it is observed that the effective bandwidth at the mobile node is improved by dynamic cost function techniques, since the mobile node is allowed to choose to hand off based on its own bandwidth criteria. The new protocol also benefits from eliminating as choices networks that do not have a large enough bandwidth channel available. In addition, it is observed that the throughput is maximized by the dynamic vertical handoff cost function with optimization, since all of the mobile node's connections can be distributed among various networks according to bandwidth availability.

Handoff policy design is a process that must provide adaptability to user interactions, becoming more complex to account for a variety of handoff indicators. Further performance analysis is important for the evaluation of different techniques, with the goal of choosing the method with the lowest processing and signaling delay that can still perform the operations outlined above. Once the handoff decision is made, the mobile node must transfer its radio link to the new point of attachment. The next section describes current research in handoff link transfer techniques.

Vertical Handoff Radio Link Transfer Design

Handoff radio link transfer is the process of rerouting a mobile user's connection path to the user's new point of attachment. It requires the network to transfer routing information about the mobile user to the new (or target) access router for the proper forwarding of packets. As described previously, 4G networks are assumed to operate in an environment of multiple standards and networks. Thus, it is expected that 4G handoff link transfer will require additional context (i.e., user- and network-specific information) to enable the mobile node to move through different networks, while maintaining multiple data flows, and to choose among a variety of options for billing and service permissions, depending on the characteristics of the transmitted data and the current network. Research problems for the transfer of a mobile node's context to a target network include the following:

Formatting and interoperability. Handoff sig-



■ **Figure 4.** *Context transfer protocol operation.*

naling messages must be specified and formatted so as to be interpretable by the target network.

Performance. The desired goal of transferring the context of a mobile node to the new network is to minimize the delay in re-establishing the mobile node's traffic flows. However, if the context transfer delay were so large as to have the same effect of the complete re-establishment, or large enough to increase the overall handoff call dropping rate, the advantages of context transfer have been removed.

Quality of service (re)negotiation. For a mobile node being handed off (or handing off) to a new network, there may be a change in service quality for better or worse, depending on such factors as bandwidth availability, congestion, and interference.

As discussed earlier for handoff decision techniques, there may be a selection process at the network or mobile node to give certain flows a priority status, or to appropriately adjust the authorization and billing constraints. A mechanism is needed to allow for internetwork and/or inter-service-provider agreements to support fast intersystem roaming that avoids an unreasonable amount of internetwork signaling exchanges to validate or institute the adjustment in services.

Context Transfer

A context transfer protocol is designed to allow access routers to exchange state information regarding a mobile node's packet treatment [11]. State information affects packet handling at each access router, and includes, as examples, protocols for managing QoS guarantees, header compression, and authentication, authorization, and accounting (AAA). In the absence of context transfer, there may be large delays because of the network signaling required to re-establish QoS flows, re-authenticate the mobile user at the new router, and set the header compression algorithms. Large handoff delays can lead to reduced QoS, disruption of TCP operation, and dropped handoff calls. Furthermore, a context transfer between access routers reduces the need for control signaling over the unreliable and bandwidth-limited wireless channel.

The context transfer protocol being developed by the IETF is illustrated in Fig. 4. As the mobile node moves from its previous access router to the new access router, the corresponding information about each of the mobile node's microflows is forwarded between the access routers. Each microflow is categorized into feature contexts, which allow the network to indicate and provide the particular context information needed per microflow. For example, a particular mobile user may be downloading streaming video while conducting a voice transmission. The context required to continue the voice call may be authentication information only, while the context needed for the video service may be QoS and header compression, in addition to authentication. The initiation of the context transfer can be triggered by several events, such as a handoff due to a change in mobile node location, a request by the mobile node to change services, or by a network need to relieve congestion at a certain router. The mechanisms and others are discussed as part of the handoff decision algorithms earlier. We discuss here the message sequences required to initiate the context transfer.

The message sequences are illustrated in Fig. 5. They are separated into two categories: proactive and reactive. For the proactive case, the context may be transferred to the new access router before the mobile node attaches, so the context is immediately available before or during handoff. In the reactive case, the new access router explicitly requests the mobile node's context information, either as part of the handoff signaling or after the handoff is completed. The *Context Transfer Start Request* message is sent from the MN to an access router to begin a context transfer. The message data consists of the mobile node's previous care-of address, the previous access router's IP address, an authorization token for the mobile node, and a list of the requested context types. It may also include the mobile node's new IP address and the new access router's IP address, if known. The *Context Transfer Request* message is sent from the new access router to the previous access router, and provides the IP addresses of the MN and the new access router, the list of feature contexts to be transferred, and a token authorizing the transfer. The token is required to authenticate the mobile node requesting the transfer. Finally, the *Context Transfer Data* message is the response from the previous access router, containing the relevant context data. In addition to the data, it includes an authorization token that is computed over the binary context data, and provides the mobile node's previous care-of address and new care-of address (if known).

The contents of the messages, as well as the messages themselves, are still a subject of research within the SEAMOBY working group, but are provided here to give some indication of the operations required for context transfer. Other open research problems include the investigation of transport layer congestion control vs. overload risks from context transfer signaling,

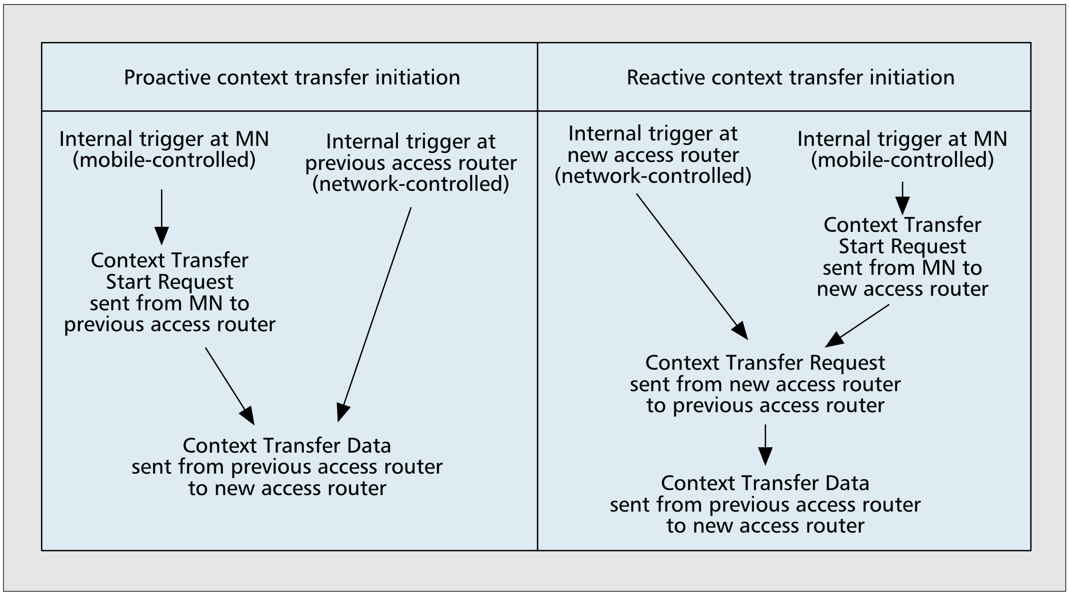


Figure 5. *Message sequences for proactive and reactive context transfer initiation.*

4G in its evolutionary and revolutionary context does not allow an exact vision of the future. However, if past evolutionary developments are an indication of the future, there is a need to promote technological adaptability and interoperability for the next generation of wireless communications.

and the problem of interdomain signaling, which requires additional efforts to establish security relationships from the previous to the new access router. The goal of ongoing research at the Wireless & Mobile Systems Laboratory in radio link transfer is to create a framework for intersystem signaling and authentication between different types of networks [12].

Conclusion

4G in its evolutionary and revolutionary context does not allow an exact vision of the future. However, if past evolutionary developments are an indication of the future, there is a need to promote technological adaptability and interoperability for the next generation of wireless communications. This article presents a tutorial on the design and performance issues for achieving an adaptable vertical handoff in a multinetwork 4G environment. Possible architectural components were described beyond 3G, and the advances and evolution of Mobile IP were envisioned as a significant step toward a 4G system based on all-IP networking technology. Traditional handoff protocols are not sufficient to deal with the goal of seamless mobility with context-aware services. However, there are advances being made via the efforts of the IETF and among researchers to create intelligent vertical handoff protocols to execute more complex handoff decision metrics and handoff policies, and context-aware handoff link transfers.

References

[1] I. F. Akyildiz *et al.*, "Mobility Management in Next Generation Wireless Systems," *Proc. IEEE*, vol. 87, no. 8, Aug. 1999, pp. 1347–84.

[2] W. W. Lu, "Fourth-Generation Mobile Initiatives and Technologies," *IEEE Commun. Mag.*, vol. 40, no. 3, Mar. 2002, p. 104–05.

[3] R. Berezdivin, R. Breinig and R. Topp, "Next-Generation Wireless Communications Concepts and Technologies," *IEEE Commun. Mag.*, vol. 40, no. 3, Mar. 2002, pp. 108–16.

[4] K. Pahlavan *et al.*, "Handoff in Hybrid Mobile Data Networks," *IEEE Pers. Commun.*, vol. 7, no. 2, Apr. 2000, pp. 34–47.

[5] D. B. Johnson, C. Perkins and J. Arkko, "Mobility Support in IPv6," IETF, May 2002, http://www.ietf.org/

[6] A. T. Campbell *et al.*, "Design, Implementation, and Evaluation of Cellular IP," *IEEE Pers. Commun.*, vol. 7, no. 4, Aug. 2000, pp. 42–49.

[7] R. Ramjee *et al.*, "HAWAII: A Domain-Based Approach for Supporting Mobility in Wide-Area Wireless Networks," *IEEE/ACM Trans. Net.*, vol. 10, no. 3, June 2002, pp. 396–410.

[8] J. Makela, M. Ylianttila, and K. Pahlavan, "Handoff Decision in Multiservice Networks," *Proc. IEEE PIMRC*, vol. 1, Sept. 2000, pp. 655–59.

[9] H. J. Wang, R. H. Katz and J. Giese, "Policy-enabled Handoffs Across Heterogeneous Wireless Networks," *Proc. IEEE Wksp. Mobile Comp. Sys. and Apps.*, Feb. 1999, pp. 51–60.

[10] F. Zhu and J. McNair, "Optimizations for Vertical Handoff Decision Algorithms," to appear, *Proc. IEEE WCNC '04*, Mar. 2004.

[11] Context Transfer, Handoff Candidate Discovery, and Dormant Mode Host Alerting (seamoby) Working Group, IETF, Apr. 2003, http://www.ietf.org/

[12] J. McNair, I. F. Akyildiz and M. Bender "An Intersystem Handoff Technique for IMT-2000 Systems," *Proc. IEEE INFOCOM 2000*, vol. 1, Mar. 2000, pp. 208–16.

Biographies

JANISE MCNAIR [M] (mcnair@ece.ufl.edu)received her B.S. and M.S. in electrical engineering from the University of Texas at Austin in 1991 and 1993, respectively, and her Ph.D. in electrical and computer engineering in 2000 from the Georgia Institute of Technology. She is currently an assistant professor in the Department of Electrical and Computer Engineering at the University of Florida, Gainesville, where she heads the Wireless & Mobile Systems Laboratory. She is active as an associate editor for *IEEE Transactions on Mobile Computing*, a reviewer for IEEE and ACM journals, a member of the organizing committee of the WMASH ACM workshop, and a session chair of several IEEE conferences. Her research interests are in the areas of handoff management, mobile user authentication, and medium access control.

FANG ZHU [StM] (fzhu@ufl.edu)received a B.S. degree from Beijing University of Posts and Telecommunications, Beijing, China, in 1999, and an M.S. degree from Colorado State University, Fort Collins, in 2002. She also worked as an engineer with China Telecom, Beijing, from 1999 to 2000. She is currently pursuing a Ph.D. degree in the Department of Electrical and Computer Engineering, University of Florida, Gainesville, where she is a research assistant in the Wireless & Mobile Systems Laboratory. Her research interests include mobility management, QoS, and resource management for next-generation wireless systems.