INTERNETWORKING WWANS AND WLANS IN NEXT GENERATION WIRELESS DATA NETWORKS

Hung-Yun Hsieh and Raghupathy Sivakumar

GNAN Research Group
School of Electrical and Computer Engineering
Georgia Institute of Technology
{hyhsieh,siva}@ece.gatech.edu

Abstract

In this paper we study the internetworking of WWANs and WLANs for next generation wireless packet data networks motivated by the capacity and coverage tradeoffs of the two networks. We compare different types of network architectures and propose a one-hop WWAN and multi-hop WLAN architecture to provide users with guaranteed service and smooth hand-offs. We then show through simulations that the proposed architecture coupled with a fine-grained packet scheduling algorithm achieves higher performance. Finally we identify several open research issues associated with the proposed architecture.

1 Introduction

With an increasing number of Internet users becoming mobile, wireless data networking technologies that facilitate tetherless communication have come under intense scrutiny and research. 2.5G wireless packet data systems such as GPRS (General Packet Radio Service) provide data rates that are a few orders of magnitude lower than those provided by the wireline Internet. In this context, the deployment of the third generation (3G) wireless systems is much anticipated to provide higher data rate. 3G wireless systems aim to offer both data and voice services (as opposed to 2.5G systems that overlay data networks over pre-existing voice infrastructures), and the data rates offered to users is expected to be as high as 384Kbps outdoors and 2Mbps indoors.

Parallel to the development of the 3G effort, several wireless local-area networking technologies such as IEEE 802.11 and ETSI HiperLAN have instigated proliferated use of WLANs at the edges of the Internet. The key advantage of WLANs over 2.5G and 3G wide-area systems is the significantly higher data rates (e.g. both 802.11a and HiperLAN/2 provide data rates of up-to 54Mbps). However, WLANs also suffer from a considerably lower coverage area. While wireless wide-area network (WWAN) cells have an approximate radius of 2-3 miles, WLANs typically have a radius of only a few hundred meters. The smaller coverage area thus precludes WLANs from being used as a ubiquitous wireless network technology. While a tremendous growth in the deployment of WLAN access-points can lead to extensive WLAN based coverage, such a deployment is infeasible because of cost and interference constraints.

An obvious solution to overcome the low capacity problem of wide-area wireless networks and the low coverage problem of local-area wireless networks is to perform a vertical hand-off [1] from the WLAN to the WWAN when a user moves out of the WLAN coverage area. Such a solution has also been proposed in several related works [2][3][4]. Specifically, the BRAN HiperLAN/2 attempts to integrate WLAN access within the same infrastructure that provides 3G wide-area access. The advantage of performing the vertical hand-off is self-evident: When the user is within the range of a WLAN access-point, a high data rate pertaining to that of the WLAN is provided. As the user moves away from the coverage area of the WLAN access-point, network connectivity is still seamlessly maintained, albeit at a significantly lower data rate.

Although the afore-mentioned approach serves as a simple solution to bridge the desirable properties of the two types of network environments, the key drawbacks of using such a simple approach are the capacity vs. coverage tradeoffs of the two network environments, and hence the service degradation the user perceives when a vertical hand-off is performed. While the curve (as shown in Figure 1) representing the capacity of a WLAN and the coverage of a WWAN is desirable at all instances, such a solution is obviously ideal but infeasible. However, the question of whether a more graceful degradation in the service can be realistically achieved when performing the vertical hand-offs is an interesting one. The answer to the question sets the context for the contributions of this work:

1. We propose a converged architecture that accommodates both wide-area and local-area networks within a single framework. The service provided to users undergoes graceful degradation when a hand-off is performed.
2. We identify several open research issues in the context of the proposed converged architecture. The research issues range from changes to the network infrastructure to changes in several network protocols including scheduling, medium access control, routing, and transport. While it is beyond the scope of this paper to provide elaborate solutions to these issues, we discuss possible directions of research.

The rest of the paper is organized as follows: Section 2 considers several alternative architectures possible and discusses the associated drawbacks and limitations. Section 3 describes the proposed converged architecture, and Section 4 presents the simulation results. Section 5 presents the open research issues within the converged architecture, and Section 6 concludes the paper.

2 Background and Motivation

In this section we consider six alternate architectures for providing wireless network access to mobile users, including a few converged architectures. For every architecture, we briefly discuss its drawbacks and limitations. We therein motivate the design of the proposed converged architecture presented in the next section. We use 3G and IEEE 802.11 as representative technologies for wide-area and local-area networks respectively for all our discussions henceforth.

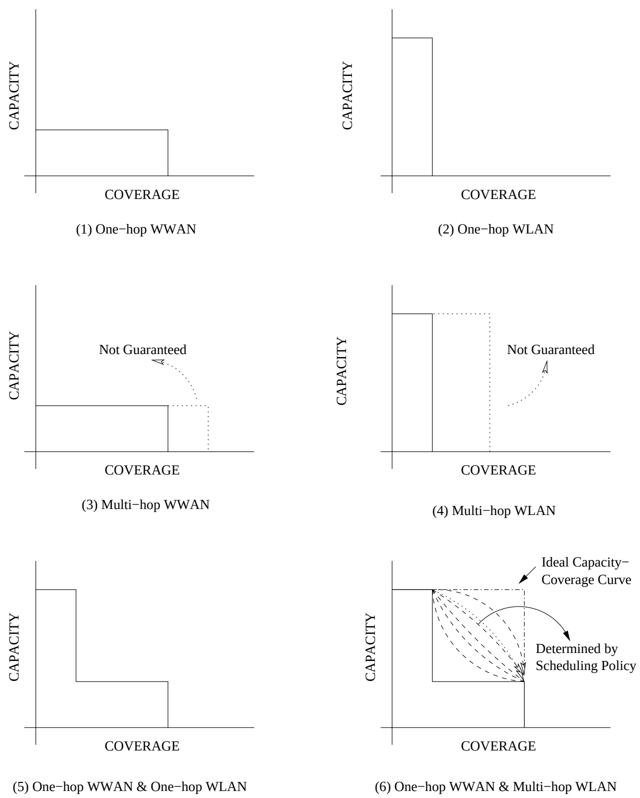

Figure 1: Illustrations of Capacity and Coverage Tradeoffs

1. **One-hop WWAN**: This is the default setup in which mobile users will have access to the 3G infrastructures. Each mobile user is directly connected to a 3G base-station. The drawback of such an approach is the limited data rates (up-to 384Kbps outdoors) that can be provided to users. A desirable property of the architecture is the extensive coverage area.
2. **One-hop WLAN**: This is the default setup in which mobile users will have access to WLAN infrastructures. Just as in 3G networks, the mobile users will have to be within the coverage cell of the IEEE 802.11 access-point. The IEEE 802.11a standard expected in the near future supports data rates of up-to 54Mbps, allowing users to experience near-wireline bandwidths. However, the reach of an 802.11 access-point is limited to about a few hundred meters precluding users from enjoying the high data rates in a ubiquitous fashion.
3. **Multi-hop WWAN**: Although several approaches have been presented in literature to adopt a multi-hop architecture for WWANs where mobile-stations participate in packet forwarding to the base-station, there are three limitations of this approach in the context of the problem addressed by this paper: (i) Using multi-hop routes to access the base-station is used primarily as a coverage enhancement mechanism such as reducing dead spot locations [5] and avoiding data rate degradation towards the boundaries of the cell [6]. Since WWANs already have a coverage area that is significantly higher than WLANs, using multi-hop routes does not considerably improve performance experienced by users within reach of WLANs. (ii) In a related work, we have demonstrated that not only can capacity not be increased by using multi-hop routes, but due to the distributed nature of a multi-hop environment, the capacity can in fact degrade when compared to a one-hop scenario. (iii) Transforming an already developed and in some case deployed one-hop infrastructure to a multi-hop infrastructure will be prohibitive cost-wise. For example, these proposed approaches require changes of channel allocation schemes such that additional channels can be allocated for multi-hop relaying. In this way, the bandwidth available to mobile users not sending traffic via multi-hop routes is further reduced.
4. **Multi-hop WLAN**: Using multi-hop routing in WLANs is motivated by two key advantages: (i) WLANs have a limited range, and using multi-hop routes can extend the coverage area of a WLAN, thus enabling users that are farther away from a WLAN access-point to enjoy higher data rates. (ii) WLANs are typically provisioned with the capability to operate in a distributed multi-hop mode. For example, the IEEE 802.11 defines the Distributed Coordination Function (DCF) mode of operation that enables mobile users to communicate in a multi-hop fashion. However, relying on solely a multi-hop WLAN environment can provide no guarantees to users in terms of connectivity and quality of service [7][8]. For this

reason, it is unlikely such an approach would be adopted as a feasible solution in the near future.

5. **One-hop WWAN and One-hop WLAN**: This architecture would comprise of the simple vertical hand-offs based approach briefly introduced in Section 1. When the user is within the range of the WLAN access-point, service is provided by the WLAN at higher data rates. When the user moves outside the WLAN coverage area, a hand-off is performed to the appropriate WWAN base-station, albeit at the cost of degrading the data rate provided. Such an approach has two disadvantages: (i) The performance degradation is significant and abrupt when a vertical hand-off occurs, and (ii) Since the coverage area of a WLAN is considerably smaller than that of a WWAN cell, it is more likely that a user is being served by the WWAN base-station than the WLAN access-point at any given instant, thus limiting average data rate experienced by the user.

6. **One-hop WWAN and Multi-hop WLAN**: This architecture would by far provide the most desirable properties when compared with the earlier considered architectures. A user inside the planned cell range of a WWAN base-station would be served by default in a multi-hop fashion through a WLAN access-point. If a multi-hop route of length $< k$ cannot be constructed to any WLAN access-point, a vertical hand-off is performed to the WWAN base-station. Such a scheme would increase the chances that a user is served at the higher data rates of WLANs. However, several issues would still persist with this approach: (i) At any given point in time, a user can strictly be served only through one of the networks. Consequently when operating in the multi-hop WLANs, a user may enjoy a wider area of high data rate but without guaranteed connectivity and quality of service. (ii) The decision to hand-off from a multi-hop WLAN environment to the WWAN is not trivial since hand-off initiation cannot be simply based on the received signal strength indicator (RSSI). While route unavailability (say due to a network partition) is one possible trigger for the hand-off, a more meaningful metric would be the throughput obtained through the multi-hop environment. Furthermore, in performing the hand-off care should be taken to avoid ping-pong switching. Therefore, current vertical hand-off schemes (and variations) proposed to internetwork WWAN and WLAN cannot be employed directly if guaranteed high data rates and a wide coverage area are to be achieved.

3 WWAN/WLAN Converged Architecture

We present a converged architecture that accommodates both the WWAN and WLAN infrastructures within the same framework. In sequence, we describe the entities involved, relationships between the entities, the different modes in which a mobile user can be served, and the underlying mechanisms.

The converged architecture consists of a super-position of the WLAN and WWAN infrastructures. The entities involved are the WWAN base-station and the associated core network [9], the WLAN access-point and associated distribution network, the mobile users within the WWAN cell, and the rest of the backbone Internet. Figure 2 illustrates the proposed architecture.

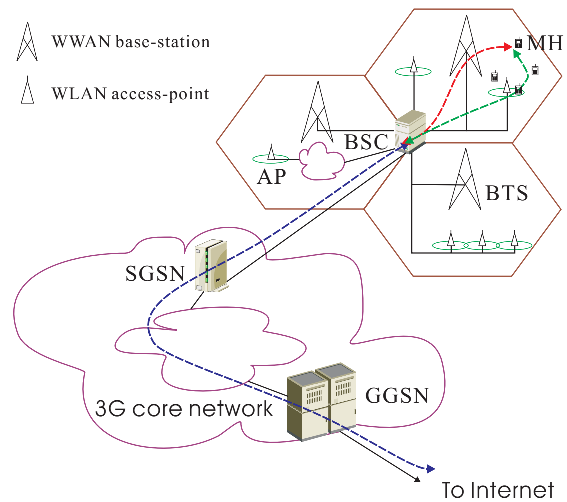

Figure 2: Converged Architecture for WWAN/WLAN

The WLAN access-points within the coverage area of a WWAN base-station are considered to be within the administrative control of the WWAN service provider. Connectivity, either via a dedicated network or the Internet backbone, is assumed between the WLAN access-points and the WWAN base-station. The mobile users are equipped with the radio interfaces for both the WWAN and WLAN environments.

Each mobile user can be in one of four states as far as connectivity within the architecture is concerned: (i) one-hop connected to the WWAN base-station, and multi-hop connected to the WLAN access-point, (ii) one-hop connected to the WWAN base-station, but not multi-hop connected to the WLAN access-point, (iii) multi-hop connected to WLAN, but not connected to WWAN, and (iv) not connected to either WLAN or WWAN. We assume a hop limit of k hops to determine whether or not a mobile user is connected to a WLAN access-point. The proposed architecture does not differentiate the scenario where the user is one-hop connected to the WLAN from the multi-hop connected scenario.

For ease of discussion, we assume that the mobile user is connected to both the WLAN access-point and the WWAN base-station. The mobile user's IP address[1] is associated (and managed) with only the WWAN, and both upstream and downstream traffic traverse through the WWAN base-station controller (BSC). We assume for the rest of the discussion that a mobile user's WWAN capacity subscription (or requirement) is *B* units. Since WLAN service is cheaper per unit bandwidth, the proposed architecture will attempt to provide the maximum possible capacity through the WLAN environment (traffic is directed via WLAN access-points). If the service through the WLAN network (say B_l) is less than *B*, the rest of the service (B-B_l) is provided through the WWAN environment. Any unused radio resource (e.g. time slots) at the base-station can be used to serve additional users, and we use this as the metric to characterize the performance enhancements of the proposed architecture in Section 4.

On the upstream, the mobile uses a unique scheduling policy to determine in a fine-grained manner which of the two networks to use for any given traffic. The scheduling policy used is imposed on a per-packet basis such that the mobile user will be able to attain its subscribed capacity with lower cost, despite the fluctuation in throughput in the WLAN environment (e.g. due to medium access delay or route failures). Any service requirement not fulfilled by the WLAN is handled by the WWAN. As mentioned earlier, even if the WLAN network is chosen to deliver some traffic, the traffic is still routed from the WLAN access-point to the backbone Internet through the WWAN core network. This precludes the need for the mobile to have multiple IP addresses (for the WWAN and WLAN environments), and supports the ability to distribute traffic on a per-packet basis between the two network environments. Note that such a fine-grained distribution comes at the cost of possible re-ordering at the static receiver. We address this issue in Section 5. Hand-offs between WLAN access-points are implicitly handled by the WLAN multi-hop routing protocols, and hand-offs between WWAN base-stations are handled as in a plain WWAN environment. Connection with the old WLAN access-point is lost if the mobile decides to hand-off to another WWAN cell. We identify the decision process involved in handing-off from one WWAN cell to another, even if connectivity is available with WLAN access-points of the old WWAN cell, as another research issue in Section 5.

On the downstream, when traffic arrives at the WWAN network, the BSC decides whether to forward it down the WWAN or through the WLAN. The decision is made based on the perceived quality of service over the two networks. As long as there is upstream traffic, the BSC keeps track of the WLAN access-point to use for downstream traffic. In the absence of upstream traffic, it is the mobile's responsibility to periodically probe for a route to the nearest WLAN access-point and the probe is delivered by the concerned WLAN access-point to the BSC for route update [10].

[1] We assume prevalence of all-IP networks both in WWANs and WLANs. It does not matter whether Mobile IP or IPv6 will be used in this context.

4 Simulation Results

4.1 Simulation Model

We use the *ns-2* [11] simulator with wireless extensions for all simulations in this paper. The channel model consists of a combination of the free space propagation model and the two-ray ground reflection model. Each node in the simulation is equipped with two network interfaces: one for WWAN operating at 2Mbps and one for WLAN at 11Mbps channel data rate. We use the IEEE 802.11 MAC protocol in the PCF (Point Coordination Function) mode for the WWAN model, and the DCF mode for the WLAN model[2]. Multi-hop WLAN routes are computed by using the DSR (Dynamic Source Routing) routing protocol. The WWAN interface uses optimal transmission power to reach the WWAN base-station, while the WLAN interface uses a fixed transmission power of 115mW (about 200m transmission range).

We use a 1600m by 1600m grid with 100 randomly distributed nodes as the network topology. Nodes are mobile and the mobility is modeled using waypoint movement model with 20m/s maximum speed and 0 pause time. All 100 nodes use either UDP or TCP to transport CBR traffic destined to the WWAN base-station.

4.2 Simulation Results

In this section, we present simulation results showing the performance improvement when multi-hop WLAN is used along with one-hop WWAN, especially with fine-grained packet scheduling mechanism. We use one WWAN base-station and several randomly distributed WLAN access-points over the network topology described before. The base-station covers the whole topology while the coverage of each access-point is only about 200m. All simulation results are obtained from averages of 10 samples with different seeds, each with a running time of 100 seconds. For lack of space we only present results when TCP is used; simulations using UDP show similar results.

[2] While we acknowledge the simulation model used for the 3G network is a simplification, we believe it does not undermine the conclusion of this paper since we are comparing the performance improvement with the use of IEEE 802.11, rather than modeling the performance of the 3G network per se.

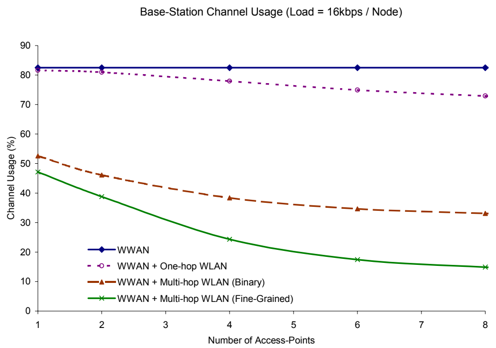

Figure 3: Base-Station Channel Usage

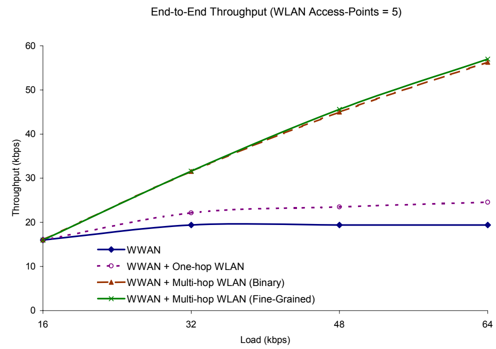

Figure 4: Per User Throughput

Figure 3 shows the percentage of channel usage at the WWAN base-station against the number of WLAN access-points, while providing 16kbps throughput to each of the 100 users. In a pure one-hop WWAN, channel usage is at least 80% since the channel rate is 2Mbps and all traffic is towards the WWAN base-station. As the number of access-points increases, in one-hop WWAN and one-hop WLAN architecture the base-station becomes less busy as seen in the figure. The improvement is, however, not significant due to the high mobility of users and limited coverage area of the access-points (which are randomly distributed). In contrast, using multi-hop WLAN helps reduce the channel usage at the base-station significantly. As discussed in Section 3 the "freed" resources can be used to serve more users. Furthermore, Figure 3 shows the fine-grained packet scheduling scheme can achieve higher performance than a binary switching scheme that involves explicit hand-offs between WWAN and WLAN.

When the number of WLAN access-points is constant, Figure 4 shows the network performance while the traffic load per user increases from 16kbps to 64kbps. It is clear that one-hop WWAN alone or with the use of one-hop WLAN does not provide users with throughput much higher than 20kbps due to the channel bottleneck at the base-station. Using multi-hop WLAN, on the other hand, can provide a per user throughput of more than 55kbps with the presence of 5 access-points. Note the fine-grained and binary schemes achieve the same throughput in this figure because a large amount of traffic is diverted to the WLAN access-points and the WWAN base-station is still under-utilized for both schemes.

5 Open Research Issues

The following are some of the open research issues in the proposed converged architecture:

1. **Infrastructure**: The architecture currently assumes the existence of connectivity between the WLAN access-points and the WWAN base-station within a WWAN cell either through a dedicated network or the Internet backbone. Dedicated networks are already in use to bridge distribution networks of different network providers [4]. However, using the Internet backbone can save on costs, but at the expense of possible longer delay and less reliable service in the wired domain.

2. **Hand-offs**: There are three types of hand-offs possible in the converged architecture: (i) WWAN-WWAN hand-offs: A mobile user who loses connectivity with the old WWAN base-station *X* and has connectivity with a new WWAN base-station *Y* might still have good connectivity with a WLAN access-point *A* belonging to the domain of *X*. Although the existence of WLAN connectivity can potentially reduce the impact of WWAN hand-offs, the decision process involved in such a hand-off needs to be investigated further. (ii) WLAN-WLAN hand-offs: Hand-offs between WLAN access points within the same WWAN coverage cell occur implicitly by the underlying routing protocol switching routes. (iii) WWAN-WLAN hand-offs: Such hand-offs are more a function of the service fluctuations a mobile experiences through the two types of networks. A mechanism that can provide fine-grained packet level distribution of traffic needs to be developed. Although we have developed such an algorithm, we are still investigating several issues such as the switching overheads, efficient data structures, etc.

3. **Routing**: The architecture does all routing through the BSC in order to avoid IP address changes at the mobile and to facilitate easier billing and monitoring. Requiring the mobile to have two IP addresses (one each for the WWAN and WLAN networks) will preclude a single connection from being multiplexed over the two types of networks (since the source will be able to recognize only

one of the IP addresses). However, the associated cost of routing traffic from the WLAN to the BSC needs to be profiled and justified. Routing on the multi-hop path between the WLAN access-point and the mobile also can be done intelligently given the presence of static access-points and an almost omniscient WWAN base-station. We believe that such a routing protocol can provide significantly better performance than general purpose ad-hoc routing protocols[3].

4. **Transport**: The multiplexing of traffic in a fine-grained manner over the two types of networks can have negative consequences at the transport layer because of re-ordering of packets. Existing transport protocols such as TCP will suffer greatly as they are designed for a single-pipe environment. This issue is, however, not unique to the proposed architecture. Several works have been proposed to address bandwidth aggregation (striping) through multiple network interfaces, especially for wireless networks [12][13]. Notwithstanding, these approaches are developed in a different context and cannot be used directly in the proposed architecture. We are currently involved in developing a transport protocol that can efficiently operate over multiple physical pipes and provide the application with the aggregate rates of the physical pipes. Since users in a mobile environment communicate with a proxy [14] for most of the time, deployment of such a transport protocol will be a feasible option for the network provider.

6 Conclusions

In this paper, we study the internetworking between WWANs and WLANs for next generation wireless packet data networks. We identify the motivation for such internetworking due to capacity and coverage tradeoffs. We also discuss different types of network architectures based on WWANs, WLANs and their integration thereof. To provide users with guaranteed service and smooth hand-offs over a wider area, we propose a one-hop WWAN and multi-hop WLAN architecture. In this architecture, the users always attempt to connect to WLAN access-points whenever possible, and use the WWAN base-stations only when service obtained via WLANs is less than the desired capacity. With a fine-grained packet scheduling algorithm the users will be able to enjoy the subscribed capacity while using only minimum WWAN base-station radio resources. We also identify several open research issues associated with the proposed architecture.

[3] Similarly, although simple CSMA/CA based approaches such as IEEE 802.11can be used for the medium access control layer in multi-hop WLANs, the proposed architecture allows a more sophisticated protocol be employed with the assistance of the base-station and access-points.

References

[1] M. Stemm and R. Katz, "Vertical Handoffs in Wireless Overlay Networks," *ACM Mobile Networks and Applications*, vol. 3, no. 4, pp. 335-350, 1998.

[2] K. Pahlavan et al, "Handoff in Hybrid Mobile Data Networks," *IEEE Personal Communications Magazine*, vol. 7, no. 2, pp. 34-47, Apr. 2000.

[3] S. Helal, C. Lee, Y. Zhang, and G. Richard III, "An Architecture for Wireless LAN/WAN Integration," in *Proceedings of IEEE WCNC*, Chicago, IL USA, Sept. 2000, pp. 1035-1041.

[4] ETSI TR 101 957 V1.1.1, "BRAN; HIPERLAN/2; Requirements and Architecture for Internetworking Between HIPERLAN/2 and 3rd Generation Cellular Systems," Aug. 2001.

[5] G. Aggelou and R. Tafazolli, "On the Relaying Capacity of Next-Generation GSM Cellular Networks," *IEEE Personal Communications Magazine*, vol. 8, no. 1, pp. 40-47, Feb. 2001.

[6] 3G TR 25.924 V1.0.0, "3GPP TSG-RAN; Opportunity Driven Multiple Access," Dec. 1999.

[7] H.-Y. Hsieh and R. Sivakumar, "Performance Comparison of Cellular and Multi-hop Wireless Networks: A Quantitative Study," in *Proceedings of ACM SIGMETRICS*, Boston, MA USA, June 2001, pp. 113-122.

[8] S. Xu and T. Saadawi, "Does the IEEE 802.11 MAC Protocol Work Well in Multihop Wireless Ad Hoc Networks?" *IEEE Communications Magazine*, vol. 39, no. 6, pp. 130-137, June 2001.

[9] 3GPP TS 23.002 V5.4.0, "3GPP TSG-SSA; Network Architecture," Oct. 2001.

[10] A. Campbell et al, "Design, Implementation, and Evaluation of Cellular IP," *IEEE Personal Communications Magazine*, vol. 7, no. 4, pp. 42-49, Aug. 2000.

[11] The Network Simulator ns-2, available via http://www.isi.edu/nsnam/ns/, June 2001.

[12] D. Maltz and P. Bhagwat, "MSOCKS: An Architecture for Transport Layer Mobility," in *Proceedings of IEEE INFOCOM*, San Francisco, CA USA, Mar. 1998, pp. 1037-1045.

[13] A. Snoeren, "Adaptive Inverse Multiplexing for Wide-Area Wireless Networks," in *Proceedings of IEEE Globecom*, Rio de Janeireo, Brazil, Dec. 1999, pp. 1665-1672.

[14] P. Sinha, N. Venkitaraman, R. Sivakumar, and V. Bharghavan, "WTCP: A Reliable Transport Protocol for Wireless Wide-Area Networks," in *Proceedings of ACM MOBICOM*, Seattle, WA USA, Aug. 1999, pp. 231-241.