Low Latency Handoff for Wireless IP QOS with NeighborCasting

Eunsoo Shim, Hung-yu Wei, Yusun Chang, and Richard D. Gitlin
Columbia University
500 West 120th Street
New York, NY 10027

***Abstract*- This paper introduces a fast handoff mechanism, NeighborCasting, for use in wireless IP networks that utilize neighboring Foreign Agent [FA] information. NeighborCasting is based on the policy of utilizing, or perhaps even wasting, wired bandwidth between Foreign Agents, while minimizing *rf* (radio frequency) bandwidth exchanges, in order that handoff latency is minimized. We demonstrate that the handoff latency is substantially reduced, while the typical overhead is minimally increased. Handoff latency is minimized by initiating data forwarding to the possible new foreign agent candidates (i.e., the neighbor foreign agents) at the time that the mobile node initiates the link-layer handoff procedure. NeighborCasting builds upon the Mobile IP handoff procedure by adding a small number of additional message types. The handoff mechanism is a unified procedure for inter-, intra-domain and inter-technology (eg, LAN/WAN or TDMA to CDMA) handoffs and provides flexible choices to the network, while maintaining transparency to the mobile node. The neighbor FA discovery process is a distributed and dynamic mechanism, and the fast handoff schemes are scalable and reliable.**

I. INTRODUCTION

It is important to architect emerging wireless IP networks to support real-time media applications such as Voice over IP [VoIP], as well as data applications. One of the key issues for VoIP networks is providing the Quality of Service (QoS) consistent with the user expectations. This is recognized as the single biggest challenge in providing high-quality voice on wired IP-based networks. In wireless networks, one of the principal additional factor affecting QoS is minimizing service disruption during handoffs of the mobile nodes. It was shown that packet loss or packet error over wireless links can cause significant throughput degradation in TCP applications, owing to the TCP flow control algorithm [2]. While buffering and forwarding packets to the new base station or attachment point from the old base station could be used to reduce packet loss due to handoff, this procedure can introduce unacceptable delay to real-time media applications such as VoIP. Therefore, it is important to minimize the handoff latency, which is defined as the period when the mobile node cannot receive application traffic during handoffs.

One could wonder whether the mobile node (MN) or the network could know or predict the new foreign agent (FA) before data transport from the old FA to the MN is disrupted. If this is possible, the network can set up data forwarding to the new FA while the MN is still communicating with the old FA and thus reduce the handoff latency significantly. Today, this is possible in IS-95 (CDMA) networks where the *soft handoff* procedure permits the MN to simultaneously receive signals from the old and new base stations. Thus the MN can inform the old base station of the identification information of the new base station. The old FA can learn the IP address of the new FA from the identification of the new base station and map it to the IP address of the new FA with the aid of a directory server. Third-Generation (3G) wireless networks, based on Wideband CDMA (W-CDMA) will likely have similar capabilities.

But this capability is neither available nor easily feasible in many current and emerging packet-based wireless networks (eg, GSM, GPRS, 802.11). It is not desirable to impose such a capability as a requirement for Mobile IP, considering the complexity of predicting the new FA or the diversity of the wireless link technologies. While it is difficult to know or predict the new FA exactly, it is not too difficult to find a reasonable set of candidate FAs (i.e., neighbors) that are likely to be the new FA after a handoff. If we allow pro-active data forwarding to these prospective FAs just before, or at the very initiation of, the handoff process, then we can achieve very low handoff latency, without introducing significant complexity in the network or MN.

Based on this idea, we have developed a fast handoff scheme, *NeighborCasting* that substantially reduces handoff latency with minimal overhead. *NeighborCasting* is built upon Mobile IP [8], that is, the motion detection and data forwarding setup schemes of Mobile IP are kept, but no longer contribute to handoff latency.

In this paper we first present our design goals and how each FA finds its neighbor FAs. Then the handoff procedure is described. Following are experiment results focused on the traffic overhead due to *NeighborCasting* and handoff latency, and lastly a summary of the paper.

II. DESIGN GOALS

We pursue the following characteristics for the fast handoff algorithms: Keep the framework of Mobile IP, low handoff latency, scalability, reliability, minimize traffic overhead on the air interface, and support for inter-technology handoffs. Since Mobile IP has been already developed and widely accepted, it is an excellent framework to build upon. Various wireless technologies are already deployed and since IP does not impose any constraints on either network topologies or on the underlying technology combination, it is preferable that a fast handoff mechanism does not bring in such constraints either. Therefore a fast handoff mechanism should also seamlessly support inter-technology handoffs. A fast handoff mechanism should scale up in terms of number of mobile nodes, number of base stations/cells, and traffic for mobile nodes. Reliability is one of the best advantages of IP and supporting mobility should not weaken the reliability of IP. A fast handoff mechanism should support fast handoff for inter-domain handoffs as well as intra-domain handoffs, or support large domains if the mechanism is applied only for intra-domain handoffs. Since bandwidth in the wireless links is a scarce resource, it should be used efficiently and it is preferable to minimize signaling overhead over the *rf* air-interface.

III. NEIGHBORCASTING MECHANISMS

A. Neighbor FA Discovery Mechanism

Here we describe how FAs discover their neighbor FAs in a dynamic and distributed fashion in *NeighborCasting*. The key idea is that each FA can learn about its neighbor FAs from the MNs since the MNs move around and come from the neighbor FAs. That is, the MN keeps the address of the old FA and transmits this identity to the new FA every time it hands over from one FA to another FA. Then the new FA learns about one of its neighbor FAs and notifies the neighbor (the old FA) about itself. This way a FA can learn all of its neighboring FAs as time goes by and many handoffs occur. Figure 1 shows the procedure through which each FA discovers its Neighbor FAs.

The first step is that the MN includes the information about the old FA in the registration message sent to the new FA. Thus the new FA is informed of the old FA as one of its neighbor and then it sends a Neighbor FA Notification message to the old FA. Each FA maintains a Neighbor FA table. This Neighbor FA information is of soft state, that is, each entry expires after the Neighbor Timeout Period has passed without any new notification for the Neighbor FA entry.

One could wonder why we wouldn't use a standard routing protocol among the FAs to establish the Neighbor FA table. We notice that the Neighbor FA table reflects the *cell* topology rather than the wired network topology. Even if the two cells are adjacent to each other, the FAs of the cells could belong to different domains or remote subnets, which makes it rarely practical or feasible to find the relationship automatically through a protocol running only in the wired network. The discovery mechanism proposed above would be far simpler than such a protocol.

Our proposed discovery mechanism is executed in a distributed fashion; that is, there is no central controller or database. The size of the Neighbor FA table is proportional to the number of adjacent or overlapping cells, which is practically bounded to be a moderate number. In overall, the mechanism is robust and scalable.

B. Handoff Procedure

In general, the Layer 2 [L2] protocol stack of the MN will be involved in the L2 handoff**,** whether the L2 handoff is network initiated or mobile initiated. We assume the Mobile IP protocol stack is notified of the impending L2 handoff. The L2 handoff notification is a local event, that is, when there is L2 handoff, the L2 protocol stack handling the L2 handoff in the MN sends notification to the Mobile IP protocol stack by a method specific to the implementation. It could be a callback function call or any other method. The L2 protocol stack just needs to provide a proper API. When and how the L2 protocol stack generates the notification depends on the L2 handoff algorithm. What is required is just a short period of time for the Handoff Notification message between the L2 handoff decision and actual disconnection of the link between the MN and the old FA. Then the Mobile IP protocol stack generates the Handoff Notification message and sends it to the old FA. The handoff procedure in this scenario is depicted in Figure 2. It shows 3 new messages are added to realize *NeighborCasting*.

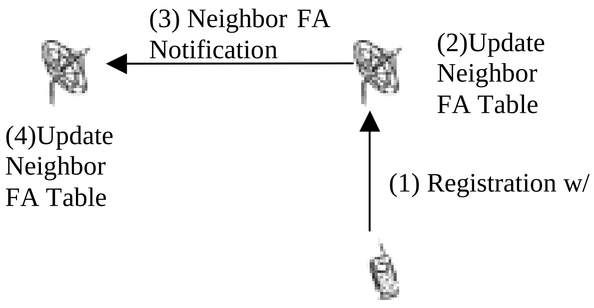

Figure 1. Neighbor FA Discovery Procedure

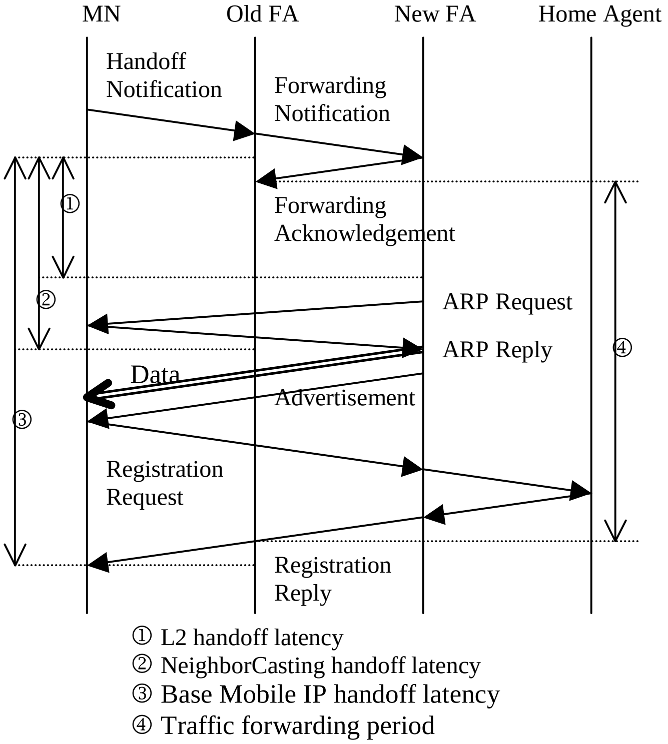

Figure 2. NeighborCasting Message Flow

We consider only break-before-make L2 handoffs in this paper [eg, as in GSM and in inter-technology GSM/802.11/Bluetooth systems]. We assume also the foreign agents are co-located with the base stations to simplify the presentation. If the time is not allowed in a certain L2 radio technology, the MN will send the Handoff Notification message to the old FA through the new base station as soon as it finishes the L2 handoff. The network should allow these messages to go through the firewall if any. If the L2 handoff is network-initiated and there is no time between when the MN is informed of the L2 handoff and the link between the MN and the old FA (or old base station) is terminated, the old FA can initiate the IP-level handoff procedure by sending the Forwarding Notification messages. Whether or not the FA will initiate the procedure, without waiting for the Handoff Notification from the MN, should be conveyed to the MN as a field in the agent-advertisement message.

As soon as the old FA gets a Handoff Notification message from a MN, it reads its Neighbor FA table and sends the Neighbor FAs the Forwarding Notification messages. Then the old FA starts forwarding data to the neighbor FAs. Note that it is required that the user traffic goes through the FA where the MN is attached. This may require using care-of-address located at FA rather than co-located care-of-address. When the new FA receives data packets forwarded by the old FA the first time, the packets contains the home IP address of the MN as the destination address and the new FA needs to know the L2 address of the MN to forward the packets to the MN. In some situations such as the 802.11 Wireless LAN, the new FA can forward IP packets to the MN by relying only on the MAC address of an interface of the MN. Since the MAC address is the same before and after the handoff, the old FA can provide the MAC addresses of the interfaces of the MN to the new FA and the new FA can forward IP packets to the MN using the provided MAC addresses. Then the new FA does not have to rely on the ARP (Address Resolution Protocol). Clearly it would save time for the new FA to start forwarding data immediately to the MN. The back-off algorithm of ARP can cause delay in finding the MAC address and thus longer handoff latency. So it is necessary to adjust the ARP request policy to remove the possibly long back-off time for the air-interface. In IS-95 or GSM networks, PPP is used for IP packet transport**,** and several options have been provided for the behavior of PPP with Mobile IP in [11]. According to [11], we believe that the proper behavior with *NeighborCasting* is to inform the FA of the home IP address of the MN as a part of IP configuration over PPP. Then after the PPP channel is established, the new FA knows where to forward the IP packets for the MN. We assume in Figure 2 that L2 handoff includes PPP channel setup if PPP is used. If the Forwarding Notification contains a valid MAC address, the new FA will try to use it. Otherwise it should try an ARP request to find the link layer address of the MN.

All the forwarded data packets to a neighbor FA are discarded if the FA is not the new FA. The old FA forwards the data to the neighbor FAs only when it cannot forward the packet to the MN. **So there is no duplicated data traffic over the air interface.**

The key point is that the MN can receive data shortly after the L2 handoff, without waiting for the agent advertisement messages or finishing the Mobile IP registration process. Figure 2 shows the case when the MN starts receiving data before the agent advertisement message. We can see that **the handoff latency of NeighborCasting can be comparable to that of L2 handoffs.** Handoffs using *NeighborCasting* are unified procedures for inter-, intra-domain and inter-technology (eg, LAN/WAN or TDMA to CDMA) handoffs and provide flexible choices to the network while maintaining transparency to the mobile node.

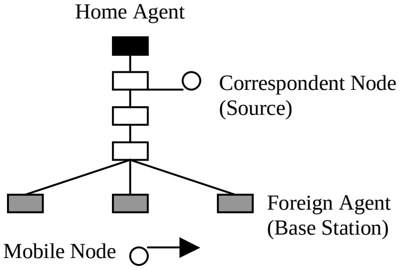

Figure 3. Network topology for the simulation on ns2

IV. PERFORMANCE

We evaluated the performance of *NeighborCasting* in two aspects: handoff latency and traffic overhead. Shortening handoff latency is a major goal of *NeighborCasting* and traffic overhead is a major cost to pay. We confirmed *NeighborCasting* results in dramatic handoff latency reduction with marginal traffic overhead compared to base MobileIP.

A. Handoff Latency

We measured the handoff latencies of the base Mobile IP and *NeighborCasting* with the topology shown at Figure 3 over the network simulator, ns2 [13]. The home agent (HA), the correspondent node (CN) and the three FAs are all connected through wires that are of duplex 10Mbps bandwidth, 2ms delay, and DropTail queuing algorithm. The wireless network is IEEE 802.11 wireless LAN with cells of radius 400m each and the cells are contiguous along a straight line but disjoint. The MN moves straight at 10m/sec crossing the touching point of the two contiguous cells. The CN is sending UDP packets at constant bit rate whose packet interval is 3.75ms. The handoff latency of Mobile IP was defined as the period between when the air-link with the old FA is disconnected and when the MN receives Mobile IP registration reply from the new FA. The handoff latency of *NeighborCasting* was defined as the time between when the air-link with the old FA is disconnected and when the new FA establishes a mapping entry of the IP address of the MN to the L2 address of the MN. Figure 4 shows the result of our simulation where the handoff latency of Mobile IP increases almost linearly as the agent advertisement period increases, while that of *NeighborCasting* is almost constant regardless of the agent advertisement period. Note that L2 handoff latency is zero in the simulation and thus the handoff latency of *NeighborCasting* is actually the time for arrival of a packet at the new FA the first time after the MN enters the cell of the new FA plus the time for ARP request/reply processing and transmission. The ARP request is sent once every time when the new FA does not know the L2 address of the newly received packet's destination, and there is no back-off mechanism in the simulator. We did not forward the MAC address of the MN to the new FA in the Forwarding Notification message in the simulation. If it were

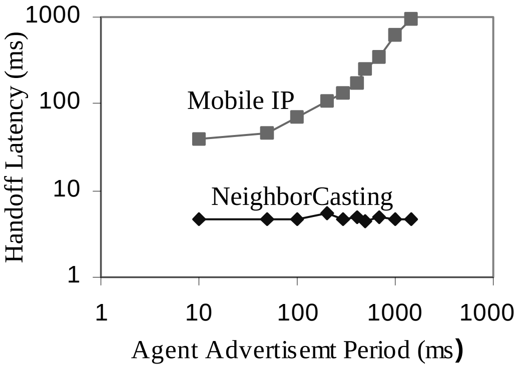

Figure 4. Handoff latencies at different ad. period

implemented, we could save the time for ARP request/reply. The handoff latency of *NeighborCasting* is around 4.5 to 5.5ms and there is dramatic improvement of handoff latency in *NeighborCasting* compared to the base Mobile IP. Note that the Mobile IP handoff latency is not decreased linearly at very short agent advertisement periods due to the time taken for registration with the home agent.

B. Traffic Overhead

Traffic overhead due to duplicating and forwarding traffic to the neighboring FAs is an obvious cost of *NeighborCasting*, while achieving very low handoff latency is its advantage. One of the main concerns of *NeighborCasting* is the traffic overhead. The total traffic overhead to the network is proportional to the handoff frequency and the forwarding period at each handoff.

We consider the case that each traffic source sends UDP packet to a mobile node at constant data rate *c* continuously. Then the amount of data forwarded to the neighboring FAs during a handoff of a mobile node is $n \cdot \tau \cdot c$ bytes where *n* is the number of neighboring FAs (e.g.: n=6 in hexagonal cell) and τ is the period during which data is forwarded to the neighboring FAs. Thus the total amount of forwarded data from a cell for a unit time period is $n \cdot \tau \cdot c \cdot \lambda$ where λ is the handoff rate and the total amount of data sent by the traffic sources to the mobile nodes in the cell for the same unit time period is $m \cdot c$ where *m* is the total number of active mobile nodes in the cell.

The fluid flow model [6,12] is widely used to analyze the cell boundary-crossing problem. According to the model, the handoff rate λ is

$$\lambda = \frac{\rho v L}{\pi} \qquad (1)$$

where λ is the handoff rate or cell boundary-crossing rate (1/sec), ρ is the active mobile node density (1/m^2), v is the mobile moving speed (m/sec), and L is the cell perimeter (m). The traffic overhead ratio ξ_f, which is defined as the number of bytes forwarded to the neighboring FAs divided by total number of bytes sent by the sources, is

$$\xi_f \equiv \frac{n \cdot \tau \cdot c \cdot \lambda}{m \cdot c} = \frac{n \lambda \tau}{m} = \frac{\rho v L n \tau}{\pi \cdot m} \qquad (2)$$

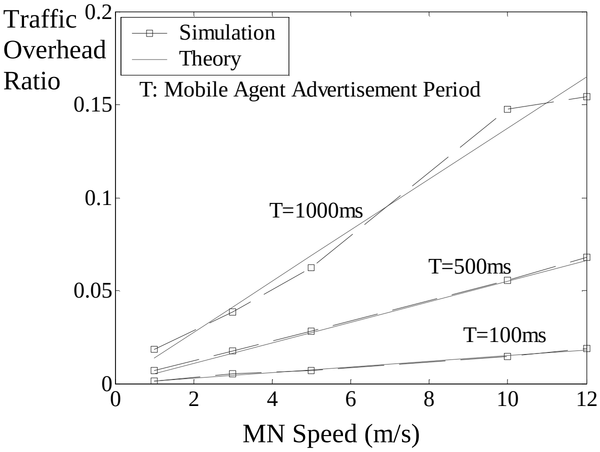

Figure 5. Traffic overhead ratio

Denoting r as the cell radius, for a hexagonal cell scenario where n=6, and L=6r we get the following.

$$\begin{aligned} \rho &= \frac{m}{cell\ \ area} = \frac{m}{\frac{3\sqrt{3}}{2} r^2} \\ \Rightarrow \lambda &= \frac{\rho v L}{\pi} = \frac{4 m v}{\sqrt{3}\, r \pi} \\ \Rightarrow \xi_f &= \frac{24\, v\, \tau}{\sqrt{3}\, r \pi} \end{aligned} \qquad (3)$$

The data-forwarding period τ, based on Figure 2, is modeled as the sum of the several components shown below.

$$\begin{aligned} \tau &= time\_L2 + u + time\_ad\_MN \\ &+ RTT\ (\text{FA - MN}) + time\_reg\_FA \\ &+ RTT\ (\text{HA - FA}) + time\_reg\_HA \end{aligned} \qquad (4)$$

where *u* is the waiting time for Mobile IP agent advertisement after L2 handoff is over, time_L2 denotes layer 2 handoff time, time_ad_MN denotes advertisement processing time at mobile node, time_reg_FA denotes registration processing time at foreign agent, time_reg_HA denotes registration processing time at home agent, RTT(FA-MN) denotes round trip time from foreign agent to mobile node, and RTT (HA-FA) denotes round trip time from foreign agent to home agent.

u is modeled as a uniformly distributed random variable with expected value T/2. If we represent T as Mobile IP advertisement period, the average value E [u] = T/2.

We simulated traffic overhead of NeighborCasting using Matlab and compared the result with the theoretical results listed above. In the user mobility model used, a mobile node moves in a constant speed *v*. It changes its direction after time t, which is a geometrical-distributed random variable. The new direction θ is chosen uniformly over [0, 2π]. Foreign agents are located in center of hexagonal cell with cell radius *r*. All active mobile nodes are receiving continuous constant bit-rate UDP flow. The serving FA is the nearest FA to mobile node. There is zero overlapping area between neighboring cells. Whenever the mobile node moves to a new cell, a handoff occurs. *NeighborCasting* forwards data flows to all the neighboring Foreign Agents. The handoff frequency is dependent on mobile moving speed *v*, cell radius *r*, and active mobile node density in a cell. There are several terms affecting the multiple user data-forwarding period τ. One of the configurable parameters in the Mobile IP is the waiting time for Mobile IP agent advertisement, *u*. Simulations with different T parameter (E [u]= T/2) are investigated.

In Figure 5, we used 200m-radius hexagonal cell topology as used in the ns2 simulation and also the handoff latency depicted in Figure 4. We can see that the traffic overhead increases as the MN speed increases and also when the agent advertisement period increases. The reason is that we have more frequent handoffs when the MN speed is faster and the forwarding period is longer when the agent advertisement period is longer. We find that the traffic overhead falls within the order of $10^{-2} \sim 10^{-1}$ in the pico-cell environment when it is defined as the ratio of total forwarded bytes over total transmitted bytes. Note that traffic overhead ratio 10^{-1} occurs around MN speed 8m/s or higher which is a somewhat extreme case in the pico-cell environment.

V. RELATED WORK

Fast handoff has been one of the research topics continuously investigated in various contexts. The Virtual Connection Tree [1] and Anticipatory Handoff Control [7] were proposed in the context of wireless ATM networks. Anticipatory Handoff Control [7] was proposed applying the concept of virtual tree connection and neighborhood to, in essence, narrow down the area covered by the virtual connection tree into the neighboring cells and the current cell. Anticipatory Handoff Control is proactive for partial connection establishment but reactive for data forwarding whereas *NeighborCasting* is proactive for data forwarding. Cellular IP [4], HAWAII [10], Proactive Handoff [3], and the Fast Handoff of El-Malki et al. [5] were proposed in the context of wireless IP networks. In Cellular IP, the gateway router plays the role of the FA for the visiting domain, and the intra-domain routers form a spanning-tree, where the root node is the gateway router and all the leaf nodes are the routers attached to the base stations. Then data path from the gateway router to each MN is kept as per-hosting routing entries in the intra-domain routers. The data path is maintained dynamically by intermediate routers monitoring the data packets or signaling messages from the MN. The requirement for spanning-tree brings in the reliability and scalability issue. The last two proposals involve initiating data forwarding to the new FA while the MN is attached to the old FA like *NeighborCasting* but both of them are focused on the case where the IP address of the new FA is known to the old FA. Nrouting [14] is also applying the concept of neighborhood. It adds an IPv6 header extension that contains the IP addresses of the neighboring access routers to the user packets as destination candidates and thus the user packets are forwarded to the right access router in the end at the time of handoff. Our neighbor FA discovery mechanism can be used to provide the access router the IP addresses of the neighboring access routers in Nrouting. The additional IP header including IP addresses of neighboring access routers at every user packet consists of traffic overhead.

VI. SUMMARY

We presented a new fast handoff mechanism for wireless IP, *NeighborCasting*, whereby the old FA *proactively* forwards data to the neighbor FAs while the MN or the network processes the L2 and Mobile IP handoff procedures. *NeighborCasting* is an extension of the Mobile IP protocol and is a unified mechanism for intra-domain, inter-domain and inter-technology handoffs. Even though it consumes more bandwidth in the wired network, the traffic overhead is limited and does not grow as the wireless network grows. Also it is shown that the traffic overhead is marginal in typical cases. The information about the neighbor FAs is collected in dynamic and distributed fashion. Also, the fast handoff mechanism does not require any special radio technology features, modification of routing mechanism in the wired network, or technology specific supports. It does not require any per-domain network entity such as a gateway FA or a gateway router that raises reliability and scalability concerns, and, most significantly, NeighborCasting can achieve a latency approaching that of the Layer 2 handoff. Consequently, NeighborCasting is a very attractive fast handoff mechanism for isochronous mobile IP-based networks.

REFERENCES

[1] Acampora, A.,Naghshineh,M.. “Control and Quality-of-Service Provisioning in High-Speed Microcellular Networks”, IEEE Personal Communications Magazine, 1994 Vol 1, No. 2.

[2] Balarkrishnan,H., Seshan,S., Katz,R.. “Improving Reliable Transport and Handoff Performance in Cellular Wireless Networks”, ACM Wireless Networks 1(4), December 1995.

[3] Calhoun, P., Hiller,T., Kempf, J., McCann,P., Pairla,C., Singh,A., Thalanany, S.. “Foreign Agent Assisted Hand-off”. draft-calhoun-mobileip-proactive-fa-03.txt (work in progress), November 2000.

[4] Campbell, A.,Gomez, J., Wan, C-Y., Kim,S., Turanyi, Z., A. Valko. “Cellular IP”. draft-ietf-mobileip-cellularip-00.txt (work in progress), January 2000.

[5] El-Malki, K., Soliman, H.. “Fast Handoff in Mobile Ipv4”. draft-elmalki-mobileip-fast-handoffs-03.txt (work in progress), September 2000.

[6] Jain,R., Puri,A., Segupta,R.. “Geographical Routing Using Partial Information for Wireless Ad Hoc Networks”, IEEE Personal Communications, February 2001.

[7] Lee, K.. “Adaptive Network Support for Mobile Multimedia”, Proc. Mobicom 95, Berkeley CA USA

[8] Perkins,C.. “IP Mobility Support”. Request for Comments (Proposed Standard) 2002, Internet Engineering Task Force, October 1996.

[9] Perkins, C.E., Johnson, D.. “Route Optimization in Mobile IP” (work in progress). draft-ietf-mobileip-optim-09.txt, February 2000.

[10] Ramjee,R., LaPorta,T., Thuel,S., Varadhan,K., Salgarelli, L.. “IP micro- mobility support using HAWAII”. draft-ietf-mobileip-hawaii-01.txt (work in progress), July 2000.

[11] Solomon, J., Glass, S.. “Mobile-IPv4 Configuration Option for PPP IPCP”, RFC2290, February, 1998

[12] Thomas,R., Glibert,H., Mazziatto,G.. “Influence of the Mobile Station on the Performance of a Radio Mobile Cellular Network”, Proc. 3rd Nordic Seminar, Denmark 1988.

[13] The Network Simulator - ns2, http://www.isi.edu/nsnam/ns/

[14] Yegin, A.,Parthasarathy, M.,Williams, C., Mobile IPv6 “Neighborhood Routing for Fast Handoff “(work in progress), draft-yegin-mobileip-nrouting-00.txt, September 2000