# TR94-020
# Virtual Cell in Mobile Computer Communications

Kyungshik Lim and Yann-Hang Lee
Computer and Information Sciences Department
University of Florida
Gainesville, FL 32611

## Abstract

This paper describes a *virtual cell* concept for the transmission of IP datagrams in mobile computer communications. A virtual cell consists of a group of physical cells whose base stations are implemented by remote bridges and interconnected via high-speed datagram packet-switched networks, supporting host mobility at the data link layer. Thus, as far as the IP layer is concerned, it appears as if the communication between two mobile hosts in different physical cells were taking place directly as in the same physical cell. It eliminates the necessity of IP-level protocols for mobility management which may interfere with the conventional IP protocol in a practical sense, and achieves a logically flexible coverage area according to mobility and data traffic patterns among physical cells. Virtual Cell Protocol (VCP) which consists of handover, address resolution, and data transfer modules is designed based on the distributed hierarchical location information of mobile hosts. To measure data which will be used as parameters to a performance model for the virtual cell, we implement VCP on a simulated environment.

Key Words: Wireless Networks, Mobile Computer Communications, IP

## 1 Introduction

As computers become more powerful and portable with the appearance of high-speed wireless interfaces, there is an increasing demand on the provision of mobile computer communications in TCP/IP environments. A fixed host in internets is always assigned an IP address which not only serves a unique identifier used by the higher layer protocols but also represents the current location of it. An inherent problem from the transmission of IP datagrams in mobile computer communications is that when a mobile host (MH) migrates, its IP address is only valid as the identification information but not able to represent the current location information. To solve this problem, various researches and developments have focused on how to integrate the functionality of host mobility into the IP layer, preserving the compatibility with the conventional IP protocol [1, 2, 3, 4, 5, 6].

Although these mobile host protocols vary on how to represent and maintain location information for efficient tracking of MHs, the techniques of using the IP options and IP encapsulation have been considered. Typical examples of the first technique are Virtual Internet Protocol (VIP) [2, 3] and the mobile host protocol using IP Loose Source Routing option[4]. IP-within-IP (IPIP)[1] and Internet Packet Forwarding Protocol[6] use the second technique. Even though the details of these mobile host protocols are different, we consider one representative mobile host protocol for each technique, VIP and IPIP, to illustrate common features and problems of integrating host mobility into the IP layer.

In VIP, the network layer is divided into two layers; the VIP layer resides on top of the normal IP layer. The VIP packet header is implemented as an option of the IP packet header. An MH keeps its permanent IP address used at the VIP layer for its identification and acquires a transient IP address used at the IP layer for its current location information when it migrates. If the MH is in its native network, both addresses are the same. After acquiring a transient IP address, the migrating MH sends its native network a notification packet whose header contains the permanent and transient IP addresses. As the packet propagates to the native network, every network entity including MHs, mobile support gateways (MSGs), and even intermediate gateways on the path snoops the header information and stores address conversion information to a cache. In the same way, the header information of all data packets in transit is also used by the network entities to maintain their caches. If the source has the cache entry for the destination, the source executes address conversion before sending a packet. The existing routing protocols of the IP layer can then correctly deliver this packet to the destination. Otherwise, the source assumes that the destination resides in its native network and sends the packet accordingly. As the packet traverses to the native network of the destination, if an intermediate gateway has the cache entry for the destination, the gateway executes address conversion and forwards the packet to the current location of the destination.

Unlike VIP, intermediate gateways in IPIP are not involved in the support of host mobility but merely in transport service. The source MSG encapsulates a network control packet for mobility management or an IP datagram from an MH into another IP datagram whose source and destination addresses specify the communicating MSGs, and transmits it over an internet. The existing routing protocols of the IP layer then correctly deliver it to the destination MSG. Thus, MSGs consider internets as a full mesh of logical point-to-point links to interconnect them. A mobile network consists of a number of MSGs, each of which maintains the location information of its own MHs. Every MH in the mobile network is assigned a unique IP address but the network part of the IP address is the same. When the MH migrates, a forwarding pointer is set from the previous MSG to the new MSG for location tracking. If the MSG has no location information for a specific MH, it broadcasts the other MSGs in the same mobile network an inquiry message asking who has the MH. However, IPIP also requires a transient IP address when the MH visits a foreign mobile network.

Regardless of which technique is used, the integration of host mobility into the IP layer reveals several implications. First, underlying networks differ widely in their network size, bandwidth, protocol, and packet size so that they or some of them may not meet performance needs for the support of host mobility such as rapid migration and tracking of MHs. Moreover, because they are usually under different administrative controls, efficient network management and optimization for the global mobile network may not be easy tasks.

Second, intermediate gateways may cause some performance problems no matter they are involved in the support of host mobility or not. If they are involved in, as in VIP, they must be modified or replaced to understand VIP so that the benefit from using existing internets as transport networks is diminished. Furthermore, because intermediate gateways have to snoop every packet in transit to maintain location information, and examine every data packet in transit to try to perform address conversion, the protocol processing and memory loads at intermediate gateways may severely affect overall performance. Even in case that they are not involved in, as in IPIP, because they usually implement packet switching operation in software, protocol processing time coupled with possible network delay between multiple hops of IP gateways may greatly affect cell switching latency.

Third, each physical cell administered by an MSG is in principle assigned an IP network address because every MH and intermediate gateway is a network entity in TCP/IP environments. It means that locating the MH in a different physical cell necessitates a different network address to represent its current location. Some undesirable features of IP-level mobile host protocols essentially comes from this fact. VIP requires a large amount of the transient IP address space which becomes a very scare network resource as internets are rapidly growing. IPIP uses one permanent IP address for each MH but relys on the broadcast inquiry among MSGs when the location information is not available by a forwarding pointer, restricting the scalability of IPIP to a local area.

Fourth, although IP-level mobile host protocols based on options keep the compatibility with the conventional IP protocol in their specifications, in a practical sense they may interfere with it because most existing fixed hosts and intermediate gateways do not implement the IP options and their implementations to support the IP options may not feasible in the near future. In addition, the current efforts to provide IP multicasting protocols and connection-oriented protocols require IP-level mobile host protocols to be compatible.

As high-speed, connectionless, packet-switched networks are emerging to extend LAN-like performance across a wide area, we believe that they can greatly affect the support of host mobility in mobile computer communications. Examples are ATM networks[13][14] and Switched Multimegabit Data Services(SMDS) networks[9][10] which are subnetworks providing an MAC service across a wide area in a large interconnected network. To take advantage of these networks, it is desirable to push the interconnection level of base stations (BSs) down to the data link layer, supporting host mobility in a distributed fashion. In Section 2, we explains the virtual cell concept to solve the problems of supporting host mobility at the IP layer. The vir-

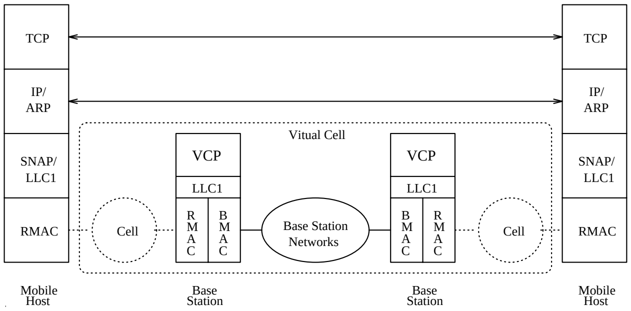


Figure 1: The Virtual Cell Concept

tual cell architecture is described in Section 3. Section 4 gives the structure of the distributed hierarchical location information, the detailed procedures of VCP, and the outline of the VCP implementation. The comparison of VCP with existing protocols and discussion are given in Section 5, and we conclude with future works in Section 6.

## 2 Virtual Cell Concept

Consider the transmission of IP datagrams between two MHs within a physical cell. Because radio links provide the broadcast ability, the source can deliver IP datagrams to the destination using the normal Address Resolution Protocol (ARP) protocol. Thus, the broadcast ability of radio links itself eliminates the necessity of locating MHs, resulting in no mobility management protocols at the IP layer and no modifications to the normal ARP protocol. To apply the similar rationale to MHs crossing physical cell boundary, we propose the virtual cell concept, as depicted in Figure 1.

A virtual cell is a logically extended cell of physical cells whose BSs are implemented by remote bridges and interconnected via a base station network. Because remote bridge BSs enable to preserve the interconnection level of physical cells at the data link layer, the whole virtual cell is assigned an IP network address. In order to make it possible for MHs in a virtual cell to directly deliver IP datagrams using the conventional IP and ARP protocols, a mobility management protocol, called Virtual Cell Protocol (VCP), is implemented at the data link layer of BSs. Each virtual cell is also allocated an ARP/Location server (ALS) which implements VCP and supports the normal gateway function with other virtual cells or fixed networks. VCP supports handover, address resolution, and location tracking for datagram delivery, based on the distributed hierarchical location information of BSs and the ALS, as described in detail in Section 4.

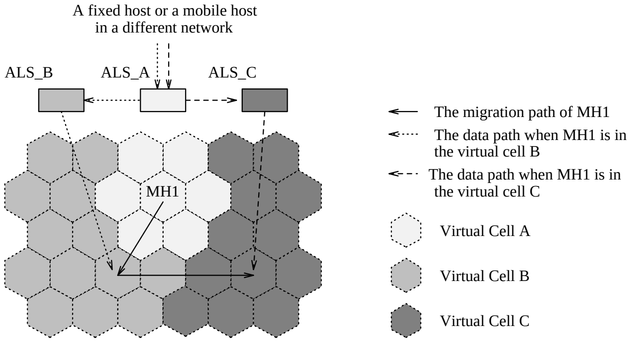


Figure 2: The Data Path Between Virtual Cells

It should be noted that the identification and location information of an MH is represented by two different data link layer addresses in the virtual cell, Radio network MAC (RMAC) address and Base station network MAC (BMAC) address, respectively. Note that the BMAC address is not only used as the location information of the MH but also as the identification information of a virtual cell. For example, the SMDS addresses can be formatted with a similar structure used for the North American Numbering Plan to represent the geographic semantics. This separation eliminates the necessity of acquiring a temporary address to represent the MH's current location as it migrates. In addition, unlike IP-level mobile host protocols, the IP network address also represents the correct location information of the MH while it moves to a different physical cell in the virtual cell. This is possible because the whole virtual cell is assigned an IP network address and host mobility is shielded from the IP layer.

Even when the MH migrates between virtual cells, the IP network address is still valid as the near-location information in the virtual cell environment. Figure 2 shows the data path from a fixed host or an MH in a different network to $MH_1$ which migrates from its native virtual cell $A$ to virtual cell $B$ to virtual cell $C$. Whenever $MH_1$ crosses the boundary of virtual cells, a one-hop forwarding pointer is maintained from $ALS_A$ in its native virtual cell to the ALS of its current virtual cell, $ALS_B$ or $ALS_C$, by the handover procedure of VCP. Note that the forwarding pointer is not maintained at the IP layer but at the VCP layer. When a fixed host or an MH in a different network wants to send an IP datagram to $MH_1$ in virtual cell $B$ or $C$, the existing routing protocols of the IP layer correctly deliver it to $ALS_A$. Then $ALS_A$ redirects it at the VCP layer to the corresponding ALS, $ALS_B$ or $ALS_C$, which keeps track of the exact location of $MH_1$ in its virtual cell.

Thus, every MH resides in its native virtual cell from the IP layer's viewpoint but practically may reside in an adjacent virtual cell because of a one-hop forwarding pointer from the native virtual cell to the current virtual cell maintained at the VCP layer. It means that the combination of the existing routing protocols of the IP layer and the forwarding pointer between virtual cells at the VCP layer gives the same effect that we have a fully duplicated location information among virtual cells for the global mobile network.

To configure neighboring and geographically distributed physical cells into virtual cells, the underlying transport network should provide high-speed datagram packet-switched services and the group addressing capability at the data link layer. The flexible configuration allows network managers and designers to customize the logical coverage area of the virtual cell to their requirements according to host mobility and data traffic patterns among physical cells.

Consider an environment that a number of physical cells are deployed in a metropolitan area. The cell size in the urban area is usually smaller than that in the suburban area in order to accommodate a larger population of mobile users. Even though user mobility is inherently unpredictable, there could be a high possibility that the daily routine of mobile users is usually confined to several physical cells in the urban area. Because of a relatively small cell and large population of mobile users in the urban area, it becomes very important to achieve the fast location tracking of mobile users and the small cell switching latency between physical cells, mitigating the effect of a large volume of mobility management information. Thus, the physical cells covering the urban area could be combined into one or a few number of virtual cell(s) for efficient mobility management and data delivery. As another example, we can consider a campus environment consisting of several campuses that are geographically distant. In order to handle both of the intracampus and intercampus traffic efficiently, the physical cells covering these campuses can be combined into a virtual cell in spite of their distance. The same rationale can be also applied to a business environment where a number of distant companies belong to one organization.

## 3 Virtual Cell Architecture

Because the virtual cell is a logical concept, the same transport network may support several virtual cells simultaneously. As shown in Figure 3, there are two roles of the transport network: base station network and backbone network. The base station network is used to interconnect a number of BSs and an ALS to build a virtual cell, and the backbone network is used to interconnect among virtual cells and fixed networks. Note that the base station network utilizes both point-to-point and multicast communications, while the backbone network does only point-to-point communication.

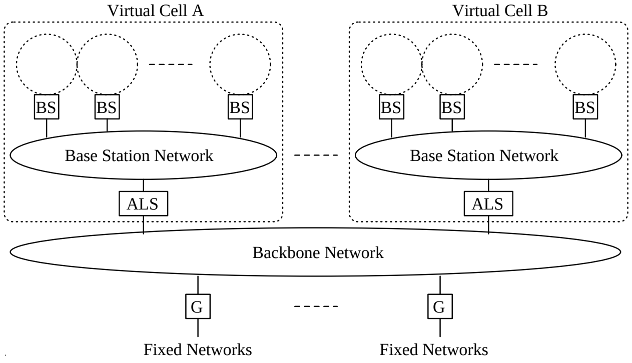


Figure 3: The Virtual Cell Architecture

### 3.1 Base Station Network

The base station network should meet some requirements from two different viewpoints:

- *Application's viewpoint.* IP operates under the assumption that packet arrivals are independent and unpredictable. If IP is coupled with the connection-oriented base station network, one connection establishment and release may be needed for each individual packet over a fixed link between a pair of BSs. Although a bundle of packets could be transmitted over one connection by some multiplexing techniques, when communicating MHs migrate to different physical cells with the connection, the problem becomes complex and even intractable. In addition, because location tracking has inherently connectionless properties, the base station network should support the datagram delivery. The current ubiquitous coverage of TCP/IP networks and applications also requires that the base station network be able to cover a wide area.
- *Mobility management's viewpoint.* Combined with the distributed location information of MHs among BSs, host mobility is a main cause of the

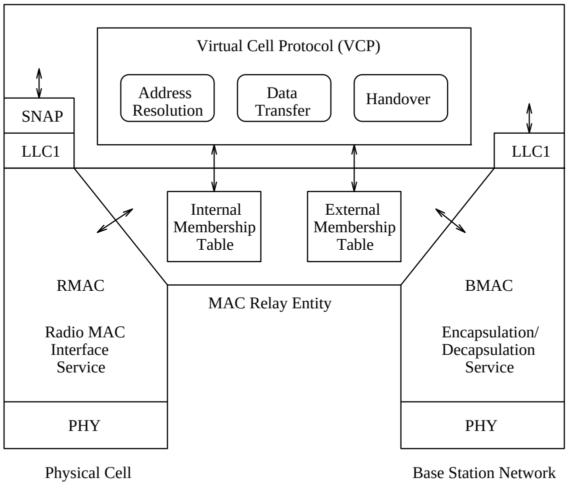


Figure 4: The Base Station Architecture

inconsistency. In order to maintain the consistency of the distributed location information efficiently, the base station network should support the multicast ability. Furthermore, because the network control information for mobility management is expected to be greatly increased as the cell size becomes smaller, the base station network should have high-bandwidth not so as to affect the normal data traffic greatly.

We consider SMDS as an example with these characteristics. Both individual and group addressing capabilities combined with a set of addressing-related service features (e.g., source address validation, source and destination address screening) enables to create a number of logical networks over SMDS. Every BS and ALS is attached to a Subscriber Network Interface (SNI) and individually addressed. At the same time, a group address identifies all BSs in a virtual cell, ensuring that each group address identifies uniquely only one set of individual addresses. The number of SNIs to be supported by a switching system is at least 256 SNIs and the future number is up to 4096 SNIs within a Local Access Transport Area. Considering that the range of a physical cell is usually 3-5 km in the urban area or 10-15 km in the suburban area, the capacity of a very few number of switching systems may be enough to support base station networking within a metropolitan area.

## 3.2 Base Station

The communication architecture in a BS is shown in Figure 4. The internal port connected to a physical cell implements an RMAC entity, while the external port connected to the base station network implements a BMAC entity. The MAC relay entity translates the information format between the physical cell and the base station network. The protocol identifier field of both the RMAC and BMAC frame headers is set for LLC type 1 Unnumbered Information format. For example, the Higher Layer Protocol Identifier field of the SMDS Interface Protocol L3_PDU used as BMAC frames is set to 1. The LLC Service Access Point of the RMAC frame is set for Subnetwork Access Protocol (SNAP), while that of the BMAC frame is set for VCP.

The VCP header has three fields: VCP type, VCP length, and RMAC address. The VCP type is set to one of *ADDRESS*, *DATA* and *HANDOVER* to specify the address resolution, data transfer and handover modules of VCP, respectively. Depending on the VCP type, the information field of the VCP frame has a different format. The *DATA* type indicates that the information field has an IP datagram, and the information fields for the *HANDOVER* and *ADDRESS* types are given in the next section. The VCP length indicates the length of the VCP header. Note that the RMAC address is only used with the *DATA* type.

Each BS maintains a partial location information of the virtual cell in its Internal Membership Table (IMT) and External Membership Table (EMT). IMT keeps track of all MHs currently roaming in its physical cell. The IMT entry for an MH has two fields, the IP and RMAC addresses, each of which is used as an identifier but has a different role. The IP address is used as an identifier by the higher network layers as usual, but the network part of the IP address represents the MH's native virtual cell to which it initially belongs. This IP address obtained by the handover procedure is only needed for fast address resolution, as described in the next section. On the other hand, the RMAC address is the other identifier used by VCP for the support of host mobility at the data link layer. EMT has a relatively small number of entries for MHs roaming in different physical cells of the same virtual cell. The EMT entry for an MH has three fields, the IP, RMAC and BMAC addresses, respectively, of which the BMAC address represents the MH's current network address.

If the source MH wants to send an IP packet to the destination MH whose native virtual cell is the same, the source performs ARP and sends the IP packet using the destination RMAC address. On the other hand, if the source wants to send an IP packet to the destination whose native virtual cell is different, the source does not perform ARP and sends the IP

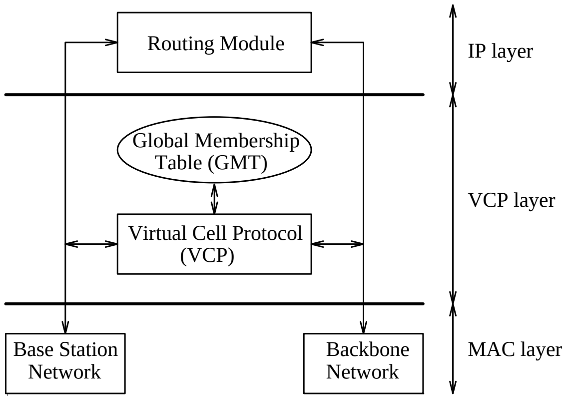


Figure 5: The Communication Architecture in an ARP/Location Server

packet using the RMAC address of its current BS, which is obtained by a beacon message. When the BS receives an SNAP/LLC/RMAC frame from the internal port, it checks the Protocol Identification (PID) field of the SNAP header. If the PID field is set for ARP, the ARP packet is sent to the address resolution module which performs Virtual Cell Address Resolution Protocol (VCARP). VCARP achieves the IP-to-RMAC address binding for the destination and also distributes the location information of the source and destination. If the PID field is set for IP, the IP packet is sent to the data transfer module with the destination RMAC address. If the destination RMAC address is the BS's RMAC address, the data transfer module sets the RMAC address field of the VCP header to the BS's RMAC address and relays a VCP/LLC/BMAC frame to the ALS without searching IMT and EMT. Otherwise, the data transfer module keeps track of the destination MH's current location using IMT and EMT. If the corresponding entry is not found, the data transfer module transmits to the ALS a VCP/LLC/BMAC frame whose VCP header contains the destination RMAC address.

### 3.3 ARP/Location Server

Each ALS maintains Global Membership Table (GMT) for MHs roaming in its virtual cell, as shown in Figure 5. The entry format of GMT is the same as that of EMT. There are two types of frames from the backbone network for data transfer. One is the VCP/LLC/BMAC frame redirected from other virtual cells that maintain one-hop forwarding pointers at the VCP layer, and the other is the SNAP/LLC/BMAC frame from fixed networks or other virtual cells. The first type of frame is sent to the data transfer module which identifies the destination BS by searching GMT with the RMAC address of the VCP header, and transmits it to the base station network. The second type of frame is, however, directly sent to the routing module which extracts the IP packet. If the network part of the destination IP address indicates the same virtual cell, the packet is sent to the data transfer module of VCP where the IP-to-RMAC address binding occurs and a VCP/LLC/BMAC frame is transmitted to the base station network or to the backbone network, depending on whether the destination MH resides in this native virtual cell or moved to another virtual cell, respectively. If the IP packet is in transit, it is transmitted to the next-hop router.

The VCP/LLC/BMAC frame from the base station network is always sent to the VCP module. If the frame carries a VCARP packet, the address resolution module performs the IP-to-RMAC address binding with GMT and sends it back to the base station network. If the frame carries an IP packet with a BS's RMAC address of the virtual cell in the VCP header, the data transfer module dose not search GMT and directly pass the IP packet to the routing module where the next-hop router is determined. Otherwise, the data transfer module searches GMT with the destination RMAC address and then transmits a VCP/LLC/BMAC frame back to the base station network or to an adjacent virtual cell following a forwarding pointer. At the same time, if the destination MH resides in this virtual cell, the ALS sends to the source BS a VCARP reply which conveys the destination BMAC address so that the subsequent data transfer can directly go to the destination BS.

## 4 Virtual Cell Protocol

### 4.1 Distributed Location Information

There are generally three different ways to distribute location information: centralized, partitioned, and duplicated. Depending on which way is used to treat location information, there is a tradeoff between location registration and paging. In a centralized system, a large volume of location updates at one physical site may degrade the network performance significantly. However, the consistency of location information is obtained and simple paging is achieved. A typical example is the first generation of mobile cellular telephone systems. In a partitioned system, the different partitions of location information are held at different sites. An example is IPIP where every BS maintains its local location information and paging is

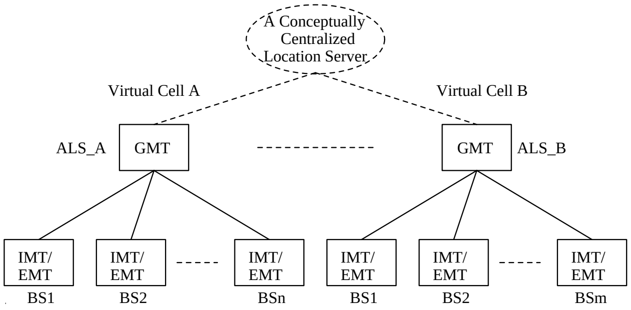


Figure 6: The Distributed Location Information in a Virtual Cell

basically required when two remote physical cells are involved in communication unless a forwarding pointer is found. Thus, this approach can alleviate the problem of location registration in the centralized system, but may need frequent paging. In a duplicated system, on the other hand, the same location information is held at the different sites. VIP is an example where the location information of an MH is held on several intermediate gateways. Although it can give an optimal routing path in the normal case, when a communicating MH migrates, the inconsistent location information may be spread over several intermediate gateways.

In order to support mobility management in a distributed manner, the virtual cell takes advantage of the distributed hierarchical location information which involves a combination of partitioning, duplication, and centralization, as shown in Figure 6. GMT is partitioned between virtual cells. On the other hand, IMT is partitioned with other IMTs and duplicated with GMT in a virtual cell. EMT partially duplicates a part of GMT for MHs in different physical cells in the same virtual cell so that address resolution and location tracking for data transfer are first tried to accomplish among BSs independent of the ALS. Thus, GMT is only referred in case that they cannot be solved among BSs.

It should be noted that there is another conceptual hierarchy above the GMT level. When $MH_1$ migrates from its native virtual cell $A$ to virtual cell $B$, forwarding pointers are maintained from $ALS_A$ to $ALS_B$ to the current BS at the VCP layer. ¿From the prospective of the IP layer, however, $MH_1$ is regarded as if it were in its native virtual cell. Thus, when a remote fixed host or an MH in a third network wants to send a packet to $MH_1$, the existing routing protocols of the IP layer correctly deliver the packet to $ALS_A$. Then $ALS_A$ traces $MH_1$ using the forwarding pointers. Therefore, the IP network address of a migrating MH represents at least the near-location information of the MH, and when coupled with the existing routing protocols of the IP layer it can give the same effect that we maintain a conceptually centralized server for the whole mobile network. If $MH_1$ migrates within a virtual cell, the IP network address of $MH_1$ gives the exact location information.

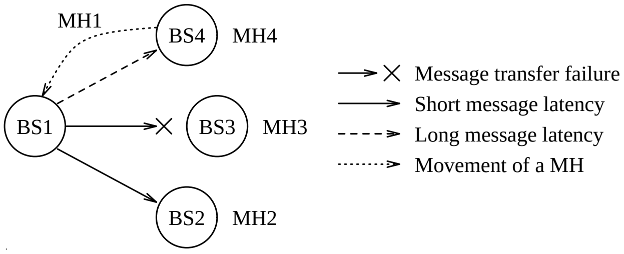


Figure 7: An Example of the Inconsistency Problem

## 4.2 Impact of Mobility

The location registration due to the handover procedure is a main source of inconsistent location information. To achieve the consistency of distributed information is not trivial at all and many researchers have been working on this problem in fixed network environments[15, 16, 17]. In the virtual cell environment, however, host mobility is integrated with fixed networks, which makes the problem even more complex. To solve this problem, the handover procedure should utilize the multicasting ability of the base station network.

Assuming that the distributed location information is initially consistent, consider that $MH_1$ moves from $BS_4$ to $BS_1$ which periodically broadcasts its identity, as shown in the Figure 7. When $MH_1$ receives a beacon packet, it sends a greeting message containing its identity to $BS_1$. $BS_1$ detects the entry of $MH_1$ to its local cell, deletes its corresponding EMT entry if exists, and inserts a new IMT entry. Then, $BS_1$ multicasts a location registration message to every BS so that it maintains the consistent location information for $MH_1$. However, some BSs may fail to receive the message due to communication errors and buffer overflows. The long propagation delay also has the similar effect as the message loss at a point in time due to mobility. To explore a practical solution, we should begin with the assumption that the base station network supports the atomic multicasting by using an existing protocol such as the Trans protocol[17]. The basic idea of the Trans protocol is that acknowledgements for multicast messages are piggybacked on messages that are themselves multicast, using a combination of positive and negative acknowledgement strategies. This piggyback feature is suitable for the handover procedure because in the steady state, if some MHs move out then there will be a high possibility that other MHs move in.

Note that although the atomic multicast is supported, the message loss can still occur if $BS_1$ multicasts the location registration message. Assume that $BS_3$ failed to receive it and $BS_4$ received it with a long message latency. Then, the following four cases can happen when other MHs transmit a message to $MH_1$ during the process of the atomic multicast:

1. If $BS_4$ has the old location information for $MH_1$, the message transmitted from $MH_4$ to $MH_1$ will be lost.
2. If $BS_4$ has the new location information for $MH_1$, the message transmitted from $MH_4$ to $MH_1$ will be received.
3. Because $BS_3$ has the old location information for $MH_1$, the message transmitted from $MH_3$ to $MH_1$ will be lost or received, depending on what value $BS_4$ has. If it has the old value, the message will be lost. Otherwise, the message will be redirected to $BS_1$ by $BS_4$.
4. Because $BS_2$ has the new location information for $MH_1$, the message transmitted from $MH_2$ to $MH_1$ will be received.

¿From the above observation, we can see that at least the previous base station $BS_4$ must have the new location information for $MH_1$ as soon as possible in order to avoid a possible message loss or to mitigate the effect of long message latency.

### 4.3 Handover Procedure

The new BS informs the previous BS of an MH's migration using point-to-point communication immediately after detecting the MH's identity so that all messages received at the previous BS during the handover procedure can be correctly redirected to the new BS. Next, the previous BS is responsible for maintaining the consistency of the MH's location information at all BSs using an atomic multicasting. Note that the ALS is excluded from the multicasting group. If the ALS is involved, every movement of the MH will need an access to the ALS and there is no difference from the centralized system. However, it may cause another inconsistency problem between the ALS and BSs because the ALS may have old location information for some MHs. Thus, the previous BS periodically backups the location information updated during a predefined time interval to the ALS using point-to-point communication. When an MH continues to move to different physical cells, there can be a redirection chain with multiple hops from the first BS to the current BS. However, the chain is eliminated when the newest location information is received through an atomic multicasting.

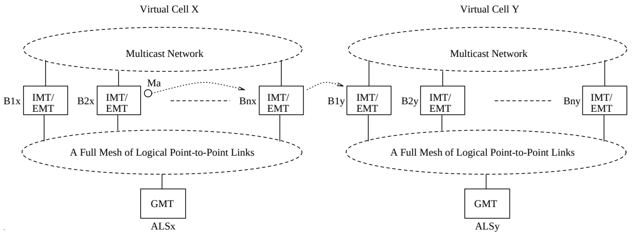


Figure 8: A Communication Model for the Handover Procedure

Figure 8 shows a communication model for the handover procedure. For convenient description, we define the following notations where the VCP header information is omitted:

- $A \Longrightarrow B, \ldots, N : \{\text{frame}\}$ implies that $A$ multicasts $\{\text{frame}\}$ to $B, \ldots, N$ over the base station network, where $\{\text{frame}\}$ consists of $\{$source address, destination address | frame data$\}$.
- $A \longrightarrow B : \{\text{frame}\}$ implies that $A$ sends $\{\text{frame}\}$ to $B$ via a point-to-point link.
- $A \leadsto B : \{\text{frame}\}$ implies that $A$ broadcasts $\{\text{frame}\}$ whose destination is B via radio links.
- $IP(A)$ implies the IP address of an MH $A$.
- $RMAC(A)$ implies the radio MAC address of $A$, where $A$ may be a BS or an MH.
- $BMAC(A)$ implies the address of $A$ in the base station network (BMAC), where $A$ may be a BS or an ALS.

**Within a virtual cell.** Consider that $MH_a$ moves from $BS_{2x}$ to $BS_{nx}$. In addition to its own addresses, $IP(MH_a)$ and $RMAC(MH_a)$, $MH_a$ initially keeps the current base station addresses, $RMAC(BS_{2x})$ and $BMAC(BS_{2x})$.

1. $BS_{nx} \leadsto MH_a$ : $\{RMAC(BS_{nx})$, broadcast address | $BMAC(BS_{nx})\}$

$MH_a$ receives a beacon including $BMAC(BS_{nx})$ from $BS_{nx}$, decides to switch from $BS_{2x}$ to $BS_{nx}$, and updates its base station address from $RMAC(BS_{2x})$ and $BMAC(BS_{2x})$ to $RMAC(BS_{nx})$ and $BMAC(BS_{nx})$, respectively.

2. $MH_a \rightsquigarrow BS_{nx} : \{RMAC(MH_a), RMAC(BS_{nx}) \mid IP(MH_a), BMAC(BS_{2x})\}$
   $BS_{nx}$ receiving a greeting message from $MH_a$ processes its IMT and EMT using the source address $RMAC(MH_a)$ as a key. If the EMT entry for $MH_a$ is found, the entry is deleted and then the IMT entry for $MH_a$ is added. $IP(MH_a)$ in the EMT entry is used for fast address resolution, as described in the VCARP section later.
3. $BS_{nx} \longrightarrow BS_{2x} : \{BMAC(BS_{nx}), BMAC(BS_{2x}) \mid IP(MH_a), RMAC(MH_a), BMAC(BS_{nx})\}$
   $BS_{2x}$ receiving the notification of $MH_a$'s migration sends an acknowledgement to $BS_{nx}$ and then processes IMT and EMT. The IMT entry for $MH_a$ is deleted and the new EMT entry for $MH_a$ is added.
4. $BS_{2x} \Longrightarrow BS_{1x}, \ldots, BS_{nx} : \{BMAC(BS_{2x})$, multicast address $\mid IP(MH_a), RMAC(MH_a), BMAC(BS_{nx})\}$
   Every BS receiving the notification of $MH_a$'s migration except $BS_{nx}$ processes its EMT. If the EMT entry for $MH_a$ already exist, the entry is updated.

**Between virtual cells.** Consider that $MH_a$ moves from $BS_{nx}$ to $BS_{1y}$. In addition to its own addresses, $IP(MH_a)$ and $RMAC(MH_a)$, $MH_a$ keeps the current base station addresses, $RMAC(BS_{nx})$ and $BMAC(BS_{nx})$.

1. $BS_{1y} \rightsquigarrow MH_a : \{RMAC(BS_{1y})$, broadcast address $\mid BMAC(BS_{1y})\}$
2. $MH_a \rightsquigarrow BS_{1y} : \{RMAC(MH_a), RMAC(BS_{1y}) \mid IP(MH_a), BMAC(BS_{nx})\}$
   $BS_{1y}$ receiving a greeting message from $MH_a$ detects from the previous base station address $BMAC(BS_{nx})$ that $MH_a$ was in another virtual cell. $BS_{1y}$ creates the IMT entry for $MH_a$.
3. $BS_{1y} \longrightarrow ALS_Y : \{BMAC(BS_{1y}), BMAC(ALS_Y) \mid IP(MH_a), RMAC(MH_a), BMAC(BS_{1y}), BMAC(BS_{nx})\}$
   $ALS_Y$ receiving a location registration message creates the GMT entry for $MH_a$ using $IP(MH_a)$, $RMAC(MH_a)$, and $BMAC(BS_{1y})$. Then, $ALS_Y$ knows that $BS_{nx}$ is a base station in virtual cell $X$, using $BMAC(BS_{nx})$.
4. $ALS_Y \longrightarrow ALS_X : \{BMAC(ALS_Y), BMAC(ALS_X) \mid IP(MH_a), RMAC(MH_a), BMAC(ALS_Y), B_MAC(BS_{nx})\}$
   $ALS_X$ receiving a location registration message updates the GMT entry for $MH_a$ using $IP(MH_a)$, $RMAC(MH_a)$, and $BMAC(ALS_Y)$. This entry serves as a forwarding pointer at the VCP layer if $ALS_X$ is the ALS of $MH_a$'s native virtual cell. If it is not, $ALS_X$ informs the ALS of $MH_a$'s native virtual cell of $MH_a$'s migration to virtual cell $Y$ so that a forwarding pointer is maintained from $MH_a$'s native virtual cell to virtual cell $Y$.
5. $ALS_X \longrightarrow BS_{nx} : \{BMAC(ALS_X), BMAC(BS_{nx}) \mid IP(MH_a), RMAC(MH_a), BMAC(ALS_X)\}$
   $BS_{nx}$ receiving the notification of $MH_a$'s migration sends an acknowledgement to $BS_{1y}$ in the reverse direction and then processes IMT and EMT. The existing IMT entry is deleted and an EMT entry using $IP(MH_a)$, $RMAC(MH_a)$, and $BMAC(ALS_X)$ is added.
6. $BS_{nx} \Longrightarrow BS_{1x}, BS_{2x}, \ldots, BS_{(n-1)x} : \{BMAC(BS_{nx})$, multicast address $\mid IP(MH_a), RMAC(MH_a), BMAC(ALS_X)\}$
   Every BS receiving the notification of $MH_a$'s migration except $BS_{nx}$ manipulates its EMT. If the EMT entry exists, it is updated.

## 4.4 Address Resolution

VCARP is designed to get the physical address of the MH in a remote physical cell of the virtual cell and should satisfy at least the following three requirements. First, in order to achieve the compatibility with ARP, VCARP must be shielded from MHs as if they were in a physical cell. Second, because the VCARP packet latency can directly affect the scalability of a virtual cell, a distributed address resolution mechanism should be considered rather than broadcasting or a centralized solution. Third, when an MH moves to an adjacent physical cell immediately after sending an ARP request and further moves to another physical cell, the ARP reply may be lost and then the ARP request may be repeatedly generated. Hence, VCARP must also support host mobility.

Figure 9 shows the ARP/VCARP packet format. The ARP packet follows exactly the same format as the existing ARP standard and has longer fields *SENDER HA* and *TARGET HA* in order to make it possible to accommodate the use of hierarchical 64-bit E.164 network addresses in radio networks. The VCARP packet format has an additional 64-bit field *BASE HA* which is used to convey the location information which is expected to be used for data transfer in the very near future. The source MH performs ARPing only when it resides in its native virtual cell and the destination MH belongs to the same native virtual cell. Otherwise, it directly transmits IP packets without ARPing.

**Within a physical cell.** The procedure is the same as in the normal ARP. The same ARP request/reply packet formats are used where an 8-bit hardware address length field allows to accommodate arbitrary radio network addresses. The IP protocol of the source MH checks the destination IP address with its address resolution table. If the address binding is successful, the source MH uses the RMAC address to transfers datagrams. Otherwise, the source MH broadcasts an ARP request. As soon as the destination MH receives the ARP request, it responds with an ARP reply that includes the requested physical address of the destination. At the same time, the BS that received the ARP

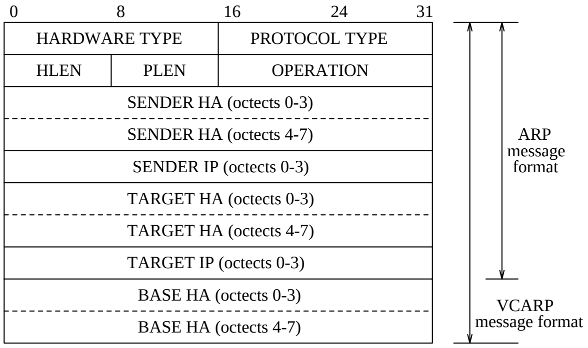
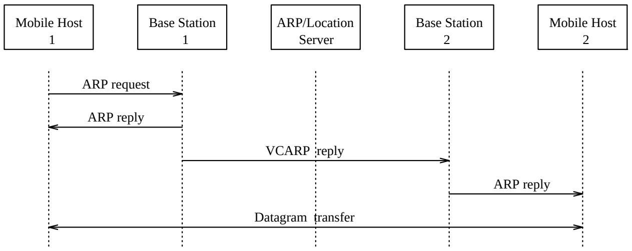
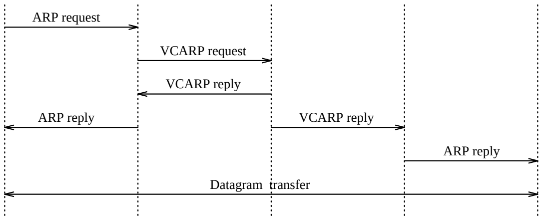


Figure 9: The ARP/VCARP Packet Format



(a) The case that base station 1 resolves the address binding



(b) The case that the ALS resolves the address binding

Figure 10: The VCARP Flow Diagram

request searches IMT and knows that the destination MH is in the local physical cell so it discards the ARP request.

**Within a virtual cell.** Figure 10(a) shows the flow of messages when $MH_1$ broadcasts an ARP request to get the physical address of $MH_2$ and it is resolved at $BS_1$. $BS_1$ that received the ARP request searches its EMT. The EMT entry for $MH_2$ already contains its physical address and $BS_1$ directly sends an ARP reply within the physical cell. At the same time, $BS_1$ sends a VCARP reply to $BS_2$. Then, $BS_2$ transforms the format of the VCARP reply to that of the ARP reply by stripping out the field BASE HA and broadcasts the transformed ARP reply in its local cell. This VCARP reply also conveys the location information of $MH_1$ to $BS_2$ through the field BASE HA which contains the network address of $BS_1$. Then, $BS_2$ creates an EMT entry so that $MH_2$ can directly transmit data to $MH_1$ without any additional address resolution in the near future communication. It is based on the observation that computer communication is usually bidirectional; if $MH_1$ has some reason to talk to $MH_2$, then $MH_2$ will probably have some reason to talk to $MH_1$.

Figure 10(b) shows the flow of messages when the address resolution is accomplished at the ALS. $BS_1$ sends a VCARP request to the ALS because it has no entry for $MH_2$. Then, the ALS sends a VCARP reply to $BS_1$ after resolving $MH_2$'s address binding using GMT. At the same time, the ALS also sends $BS_2$ a VCARP reply with the same reason as in the above case. The location information of $MH_1$ and $MH_2$ is also conveyed via both VCARP replys to establish a logical bidirectional link between them. Now we describe the support of host mobility in VCARP. Consider the case that $MH_1$ requested an address resolution to send datagrams to $MH_2$ and then moved to a third $BS_n$ before the VCARP reply is arrived. By the handover procedure, $BS_1$ knows that $MH_1$ currently belongs to $BS_n$. As soon as $BS_1$ receives the VCARP reply, it redirects the packet to $BS_n$.

## 4.5 Data Transfer

For the transmission of packets between virtual cells and fixed networks, we have already described in the previous sections. This section deals with the packet transmission within a physical cell and within a virtual cell. Let $BS_{ij}$ denote the $i$-th BS of virtual cell $j$ in Figure 6, where $i = 1, \ldots, n$ when $j = A$ and $i = 1, \ldots, m$ when $j = B$. We assume that $MH_{ij}$ belongs to $BS_{ij}$ and $M_x$ initially belongs to $BS_{1A}$. We also assume that the address binding is achieved by VCARP.

**Within a physical cell.** Consider the case that $MH_{1A}$ wants to send a packet to $MH_x$. By the address resolution protocol, $MH_{1A}$ knows the RMAC address of $MH_x$ and directly broadcasts the packet to it. $BS_{1A}$ knows that $MH_x$ resides in its physical cell by checking IMT using the RMAC address of $MH_x$ as a key, and discards the received packet.

**Within a virtual cell.** Consider the case that $MH_x$ moves from $BS_{1A}$ to $BS_{2A}$ and each of $MH_{1A}$, $MH_{2A}$, and $MH_{nA}$ wants to send a packet to $MH_x$. By the handover procedure, a forwarding pointer is set from $BS_{1A}$ to $BS_{2A}$ at the EMT entry of $BS_{1A}$ and the packet from $MH_{1A}$ is correctly delivered to $MH_x$ by the forwarding pointer. The packet from $MH_{2A}$ is directly delivered to $MH_x$ and $BS_{2A}$ discards it because the IMT entry for $MH_x$ is found. The delivery of the packet from $MH_{nA}$ has three cases. If the EMT entry for $MH_x$ is found before the completion of $BS_{1A}$'s multicasting operation, the packet will be delivered to $BS_{2A}$ via $BS_{1A}$. If the EMT entry is found after the completion of its multicasting operation, the packet will be directly delivered to $BS_{2A}$. If the EMT entry is not found, the first packet from $MH_{nA}$ will be delivered to $BS_{2A}$ through the ALS. At the same time, the ALS generates a VCARP reply in order to have $BS_{nA}$ learn that $MH_x$ is currently belongs to $BS_{2A}$. It makes it possible to directly transfer the subsequent packets following the first one to $BS_{2A}$ without going through the ALS repeatedly.

## 4.6 VCP Implementation

The virtual cell is implemented as a distributed application that uses TCP/IP as a transport mechanism. There are five types of distributed network component processes, each of which corresponds to a network entity in the virtual cell: mobile host process, physical cell process, base station process, base station network process, and ARP/Location server process. The physical cell process simulates the broadcast ability of a physical cell while the base station network process simulates a full mesh of logical point-to-point links with the multicast ability. Both are implemented as servers and the other processes are implemented as clients. The mobility and data traffic are generated at the mobile host process. When a client process is initiated, the user defines a simulated IP address, the destination machine and protocol port number of the corresponding server process(es). Because the socket address structure contains an internet address and a protocol port number to identify a communication endpoint between a TCP connection, the <IP address, port number> pair is used as an RMAC or BMAC address of a client process. The purpose of this implementation is to demonstrate the correct operation of VCP and measure data including the protocol processing time of the VCP layer at a BS and an ALS, which will be used as parameters to a performance model for the virtual cell in the future research.

# 5 Comparison and Discussion

Figure 11 summarizes the characteristics of the virtual cell and compares with IPIP and VIP. The figure also includes Cellular Packet Switch (CPS)[7, 8] which is designed for digitized voice and data services in personal communications. The basic idea of CPS is to interconnect BSs with high-bandwidth Metropolitan Area Networks based on the IEEE 802.6 standard and support mobility using the virtual circuit packet-switching technique.

| Networks / items | IPIP | VIP | CPS | VCP |
|---|---|---|---|---|
| Target Applications | TCP/IP applications | TCP/IP applications | packetized voice and data services | TCP/IP applications |
| Target environment | campus area | wide area | metropolitan area | wide area |
| Base station networks | point-to-point physical links (Internet) | point-to-point physical links (Internet) | DQDB MAN | packet-switched connectionless net with multicast ability (SMDS) |
| Mobility support layer | network layer | network layer | data link layer | data link layer |
| Type of movement | mobility | portablility/ mobility | mobility | mobility |
| Type of packet switching | datagram | datagram | virtual circuit | datagram |
| Address resolution | Proxy ARP | address mapping | no | VCARP |
| Scalability | low | high | high | high |
| Cell coverage area | physically bounded | physically bounded | physically bounded | logically flexible |

Figure 11: Comparisons of the Virtual Cell With Other Networks

When considering the ubiquitous coverage of TCP/IP networks and applications, the virtual cell has several advantages against IP-level mobile host protocols. Because the virtual cell uses a underlying datagram network to interconnect BSs and to form flexible coverage, the difficulties arising from diverse underlying networks can be avoided in terms of the performance of mobility management function and the complexity of network management function. Moreover, because the virtual cell is a logical cell and the bridge function is usually much faster than the gateway function, it is possible to increase performance significantly if we can properly engineer the deployment of virtual cells so that the intra virtual cell traffic is as much larger than the inter virtual cell traffic as possible. Preserving the interconnection level of BSs at the data link layer also makes it possible to shield host mobility from the IP layer, and using the base station network address as a contact point of an MH makes it possible to avoid acquiring a temporary address by some kind of local utilities whenever the MH migrates. Thus, the problems of interfering with the conventional IP layer and of reserving a large amount of IP address space for location information can be solved.

There are two major issues related to the performance of VCP. Within a virtual cell, a BS first tries to perform address resolution and data transfer locally without the intervention of the ALS, based on its two membership tables, IMT and EMT. Thus, the hit ratio of the EMT search directly affect the degree of referring to the ALS when remote MHs in different physical cells are involved in communication. This issue directly addresses the size and update scheme of EMT and also closely related to the IP traffic and host mobility pattern. A number of research results on the internet protocol traffic analysis in fixed network environment reveals that certain hosts communicate more with one another than with other hosts[19, 20, 21]. Even in the virtual cell environment, the locality of destination hosts can give an important implication for each of the address resolution and data transfer functions. A small ARP cache at each MH can greatly reduce VCARP traffic within a virtual cell and then address resolution should not severely burden the size of EMT. If we do not consider host mobility, the locality property also means that data transfer does not severely burden the size of EMT. When considering host mobility, the update scheme of EMT could be rather important if we can properly deploy virtual cells based on mobility and traffic patterns among physical cells.

The other issue is the amount of multicast traffic among BSs, which is generated only by the handover function of the virtual cell. Because the multicasting ability is supported by switches in the base station network, each BS is not related to the load of generating the multicast traffic, and the bandwidth usage of the base station network is affected by it. The relatively short length of the handover message should not severely affect the high-bandwidth base station network. This should be addressed quantitatively based on mobility models.

## 6 Conclusion and Future Works

We have presented a new network infrastructure for the transmission of IP datagrams in mobile computer communications. With the virtual cell concept, physical cells are grouped into larger logical cells where host mobility is shielded from the IP layer. It eliminates the need of IP-level protocols for host mobility within a virtual cell and provides the flexible coverage area that can be properly engineered according to host mobility and data traffic patterns among physical cells. In order to achieve this concept, we have described the virtual cell protocol which consists of handover, address resolution, and data transfer modules, based on the distributed hierarchical location information. We are developing a performance model of the virtual cell to capture protocol processing loads at the base station and ARP/Location server and scalability in terms of the number of physical cells. Based on the analysis, we will deal with how to deploy virtual cells to minimize the total cost of tracking mobile users in a given environment.

## References

[1] John Ioannidis,Dan Duchamp, and Gerald Q. Maguire Jr., "IP-based Protocols for Mobile Internetworking," *ACM SIGCOMM'91*, pp. 235-245, 1991.

[2] Fumio Teraoka, Yasuhiko Yokote, and Mario Tokoro, "A Network Architecture Providing Host Migration Transparency," *ACM SIGCOMM'91*, pp. 209-220, 1991.

[3] Fumio Teraoka, Kim Claffy, and Mario Tokoro, "Design, Implementation, and Evaluation of Virtual Internet Protocol," *IEEE DCS'92*, pp. 170-177, 1992.

[4] Charles Perkins and Yakov Rekhter, T. J. Watson Research center, IBM Corp., "Short-cut Routing for Mobile Hosts," draft RFC, July 1992.

[5] Charles Perkins and Yakov Rekhter, T. J. Watson Research center, IBM Corp., "Support for Mobility with Connectionless Network Layer Protocols," draft RFC, Jan. 1993.

[6] Hiromi Wada, Tatsuya Ohnishi, and Brian Marsh, "Packet Forwarding for Mobile Hosts," Matsushita Corp., draft RFC, Nov. 1992.

[7] David J. Goodman, Gregory P. Pollini, and Kathleen S. Meier-Hellstern, "Network Control for Wireless Communications," *IEEE Commun. Mag.*, pp. 116-124, 1992.

[8] Kathleen S. Meier-Hellstern, Gregory P. Pollini, and David J. Goodman, "A Wireless Service for the IEEE 802.6 Metropolitan Area Network," *IEEE GLOBECOM'91*, pp. 1964-1968, 1991.

[9] David M. Piscitello and Michael Kramer, "Internetworking Using Switched Multi-megabit Data Service in TCP/IP Environments," *Computer Networks and ISDN Systems*, pp. 62-71, 1991.

[10] F. R. Dix, M. Kelly, and R. W. Klessig, "Access to a Public Switched Multi-megabit Data Service Offering," *Computer Networks and ISDN Systems*, pp. 46-61, 1991.

[11] D. Piscitello and J. Lawrence, "The Transmission of IP Datagrams over the SMDS Service," *RFC 1209*, 1991.

[12] George H. Clapp, "LAN Interconnection Across SMDS," *IEEE Network Magazine*, pp. 25-32, September 1991.

[13] Masatoshi Kawarasaki and Bijan Jabbari, "B-ISDN Architecture and Protocol," *IEEE J. Select. Areas Commun.*, Vol. 9, No. 9, pp. 1405-1415, December 1991.

[14] Jean-Yves Le Boudec, "The Asynchronous Transfer Mode: a tutorial," *Computer Networks and ISDN Systems*, Vol. 24, pp. 279-309, 1992.

[15] Susan B. Davidson, "Consistency in Partitioned Networks," *ACM Computing Surveys*, Vol. 17, No. 3, pp. 341-370, September 1985.

[16] Jo-Mei Chang and N. F. Maxemchuk, "Reliable Broadcast Protocol," *ACM Trans. on Computer Systems*, Vol. 2, No. 3, pp. 251-273, August 1984.

[17] P. M. Melliar-Smith, Louise E. Moser, and Vivek Agrawala, "Broadcast Protocols for Distributed Systems", Vol. 1, No. 1, pp. 17-25, 1990.

[18] Samuel Sheng, Anantha Chandrakasan, and R. W. Brodersen, "A Portable Multimedia Terminal," *IEEE Communications Magazine*, pp. 64-75, December 1992.

[19] Andrew Schmidt and Roy Campbell, "Internet Protocol Traffic Analysis with Applications for ATM Switch Design," *ACM SIGCOMM'92*, pp. 39-52, 1992.

[20] Ramon Caceres, Peter B. Danzig, Sugih Jamin, and Danny J. Mitzel, "Characteristics of Wide-Area TCP/IP Conversations," *ACM SIGCOMM'91*, pp. 101-112, 1991.

[21] K. M. Khalil, J. C. Hand, and M. Mariswamy, "Analysis and Traffic Characterization of A Wide Area Network," *IEEE ICC'93*, pp. 1829-1835, 1993.