Cellular Universal IP: A Low Delay Mobility Scheme based on Universal IP Addressing

Patrick P. Lam
Chinese University of Hong Kong
lampk3@ie.cuhk.edu.hk

Soung C. Liew
Senior Member, IEEE
Chinese University of Hong Kong
soung@ie.cuhk.edu.hk

Jack Y. B. Lee
Senior Member, IEEE
Chinese University of Hong Kong
yblee@ie.cuhk.edu.hk

ABSTRACT
The concept of care-of-address (CoA) is a major cause of excessive handoff delay in Mobile IPv6 for real time multimedia traffic. Many schemes eliminate the use of CoA at the micro-mobility scale, but leave the macro-mobility unsolved. This paper proposes a novel alternative IPv6 mobility scheme based on *universal addressing* – Cellular Universal IP (CUIP) – for real-time traffic in wireless access networks. In CUIP, a mobile node is addressed with a universal IP address regardless of its location, making CoA and tunneling unnecessary in micromobility and even macromobility handoffs. CUIP manages roaming and handoff differently – whereas explicit signaling is used for roaming, a *handoff-on-the-fly* route-update scheme is used during handoff to embed signaling information into the outgoing data packets to minimize handoff delay. We prove analytically that, on average, fewer than three routers need to be updated per handoff. As a result, CUIP incurs an expected network layer handoff delay on the order of milliseconds only. In addition, the support of QoS is possible. A simple security scheme is also proposed to enable mutual authentication at the network layer.

Categories and Subject Descriptors
C.2.1 [**Computer Communication Networks**]: Network Architecture and Design – *wireless communication*.

General Terms
Algorithms, Performance, Design and Security.

Keywords
IP mobility, handoff, macro-mobility and micro-mobility.

1. INTRODUCTION

1.1 Fundamental Problem of MIPv6
Mobile IPv6 [4] (MIPv6) is the *de facto* standard for IP mobility in IPv6 networks. In MIPv6, whenever a mobile node (MN) moves to a foreign network, it must acquire a globally unique CoA. According to [16] and [17], the CoA acquisition alone already contributes about two seconds to the handoff delay in MIPv6. When the extra delay for home agent (HA) binding is included, the blackout period incurred by a handoff will likely be considerably longer than the tolerable limit of most real-time applications (i.e., 150 ~ 400ms [7]). The concept of CoA also causes the well-known triangular routing problem [3]. Although Route Optimization (RO) [3] has been standardized to eliminate this shortcoming, RO itself requires at least 1.5 roundtrips [4] of message exchanges between the MN and correspondent node (CN) to complete. Furthermore, RO requires specific support from the CN, which may not be readily available in many cases.

Therefore, the requirement of acquiring a new CoA after handoff is the fundamental cause of the suboptimal performance of MIPv6 in many aspects. In this paper we propose a novel approach that eliminates the need of CoA by allowing a MN to be addressed with a universal IPv6 address regardless of its location. We also prove analytically that this scheme gives a tight upper bound on the expected handoff delay.

1.2 Roaming vs. Handoff
Although MIPv6 enables user mobility across different networks, it is only good for "roaming", rather than for "handoff", especially for real-time applications. Roaming and handoff are differentiated by the channel activity when a user moves from one network to another. If the user is not involved in an active communication session during the movement, the user is said to be "roaming" to the new network. If the user is in an active session and the continuity of the session must be maintained, then the user is said to be "handing off" to the new network. The main requirement for roaming is to allow newly initiated call sessions to reach the user at the new location. Roaming is relatively less delay sensitive.

The delay requirement for handoff, however, is much more stringent than roaming. Consider a user in an active voice session with continuous data streams in both directions. The handoff signaling delay must be kept to a minimum so that the user will not notice the blackout period for the ongoing session. That is, the network must continuously and accurately route packets to the user's new location "almost immediately" after the handoff, preferably within 150ms. In terms of "handoff", MIPv6 is ineffective because of the excessive handoff delay.

This paper primarily focuses on improving the handoff performance for IP mobility. Unless specifically noted otherwise, handoff scenarios will be assumed in our discussions.

1.3 Major contributions of this paper
In summary, two key contributions of this paper are as follows:

1. To overcome the fundamental deficiencies of MIPv6 in handoff, we propose:
   a. An IP mobility scheme, Cellular Universal IP (CUIP), using universal IPv6 addressing that

This work is sponsored by the Areas of Excellence scheme established under the University Grant Committee of the Hong Kong Special Administrative Region, China (Project Number AoE/E-01/99).

MSWiM'05, October 10–13, 2005, Montreal, Quebec, Canada.

eliminates the need of CoA and tunneling at both micro- and macro-mobility scale handoffs.

b. A handoff-on-the-fly signaling scheme that eliminates the need for explicit handoff signaling to achieve minimal handoff delay.

2. We prove analytically that the expected number of routers being updated per handoff in CUIP is upper bounded by three. Consequently:

   a. The impact of IP mobility on quality-of-service overhead is reduced.

   b. The "blackout" time of real-time connections due to handoff is minimal.

1.4 Organization of the Paper

This paper introduces a universal IPv6 addressing scheme for IP mobility in the wireless access networks. The remainder of this paper is organized as follows. Section 2 discusses the previous work of IP mobility. Section 3 introduces the architecture of CUIP and the major supporting algorithms. Section 4 analyses the performance of CUIP and section 5 concludes the paper.

2. Related Work

Among the IP mobility proposals, the most well-known ones include Fast-Handovers for Mobile IPv6 (FH) [1], Hierarchical Mobile IPv6 (HMIPv6) [8], Handoff-Aware Wireless Access Internet Infrastructure (HAWAII) [10] and Cellular IP (CIP) [9]. Virtually all of these proposals are along one or both of the following directions.

2.1 Micromobility Oriented Architecture

This approach reduces the handoff delay at the micro-mobility scale, and leaves the macro-mobility for the traditional MIPv6 to handle. Examples of this approach include HMIPv6, HAWAII and CIP. The major problem of this approach is that it still requires the acquisition of CoA in macro-mobility or even micro-mobility handoffs (e.g., HMIPv6). Hence, the overall delay caused by CoA acquisition is only reduced, rather than eliminated.

CUIP, on the other hand, does not require a CoA after roaming or handoff in either micro- or macro-mobility. In addition, macro-mobility across CUIP enabled networks is also handled through CUIP itself, without relying on MIPv6.

2.2 Explicit Handoff Signaling

Another common approach is to use explicit signaling protocols to set up the handoff process and/or the packet forwarding tunnel between the home and the visiting networks. Examples of this approach include FH, HMIPv6, and HAWAII. The major drawback is that the handoff signaling protocols are most likely invoked along the cell boundaries when the MN needs to be handed off. Since the signal strength is usually weak along the cell boundaries, the signaling operation may face a considerable failure rate. Waiting for time-out and retransmission of signaling messages only add more latency to the handoff process. After all, the explicit signaling protocols themselves will also inevitably introduce signaling delay to the process.

CUIP is a pure Layer-3 scheme that embeds the handoff signaling information into the IPv6 hop-by-hop option header of the outgoing data. Although data packets may still face high loss rate at the cell boundaries, waiting for signaling time-out and retransmission are not necessary because additional signaling information will reach the network with subsequent data packets. This is especially effective for real-time applications characterized with continuous stream of data packets, as small packet loss is usually favored over excessive handoff delay [11].

3. Cellular Universal IP (CUIP)

3.1 Overview of CUIP

The concept of universal addressing is borrowed from the traditional mobile phone communication which allows a MN to be addressed anywhere with a single phone number. We believe that a similar scheme should be applicable to the IP network as well if enough addressing space is available. IPv6, with its huge addressing space, is therefore a good enabler of such a scheme.

Thus, CUIP allows an MN to be addressed and located universally by a single IPv6 address. An MN does not need to acquire a CoA when traveling to a foreign network and therefore does not need to register the CoA with the home network for its movements. Consequently, no tunneling of packets is involved.

CUIP relies on Layer-2 signaling to trigger Layer-3 handoff. After that, it is a pure Layer-3 scheme that does not depend on explicit signaling mechanism. The underlying mechanism of CUIP is based on the following observations.

1. In a hierarchical network structure, all handoff scenarios must consist of exactly one cross-over router (COR) between the previous and the new routes, where the COR is defined as the router at the closest forking point of the two routes with respect to the MN. After handoff, only the routers on the new route and the previous route, up to the COR, need to be updated. The handoff is therefore transparent to the rest of the network, including the CN (see Figure 1). In addition, the handoff is completed as soon as the new route is updated.

2. The routers to be updated for handoff are along the data path. Therefore, signaling can be piggybacked on outgoing data packets for more efficient handoff, particularly for real-time applications with continuous stream of data packets. The signaling delay is therefore proportional to the time interval between two consecutive data packets. That is, the signaling delay scales naturally with the blackout delay requirements of the data stream.

For roaming, CUIP relies on a simple network layer signaling scheme to update the new and previous routes. For handoff, which is considered to be more time critical, the routing update of CUIP is enabled by a *handoff-on-the-fly* route-update scheme. Route-update information is embedded in the IPv6 hop-by-hop option header of the outgoing data packets (e.g., voice and video packets) during handoff, so that the handoff can be processed as soon as the outgoing data packets travel along the routers on the new route, and completed as soon as the first data packet reaches the COR. Such a route-update mechanism eliminates the delay encountered by explicit signaling. At the COR, route-update information will be removed from the IPv6 header of the data packets as they continue to travel to the CN, so that backward compatibility with external networks is ensured. In section 4, we will prove that on average less than three routers need to be updated per handoff regardless of the network layout. Consequently, the Layer-3 handoff process will have an expected latency that is on the order of milliseconds. Furthermore, the work load of the handoff operation is distributed throughout the entire network, and thus there will be no single-point-of-failure.

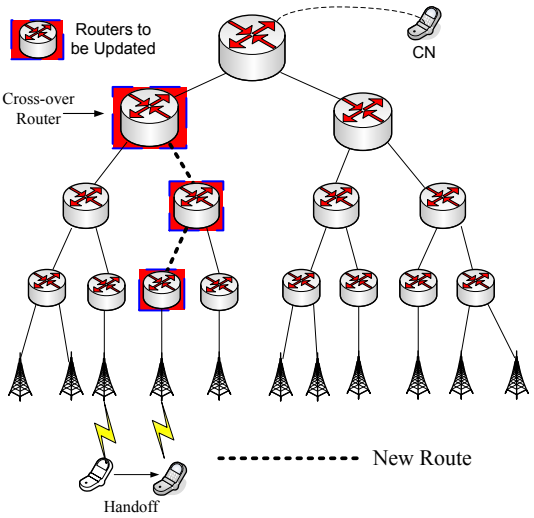

Figure 1. Only a few routers are required to be updated during handoff.

3.2 Terminology

In the following section we define a few terms used in this paper (please refer to Figure 2).

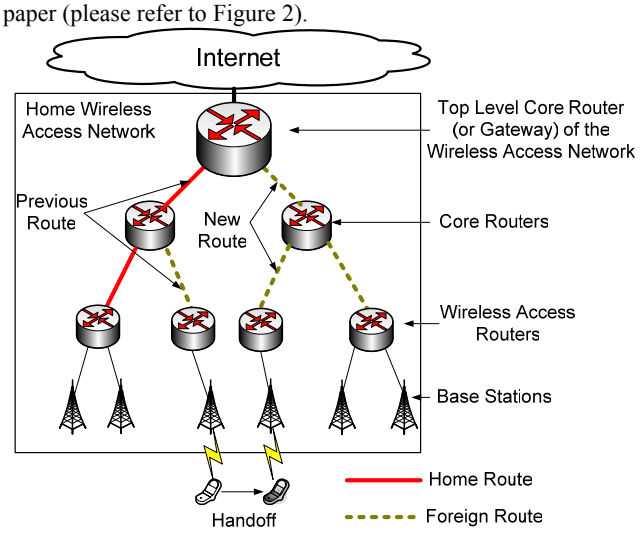

Figure 2. A sample wireless access network

A *wireless access network* is the network owned by the cellular service provider. A *universal address* is a globally unique IPv6 address assigned to an MN when a user subscribes to the IP service from a provider. This universal address is invariant under mobility and is used to address the MN regardless of its location.

A router in the wireless access network is defined as an IPv6 router with CUIP support. A router can further be differentiated as a *Wireless Access Router (WAR)* and a *core router*. A WAR is an access router that connects the base stations to the core routers of the wireless access network, and provides IP connectivity to the MNs. The core routers connect the WARs to the top level core router, or the gateway, of the wireless access network.

A *home network* of a MN refers to the wireless access network that an MN subscribes service from, whereas a *foreign network* is any wireless access network other than the home network.

A *home route* of an MN is the entire route, from the top level core router, or the gateway, all the way to the WAR, to which the MN is assigned to during subscription. A *foreign route* is a route, or the portion of a route, that deviates from the home route. All packets for a particular MN are always forwarded to or along its corresponding home route. If the MN is away from the home route, the packets will hit a COR along the way to the MN and from there they will be directed to the appropriate foreign route where the MN is currently located.

A *previous route* is the route an MN was attached to before handoff, and a *new route* is the route an MN will be attached to after handoff. The home route and foreign route must not be confused with the previous route and new route. In CUIP, a previous route may or may not be a home route, whereas a new route may or may not be a foreign route. Note also that the base station is excluded from the definition of a route.

A *handoff scenario* accounts for the network layer movement across two wireless access routers. Note that handoffs between base stations that do not invoke network layer handoffs are not within the scope of this paper.

3.3 The Logical Components

3.3.1 Cross-over Router (COR)

The COR is a core router in the wireless access network that happens to be the cross-over point for a particular handoff scenario. Its main responsibility is to correctly forward the future incoming packets to the new route of the corresponding MN after handoff, and to notify the routers on the previous route about the MN's handoff. Figure 3 illustrates the concept of COR. Note that different handoff scenarios may be associated with different CORs, and a COR may or may note be along the home route of the corresponding MN.

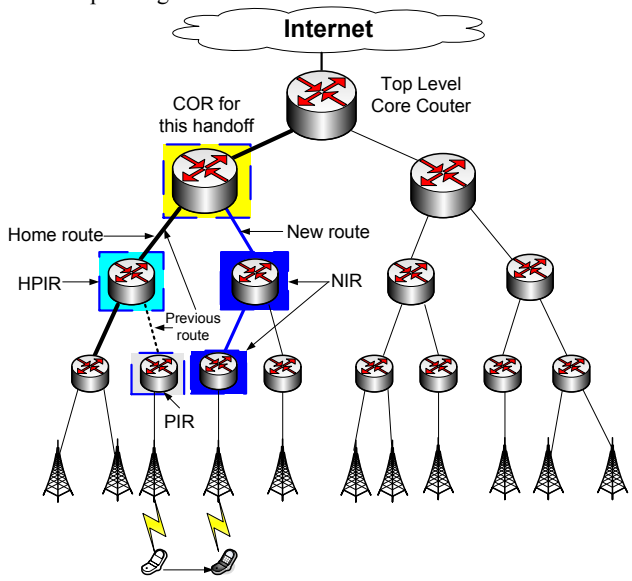

Figure 3. Illustration of the concept of COR

3.3.2 Intermediate Router (IR)

An IR is a router underneath the COR. A new IR (NIR) lies on the new route of a MN after handoff, and is responsible for directing packets addressed to the visiting MNs appropriately. A previous IR (PIR) lies on the previous route of a MN. Unless the PIR also lies on the home route of the MN, the host entry for the MN will be removed from the PIRs once the MN is handed off to

a new route. A home PIR (HPIR) is a PIR on the home route of the MN. Note that the entry of the MN will not be removed from an HPIR. Instead, HPIR will forward the packets addressed to the MN initiated from a WAR on the home route (i.e., the CN is on the home route of the MN) to the default route until it reaches the COR (recall that a default route always forwards packets to the upper level core router), so that the COR can relay the packets to the new route of the MN. A router can simultaneously be a COR, a NIR, PIR and HPIR for different MNs.

3.3.3 Top Level Core Router (TLCR)

A TLCR refers to the core router at the top of a wireless access network hierarchy. It can also be considered as the gateway to the Internet as described in [14]. According to [15], the TLCRs of fixed-line ISPs are directly connected in a fully meshed topology. For wireless ISPs that support MN mobility they will likely have set up appropriate service agreements among themselves. Therefore we can safely assume that their wireless access networks will also be inter-connected directly through their TLCRs to facilitate macromobility of the MNs.

3.4 The Mobility Routing Table (MRT)

It is worth noting that the concept of COR requires the routers to lookup 128-bit long IPv6 host addresses of the MNs frequently, because CORs need to correctly route packets to the MNs according to their host addresses. If not done right, this could impact the efficiency of the routing tables negatively. With this in mind, we have designed a new routing table structure, called mobility routing table (MRT), to handle the IP route lookup with enhanced scalability and efficiency even in the wireless access networks characterized by high user mobility.

	Fixed Routing Part				
	Destination	**Gateway**	**Flags**	**Interface**	**Away**
1	aabb::fe11:0/112	aabb::fe11:0001	X	eth1	
2			C		aabb::fe11:25bd/128
4			F		aabb::fe11:aef3/128
5	aabb::fe12:0/112	aabb::fe12:0001		eth2	
6	default	aabb::fe00:0001		eth0	
7					
8		**Mobility Routing Part**			
9	**Destination**	**Gateway**	**Flags**	**Interface**	
10	aabb::fe11:25bd/128	aabb::fe12:0001		eth2	
12					
13		**Visitor Routing Part**			
14	**Destination**	**Gateway**	**Flags**	**Interface**	
15	aabb::23da/128	aabb::fe11:0001		eth1	
16	aabb:1111::3df2:13ff/128	aabb::fe11:0001		eth1	

Fixed Routing Part
-- Containing the prefix entries of the home routes for the MNs.
--The "Away" field contains host entries for the MNs currently away from home.
-- "X" flag indicates that some MNs with this prefix is currently away from home.
-- "C" flag indicates that this router is a COR for the MN in the "Away" field.
-- "F" flag indicates that this router is a HPIR for the MN in the "Away" field.
Mobility Routing Part
-- Containing the host entries of the MNs for which this router serves as the COR.
Visitor Routing Part
-- Containing the host entries for the MNs for which this router serves as a NIR.
-- Only searched when the Fixed Routing Part returns no result for the query.
-- If this part also finds no result for the query, the Default route will be chosen.

Figure 4. A sample MRT

Instead of using the non-scalable flat routing table structure, MRT preserves the prefix routing efficiency to a certain degree so that the impact of host address lookup on the overall routing performance can be reduced. It can be shown that, with MRT, the routing table lookup performance largely depends on the ratio between the number of MNs staying on their home routes and the number of MNs away from their home routes. Due to space limit, we only provide a sample MRT structure in Figure 4 without going into the details. The only thing we need to know about the MRT at this point is that, each router must contain exactly one default route, and this default route must point to the upper level core router. That is, all the packets of failed queries must be forwarded to the corresponding upper-level core router, because they must be forwarded to the default route.

3.5 Self-Identification of COR

A core router must be able to identify itself to be the COR for the particular handoff scenario. This self-identification process is enabled by the unique characteristic of the COR – a COR is on the new route, and it must also contain the previous route entry of the concerned MN. During the traversal of the CUIP route-update packets from the MN toward the COR on the new route, the routers check whether there is an existing entry, excluding the default-route entry, for the MN in the MRT. If so, that entry corresponds to the previous route for the MN, and therefore the core router is the COR for this handoff; otherwise it is a NIR. Figure 5 depicts this algorithm.

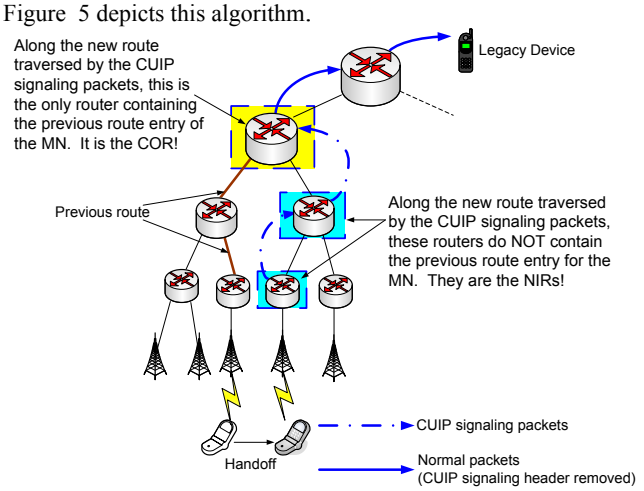

Figure 5. Illustration of the COR self-identification algorithm

3.6 CUIP Route-Update Mechanisms

The *Cellular Universal Roaming Update* (CURU) and the *Cellular Universal Handoff Update* (CUHU) are two route-update schemes CUIP uses to update the MRTs along the new and the previous routes of the MNs after a roaming and handoff, respectively, has occurred. We will first define the IPv6 hop-by-hop option header that enables them.

3.6.1 CUIP IPv6 Hop-by-Hop Option Header

Based on the guideline given in [12], Table 1 shows the definition and format of the CUIP Hop-by-Hop Option header.

The three highest-order bits (111) in the option type field are defined in [12] to provide the following two features:

1. When an IPv6 node does not recognize this option type, it must discard the packet and send an "ICMP Parameter Problem", Code 2, message to the packet's source address. Therefore, if the CUIP route-update packet reaches a node that does not support CUIP, the packet must be discarded, and an ICMP Parameter Problem message must be returned to the MN. The MN can then choose to notify the user of its inability of maintaining the session or to use other schemes (e.g., MIPv6) to continue the session if possible.

2. The option data can be changed en-route. This is set so that the COR can remove this option header after processing the CUHU route-update packets.

The next five bits in the option type field (00011) simply serves as the identification of CUIP. The option data length field specifies the length of the option data in number of bytes, which is unity in our case. There are two possible values of option data, 0 or 1, representing "Notification" or "Acknowledgement". The usage of these values will become clear as we describe the route-update schemes in the following subsections.

Table 1. CUIP IPv6 Option Header Definition

Option Field	Field Name	Field Length	Field Value
Type	CUIP Option	8 bits	11100011
Data Length	-	8 bits	00000001
Data	Notification	8 bits	00000000
	Acknowledgement	8 bits	00000001

3.6.2 CUIP Roaming Update (CURU)

CURU enables a user to roam from one WAR to another. Its main duty is to update the MRTs along the new path up to the COR, and from the COR to the previously associated WAR along the previous route. Note that CURU is only designed for roaming purpose which is relatively insensitive to delay. Figure 6 depicts the path that CURU route-update packet travels.

CURU makes use of the IPv6 hop-by-hop option header to handle the MRT update. As soon as an MN is notified by the link layer [1] that it has roamed to a new route, it will send an empty packet with the CUIP Notification header, which is the CUIP option header with the data field set to "Notification", to the newly associated WAR. Inside this CUIP Notification embedded packet (referred to as CURU-N packet hereinafter), the source address is set to the universal address of the MN. Since the destination of the CURU-N packet is supposed to be the location of the COR, which is not known at this point, the destination field of the CURU-N packet must be set to the special CURU address (SCA) so that the routers still "know" that the packet must be forwarded to the COR of this handoff. Note that the SCA must be an address that is not normally routable, and should eventually be reserved by IANA [13]. For the sake of illustration, we tentatively assign "::AA11/128" to the SCA for now. It is worth noting that the use of the SCA in the destination field will not cause incompatibility with normal IPv6 networks because the CURU-N packet will simply be dropped by routers without CUIP support. Furthermore, the SCA is used in CURU for signaling purpose only, not for routing.

When a router receives a CURU-N packet from a handing-off MN, which can be identified by the existence of CUIP Notification header and the SCA in the destination address field, the CURU mechanism takes place. If the router is the COR, it will modify the corresponding entry in the MRT so that incoming packets for the MN will be forwarded to the appropriate route accordingly. The COR then forwards the CURU-N packet to the PIR or HPIR through the interface of the previous route, so that the corresponding route entry in the previous route can be removed or modified, respectively. If the router is not a COR, it is a NIR. It will record to the MRT the interface from which the packet is received, so that incoming packets for the MN will be directed to this interface. This NIR then forwards the CURU-N packet to the upper-level core router through the default route.

To ensure the signaling packet does get to the previous WAR correctly, the MN will resend the CURU-N packet at a one second interval, until the corresponding CUIP Acknowledgement embedded packet (referred to as CURU-A packet hereinafter) returned by the previous WAR is received. Note that the routers along the new route, if already containing the MN's routing information, can simply ignore the CUIP Notification header in the subsequent CURU-N packets sent from the same MN.

After the CURU route-update process has been completed, all the future packets will be routed to the MN accordingly.

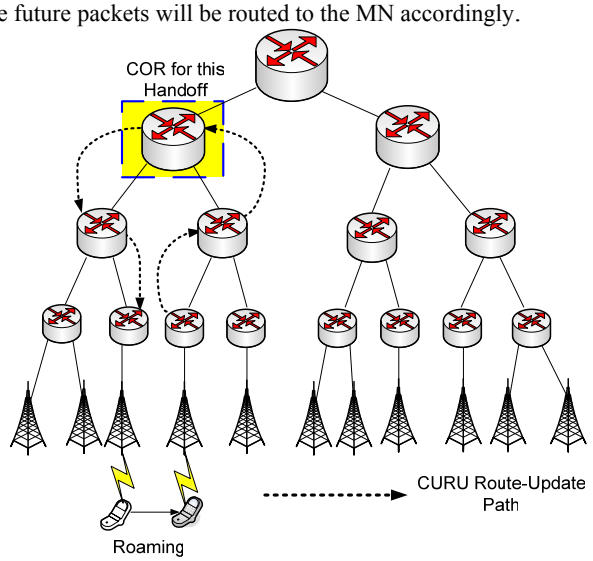

Figure 6. Basic idea of CURU mechanism

3.6.3 CUIP Handoff Update (CUHU) – Handoff-on-the-fly

CUHU implements the concept of "handoff-on-the-fly". In CUHU, the CUIP Notification header is embedded within the outgoing data packets (referred to as CUHU-N packet hereinafter). That is, the route-update information is piggybacked on the data packets (e.g. voice packets) through the hop-by-hop option header. The main duty of CUHU is to update the MRT of the routers along the new route after handoff, up to the COR. Once the CUHU-N data packets reach the COR, the CUIP Notification header will be removed from them, so that they will travel from the COR to the CN just like normal packets. This ensures the backward compatibility of CUIP with legacy devices on the Internet. Figure 7 depicts the CUHU mechanism.

As soon as the MN detects a handoff, if the MN is actively sending data, CUHU will be initiated and CUIP Notification header will be embedded in the outgoing data packets to form the CUHU-N packets. Unlike the CURU case, the source and destination addresses of the CUHU-N packets will be set to the MN's address and the CN's address respectively. That is, the source and destination addresses of the active session are not changed. When a router receives a CUHU-N packet, which can be identified by the existence of CUIP Notification header in the data packet, the CUHU mechanism takes place. The mechanism is similar to CURU described above, except that the COR will remove the CUIP option header before forwarding the packet toward the destination. To ensure that the CUHU-N packet gets to

the COR, the MN will continue to embed the CUIP Notification header into the outgoing data packets until it receives the CUIP Acknowledgement embedded packet returned from the COR. Similar to CURU, the routers along the new route, if already containing the MN's routing information, can simply ignore the CUIP Notification header in the subsequent CUHU-N packets sent from the same MN.

It is worth noting that CUHU only updates the MRTs along the new route up to the COR. Modifications of the MRTs on the previous route, as required for user roaming purpose, will be immediately initiated by the COR through a "half-way" version of CURU. Strictly speaking, after the first CUHU-N packet reaches the corresponding COR, although the route-update is not yet completed on the previous route, the handoff is already completed and the incoming packets from the CN can already be routed to the MN accordingly.

Note that if the CN happens to be on the MN's previous route, before the CURU completes the update of the PIRs, its packets sent to the MN will fail the route query at the PIRs. These packets will be forwarded to the upper level core router through the default route until they hit the COR, and from there they will be directed to the new route and eventually to the MN.

If the MN is not actively sending data at the time of handoff, CURU route-update will be initiated instead.

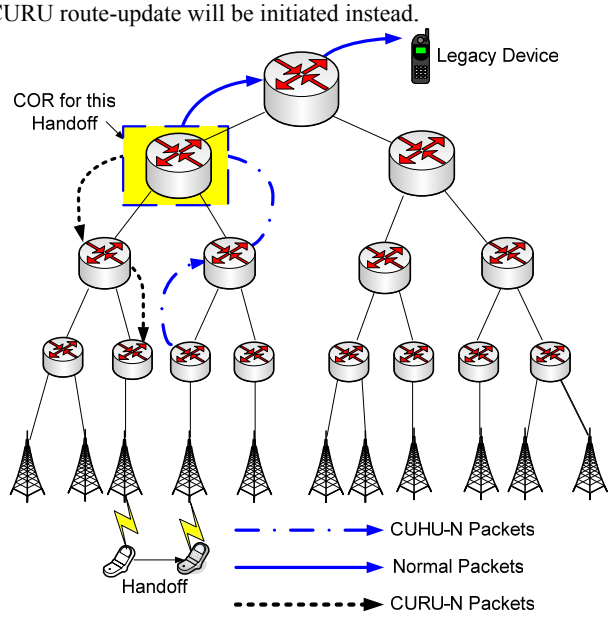

Figure 7. Basic idea of CUHU mechanism

3.7 Macro-mobility for CUIP

We have focused on handoff scenarios within a single wireless access network so far, which can be considered as micro-mobility. We now look at how CUIP handles macro-mobility.

Recall that all TLCRs are assumed to be directly connected in a meshed manner, and therefore they all have a route entry leading to each other. It can be seen from Figure 8 that, when an MN is handed off across multiple wireless access networks, which is considered as macro-mobility, the two wireless access networks can actually be logically viewed as one hierarchical structure with the home TLCR acting as the COR for the handoff, whereas the foreign TLCR as an NIR. When the foreign TLCR realizes that it is not the COR for a handoff after receiving a CUHU-N/CURU-N route-update packet, it will forward the route-update packet to the MN's home TLCR determined by the source address in the packets. Consequently, the home TLCR will always be updated for the whereabouts of its MNs. Note also that the foreign TLCR is only required to do this for the CURU-N/CUHU-N route-update packets. For normal outgoing packets, the foreign TLCR can forward them directly to the CN.

When the home TLCR receives the CUHU-N/CURU-N route-update packets from adjacent TLCRs, it will take the responsibility to be the COR for this handoff, and update the MRTs along the previous route accordingly. Note that the previous route of a macro-mobility handoff could be in a foreign network. Figure 8 shows the case when CUHU is used. It is important to observe that, after macro-mobility handoff, the concept of COR will also be applicable in the foreign network. That is, the home TLCR will only be involved when another macro-mobility handoff occurs.

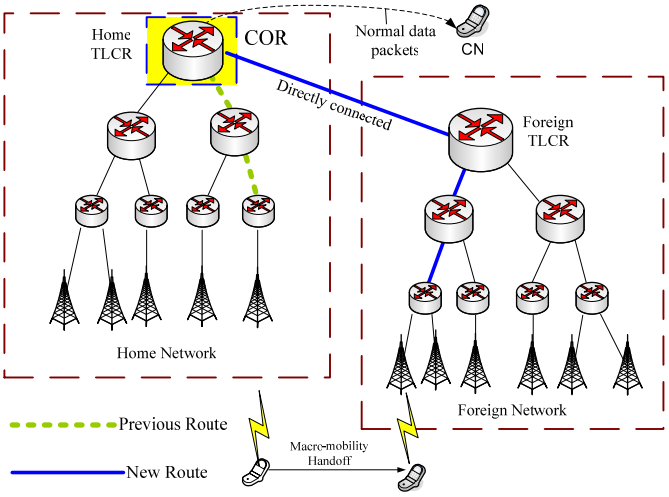

Figure 8. Macro-mobility can be logically viewed as handoff within one hierarchical structure

Therefore, unlike most other MIPv6 based micro-mobility schemes, the macro-mobility of CUIP is enabled through CUIP itself and the MNs are still addressed by the universal address. In other words, as long as the originating and destination wireless access networks are CUIP enabled, macromobility can be manipulated without relying on MIPv6.

Finally, one may have noticed that a looping situation could occur during macro-mobility. This is caused by the fact that all TLCRs are directly connected together, and therefore each of them would have a route entry pointing to each other. Imagine that a MN is handed off from its home network to a foreign network. The home TLCR, after CUHU/CURU route update mechanism, will direct all the incoming packets for the MN to the foreign TLCR whose network underneath is now serving the MN. At the same time this foreign TLCR is supposed to route the packets for this MN back to the home network because there is a route entry for the MN's prefix pointing to the its home network. Thus, a loop has occurred. CUIP handles this situation by having CURU/CUHU turn on the "V" flag on the entry corresponding to this MN's prefix in the MRT whenever the signaling packets passes through a foreign TLCR. As a result, when the foreign TLCR sees the "V" flag in the query result of an incoming packet, it will avoid looping the packet back to the home TLCR, and will

search for the MN's entry in the MRT and forward the packet to the MN accordingly. Figure 9 depicts this idea.

		Fixed Routing Part			
1	**Destination**	**Gateway**	**Flags**	**Interface**	**Away**
2	bbdd::fe10:0/112	bbdd::fe10:0001		eth1	
3	3810:d012::0/64	3810:d012::0	V	eth4	3810::fe12:0123/128
4					
5		**Mobility Routing Part**			
6	**Destination**	**Gateway**	**Flags**	**Interface**	
7					
8		**Visitor Routing Part**			
9	**Destination**	**Gateway**	**Flags**	**Interface**	
10	3810::fe12:0123/128	bbdd::fe10:0001		eth1	

Figure 9. A sample MRT in a TLCR -- the "V" flag indicates that a MN from a neighbor TLCR is visiting this network

3.8 Security for CUIP

We now consider a security scheme that is largely based on the security architecture proposed by 3GPP [18]. In 3GPP's proposal, a subscriber is identified by a globally unique international mobile subscriber identity (IMSI) as the permanent identifier which is stored on a user service identity module (USIM) inserted into the MN. In CUIP, we will replace the IMSI with CUIP-IMSI, which is similar to the IMSI, but with the universal IPv6 address appended to it. Since the universal IPv6 address also serves as a globally unique identifier of the network layer, appending such an address into the IMSI allows the mutual authentication procedures used in [18] to be applied on the network layer as well. The following two entities defined in 3GPP's security architecture [18] will be used with minor modifications to adapt to CUIP.

1. The home location register (HLR) -- An HLR is associated with every wireless access network and is responsible for providing identity information of the subscribers for whom it represents their home networks.
2. The authentication center (AuC) -- An AuC is associated with every HLR and is responsible for storing the identity keys for the MNs registered with the HLR.

Figure 10 depicts the security mechanism in the CUIP network, which is evolved from the one given in [18]. When a MN is handed off to a foreign network, the foreign WAR invokes the authentication procedure by requesting the MN to identify itself through the user identity request. The MN then provides the CUIP-IMSI with the user identity response to it. The WAR will transmit the received CUIP-IMSI to the MN's corresponding HLR/AuC at its home network. Note that the USIM and the AuC share a master key *K*. Based on *K* and the CUIP-IMSI, the mutual authentications of both the link layer and network layer can be carried in one step. After that, the IPv6 packets from the MN will be trusted by the foreign network and vice versa. Ingress filtering of source IP address [19], which is normally used to defend the network from attacks initiated by bogus IP addresses, is therefore not necessary. The rest of the security procedures shown in Figure 10 simply follow what have been defined in [18] and we will not provide the details in this paper.

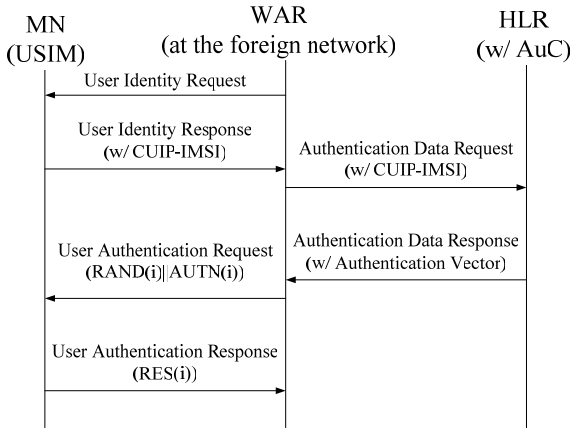

Figure 10. Authentication mechanism in CUIP based Wireless Access Network

3.9 Interoperability for CUIP

CUIP is only required to operate in the wireless access networks that intend to provide seamless IP mobility. The public Internet and the fixed line IP networks are not required to support it in order to communicate with CUIP based wireless access networks.

CUIP can also be backward compatible with MIPv6 seamlessly. Assuming that a CUIP based MN and the corresponding home TLCR are also MIPv6 capable. When MIPv6 mode is enabled, the universal address will serve as the home address and the home TLCR will serve as the home agent of the MN. For example, imagine that a CUIP enabled MN moves into a foreign network that supports MIPv6, but not CUIP. All the CURU/CUHU packets will be dropped by the non-supporting routers, but the MN will then receive the periodic router advertisements [20] from this foreign network. The MN can switch to MIPv6 mode and begin the CoA acquisition, home agent binding update (with its home TLCR) and other MIPv6 procedures defined in [4]. Similar procedures can be applied to other MIPv6 related schemes as well.

4. Performance Analysis for CUIP

4.1 Handoff Latency Analysis

One of the key characteristics of CUIP is that the handoff will be completed as soon as the first CUHU-N route-update packet reaches the corresponding COR on the new route. Since updating the routers on the previous route is not required for the handoff to complete, we will not consider the latency introduced on the previous route here.

Let us consider a general handoff scenario depicted in Figure 11, there are a total of *N* routers from the new WAR to the COR inclusive along the new route, and the handoff is said to be updating *N* routers. Assume the time to update the MRT in a router is T_{MRT}, the transmission delay between the MN and the base station is T_{AIR}, and that between each hop along the network, including the path between the BS and the WAR, is T_{HOP}. The total handoff delay is then

$$T_{HANDOFF\_CUIP} = N * T_{MRT} + N * T_{HOP} + T_{AIR} \quad (1)$$

For a reasonably sized routing table in a common router, the average time for inserting/deleting an entry is about 4μs [2]. Since

the entry insertion/deletion operation in MRT is identical to those in the traditional routing table, we can assume $T_{MRT} \approx 4\mu s$. According to [14], T_{AIR} is typically around 20ms, and according to [15], T_{HOP} is about 153μs in an operational backbone network. Plugging these values into (1), the handoff latency is approximately given by (in ms)

$$T_{HANDOFF\_CUIP} = N * 0.157 + 20 \quad (2)$$

From (2), we see that the handoff latency in CUIP only depends on N, the number of routers being updated in a particular handoff scenario. We are therefore interested in finding the expected value of N, or the average number of routers updated per handoff, denoted by $E[N]$.

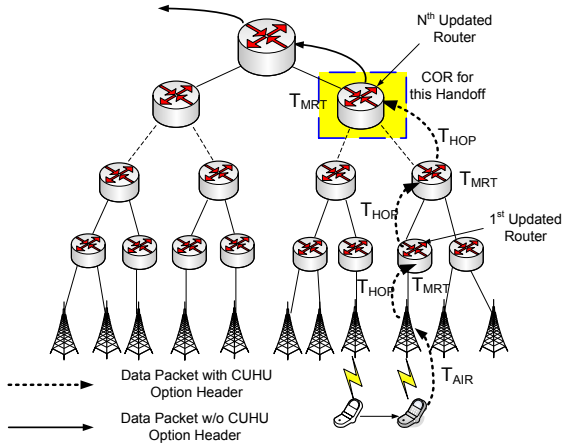

Figure 11. Sample handoff scenario for performance analysis

In the following subsections, we will prove that $E[N]$ is upper bounded by three. The proof proceeds in two steps. First, we calculate $E[N]$ based on the assumption of a one dimensional (1-D) *balanced* hierarchical structure. We obtain an upper bound of three for $E[N]$. Then we prove by mathematical induction that the upper bound is also applicable to a 1-D *unbalanced* hierarchical structure. A uniform probability distribution of handoff scenarios is assumed in all the analyses. That is, all handoff scenarios in the wireless access network are equally likely to happen.

4.1.1 1-D Balanced Hierarchical Structure

Consider the 1-D hierarchical network architecture with a balanced placement of routers as shown in Figure 12. Each node in the structure corresponds to a router, and we assume that each node has exactly M child nodes and the entire hierarchy has a depth of K levels. Therefore at the lowest level, where all the handoffs occur, there are a total of M^K WARs.

In this subsection, we assume that (i) an MN can only be handed off to adjacent WARs, and (ii) consider a "ring" fashioned structure so that the rightmost WAR in the hierarchy is considered to be adjacent to the leftmost WAR.

Let P_i be the probability of the occurrence of a handoff scenario which needs to update i or more routers on the new route (or having a COR at level K-i+1 or higher). That is, $P_i = \Pr[N \geq i]$. We therefore have

$$E[N] = \sum_{i=1}^{\infty} P_i \quad (3)$$

Since every network layer handoff needs to update a minimum of two routers, $P_1 = P_2 = 1$. Imagine that a MN begins its journey from the leftmost WAR toward the right hand side in Figure 12, a handoff scenario that updates three or more routers occurs every M WARs apart. Therefore $P_3 = 1/M$. Similarly, a handoff that updates four or more routers occurs every M^2 WARs apart. Therefore $P_4 = 1/M^2$. In general, we have $P_i = 1/M^{i-2}$ for $2 \leq i \leq K+1$. Note that the maximum number of routers being updated in a K level hierarchy is $K+1$. From (3),

$$E[N] = 2 + \sum_{i=3}^{K+1} \frac{1}{M^{i-2}} = 2 + \sum_{i=1}^{K-1} \frac{1}{M^i} < 3 \quad (4)$$

for $M \geq 2$ and $K \geq 2$. In (4), we have made use of the fact that $\sum_{i=1}^{K-1} 1/M^i < 1, \forall M \geq 2$. Note that we have taken into account the wrap-around scenario from the rightmost WAR to the leftmost WAR in P_{K+1}. Finally, we have only counted the number of handoff scenarios in one direction, i.e., from left to right. By symmetry, if we had counted the scenarios in both directions, we would have arrived at the same result. This is sufficient under the balanced deployment assumption due to the symmetry of both directions. In the following asymmetric deployment architecture, we will need to count the handoff scenarios in both directions.

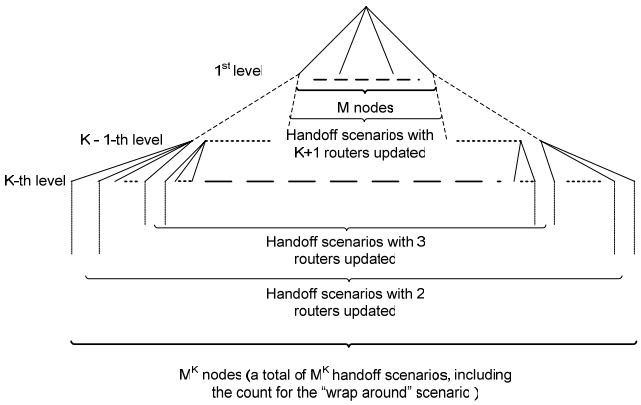

Figure 12. Derivation of *E[N]* under a 1-D balanced hierarchical tree structure

4.1.2 1-D Unbalanced Hierarchical Structure

We now derive $E[N]$ for the 1-D unbalanced hierarchical structure. In an unbalanced hierarchical structure, we assume that each router can have an arbitrary number of children nodes, denoted by Y. We further assume that $Y = 0 \, or \, Y \geq 2$ so that, if there are children nodes underneath a router, at least one handoff scenario can occur between them. Replacing Figure 12 with Figure 13, we now have an unbalanced hierarchical structure.

We now prove by mathematical induction that $E[N] < 3$ for all K. For $K = 1$, we see that $E[N] = 2$, and therefore less than 3. Assume $E[N] < 3$ is true for $K = d$, and denote E_d to be $E[N]$ at level $K = d$, then we need to prove that $E_{d+1} < 3$ also holds.

Let us separate the handoff scenarios into two groups. The first group refers to all the handoffs updating only the routers at lowest 2 levels in the hierarchy, i.e., updating exactly two routers. We denote the total number of handoff scenarios belonging to this group by S_2. The other group refers to all handoffs updating three or more routers. We denote the total number of handoff scenarios

belonging to this group by S_{UPPER}. Figure 13 illustrates this idea.

Now we can write

E_{d+1} = Prob{a handoff needs to update exactly 2 routers}*2 +
Prob{a handoff needs to update 3 or more routers}*
(expected number of routers updated for this handoff)

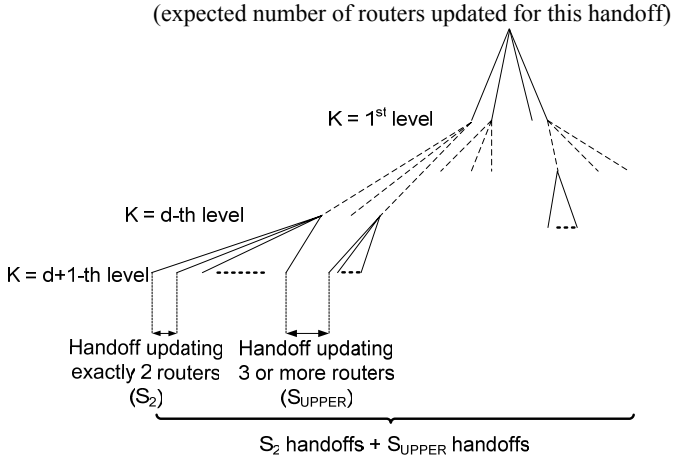

Figure 13. Derivation of *E[N]* under a 1-D unbalanced hierarchical tree structure

With our definitions of S_2 and S_{UPPER}, plus the fact that the average number of routers updated at level d+1 is one more than the average number of routers updated at level d, we have

$$E_{d+1} = \frac{S_2}{S_{UPPER} + S_2} * 2 + \frac{S_{UPPER}}{S_{UPPER} + S_2}(E_d + 1) \quad (5)$$

Since we have assumed $E_d < 3$, (5) becomes

$$E_{d+1} < 2 + \frac{S_{UPPER}}{S_{UPPER} + S_2} * 2 \quad (6)$$

Since there are at least two children nodes per intermediate router, we must have

$$S_{UPPER} < S_2 \quad (7)$$

The validity of (7) can be seen by constructing the worst case unbalanced hierarchy of routers in which S_2 is minimized, as shown in Figure 14. In Figure 14, there are two child nodes per parent router. We can easily see that, for an L level hierarchy, $S_2 = L + 1$ and $S_{UPPER} = L - 1$. Obviously, (7) is valid even when S_2 is minimized. Substitute (7) into (6), we obtain $E_{d+1} < 3$. By induction, $E_K < 3$ is true for all K.

Recalling that macro-mobility can logically be viewed as a handoff inside one hierarchy (see Figure 8), the same result given above therefore accounts for macro-mobility as well. Furthermore, the key assumption in the above proof is the condition $S_{UPPER} < S_2$ in (7). In fact, careful examination shows that in no other places did we assume the 1-D structure. Therefore, we can easily draw the same conclusion, using mathematical induction based on the similar principle given above, for 2-D hierarchical structures. Due to space limitation, however, the details of the proof are not provided in this paper.

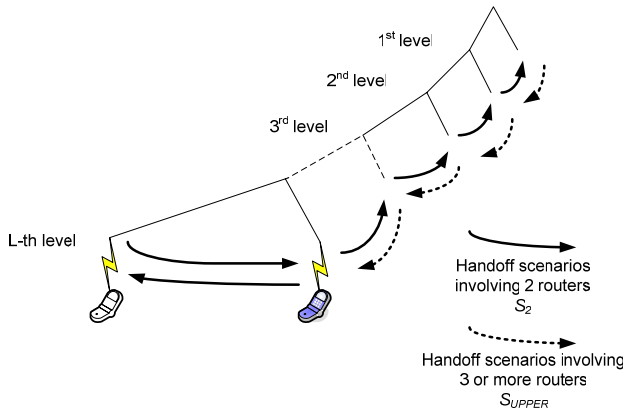

Figure 14. Worst case scenario for unbalanced hierarchy of routers in which S_2 is minimized

4.2 Implications

4.2.1 Minimal handoff delay achieved

From the analyses given above, we can conclude that the expected number of routers involved in each handoff is upper-bounded by three. If we apply this upper bound into (2), we have

$$E[T_{HANDOFF\_CUIP}] < 3 * 0.157 + 20 = 20.471\,ms \quad (8)$$

The expected latency per handoff is approximately 20.5ms at the network layer, which is considerably lower than the tolerable delay range of real-time traffic (~150ms).

4.2.2 Impact of IP mobility on QoS reduced

As per-flow quality of service (QoS) is an important parameter in real-time multimedia applications, it must also be considered. Ramjee *et al.* [10] has pointed out that per-flow QoS reservation requires the identification of the addresses at both endpoints of a flow, and protocols like RSVP also assumes the consistency of endpoint addresses. It is therefore reasonable to suggest that the nature of universal addressing will be beneficial to QoS support in the highly mobile environment.

Imagine that a handoff has occurred in a *K*-level hierarchical wireless access network similar to the one depicted in Figure 13. As we have proven in the previous subsection that, with CUIP, each handoff on average needs to update the QoS parameters in less than three routers regardless the value of *K*. Due to the universal addressing nature of CUIP, the handoff and the QoS parameter changes are transparent to the rest of the network. Thus, we consider that CUIP is capable of supporting QoS with minimal overhead.

5. Conclusion

This paper has introduced a novel IP mobility scheme for the wireless access networks, *Cellular Universal IP* (CUIP), in which each MN is assigned a universal IPv6 address at subscription that is invariant under mobility. With CUIP, the traditional use of care-of address and IP tunneling are no longer required. Unlike most existing IP mobility schemes, CUIP does not rely on explicit handoff signaling mechanism, neither does it rely on MIPv6 for macro-mobility handoffs. As a result, the excessive handoff delay and triangular routing of IP mobility are eliminated.

CUIP minimizes the latency of IP mobility based on two observations:

1. Every IP handoff under a hierarchical network structure has an associated COR, and that only the routing tables of the COR and the routers below it need to be updated.
2. The routers to be updated for handoff are along the data path. Therefore, signaling can be piggybacked on outgoing data packets for more efficient handoff, particularly for real-time applications with continuous stream of data packets. The signaling delay is then proportional to the time interval between two consecutive data packets – that is, the signaling delay scales naturally with the blackout delay requirements of the data stream.

Based on the above observations, CUIP accomplishes IP routing update by an intelligent *handoff-on-the-fly* route-update scheme that handles the change of routes without explicit signaling. We have shown that the expected number of routers to be updated per handoff is upper bounded by three, even when macro-mobility is taken into account. Analytically, we conclude that the scheme requires about 20.5ms on average to complete a handoff at the network layer. The fact that MNs can always be addressed with a universal address also reduces the impact of IP mobility on QoS management. We therefore believe that a universal addressing scheme, such as CUIP, can be an alternative for IP mobility to the two-tier addressing scheme currently being used by mobile IPv6 and the schemes based on it.

Finally, we have suggested a security mechanism evolved from the one proposed in 3GPP to handle the mutual authentication between the MNs and the CUIP network. Ingress filtering of source IP addresses can then be avoided.

6. REFERENCES

[1] R. Koodli, “Fast handovers for mobile IPv6,” draft-ietf-mobileip-fast-mipv6-08.txt, IETF, 2003.

[2] V. Srinivasan and G. Varghese, “Faster IP lookups using controlled prefix expansion,” ACM Sigmetrics, 1998.

[3] C. Perkins and D. Johnson, "Route optimization in mobile IP," Internet draft, draft-ietf-mobileip-optim-08.txt, IETF, Feb. 2000.

[4] D. Johnson, C. Perkins and J. Arkko, “Mobility support in IPv6,” RFC 3375, IETF, June 2004.

[5] S. Thomson, T. Narten, "IPv6 stateless address autoconfiguration", RFC 2462, IETF, December 1998.

[6] H. Yokota, A. Idoue, T. Hasegawa, and T. Kato, “Link layer assisted mobile IP fast handoff method over wireless LAN networks,” MOBICOM, 2002.

[7] T. Janevski, Traffic Analysis and Design of Wireless IP Networks. Artech House, 2003, ch. 2 and ch. 5.

[8] H. Soliman, C. Castelliccia, K. El-Malki, L. Bellier, “Hierarchical mobile IPv6 mobility management (HMIPv6),” Internet Draft, draft-ietf-mobileip-hmipv6-08.txt, IETF, June 2003.

[9] A. Valko, “Cellular IP: A New approach to internet host mobility,” ACM SIGCOMM Computer Communication Review, 29(1):50--65, January 1999.

[10] R. Ramjee, K. Varadhan, L. Salgarelli, S. Thuel, S. Wang and T. Porta, “HAWAII: A domain-based approach for supporting mobility in wide-area wireless networks,” IEEE/ACM Transactions on Networking, June 2002.

[11] P. Lam and S. C. Liew, "UDP-Liter: An Improved UDP Protocol for Real-Time Multimedia Applications over Wireless Links," The 1st IEEE ISWCS, Sept 2004.

[12] S. Deering and R. Hinden, “Internet protocol, version 6 (IPv6) Specification,” RFC 2460, IETF, December 1998.

[13] IANA, The Internet Assigned Numbers Authority, http://www.iana.org/

[14] D. Wisely, P. Eardley and L. Burness, IP for 3G: Networking technologies for mobile communications. John Wiley & Sons, 2002, ch. 6.

[15] K. Papagiannaki, S. Moon, C. Fraleigh, P.Thiran, F. Tobagi, C. Diot, “Analysis of measured single-hop delay from an operational backbone network,” IEEE Infocom, June, 2002.

[16] N. Nakajima, A. Dutta, S. Das and H. Schulzrinne, “Handoff delay analysis and measurement for SIP based mobility in IPv6,” IEEE ICC, May 2003.

[17] J. Vatn, "Long random wait times for getting a care-of address are a danger to mobile multimedia", IEEE International Workshop on Mobile Multimedia Communications (MoMuC'99), November 1999.

[18] 3GPP TS 33.102 v5.5.0, “3GPP; technical specification group services and systems aspects; 3G Security; Security Architecture (Release 5),” 3GPP, 2004.

[19] P. Ferguson and D. Senie, “Network Ingress Filtering: Defeating Denial of Service Attacks which employ IP Source Address Spoofing,” RFC 2267, IETF, January 1998.

[20] T. Narten, E. Nordmark and W. Simpson, "Neighbor Discovery for IP Version 6 (IPv6)", RFC 2461, IETF, December 1998.