A Next Generation QoS Signaling Protocol for IP-based Mobile Networks

Sung-Hyuck Lee*, Seong-Ho Jeong**, Byoung-Joon Lee*, and Jongho Bang*

Abstract— **The mobility of an IP-based node affects routing paths, and as a result, can have a dramatic effect on protocol operation and state management. In this paper, we analyze the effects of node mobility on the interaction between QoS signaling protocols (e.g., IETF NSIS protocols) and mobility management protocols, and how the protocols operate in different mobility scenarios. We also propose an efficient QoS signaling protocol which operates adaptively in IP-based (especially, IPv6-based) mobile networks. The key features of the protocol include crossover node discovery and local repair to achieve seamless services. Our simulation and experimental results show that the proposed/implemented protocol works well in mobile environments.**

keywords— **Resource Reservation, Mobile IP, Crossover Node**

I. Introduction

Ambient Networks (AN) aiming at an innovative new network vision based on the dynamic composition of networks requires a solution that provides easy-to-use, rich and trustworthy multimedia communication services for all. In what QoS is concerned, the AN also aims to provide dynamic QoS control in ubiquitous mobile environments. In this context, the mobility-aware QoS signaling is a crucial element of AN's QoS functionalities to achieve seamless QoS between ANs [1].

The IETF Next Steps in Signaling (NSIS) working group is standardizing a signaling protocol suite with QoS signaling (which AN aims) as the first use case. The overall signaling protocol suite is decomposed into a generic lower layer with separate upper layers for signaling applications. The lower layer, NSIS Transport Layer Protocol (NTLP), is intended to provide a generally useful transport service for such signaling messages. The actual signaling messages are in general originated within upper layer signaling applications, each having its own NSIS Signaling Layer Protocol (NSLP), and the role of the NTLP is primarily to move these messages around the network to the appropriate nodes. The general description of the NSIS protocol suite, including its two-layer architecture, can be found in [2].

One of the important features that the NSIS protocol needs to provide is mobility support [3]. In highly mobile environments, frequent handovers may result in a significant degradation of QoS performance if the wireless/mobile access network is unable to provide enhanced solutions for prompt QoS re-establishment. Especially, how QoS signaling interacts with Mobile IP may have a significant impact on QoS performance. This paper mainly identifies the key issues on the interaction between Mobile IPv6 [10] and QoS signaling, and proposes desired QoS signaling functions for mobile environments. The rest of the paper is organized as follows: Sections II and III describe the proposed protocol called mobility-aware QoS signaling protocol (MQSIG), followed by performance results and concluding remarks in sections IV and V, respectively.

*Comm. & Network Lab., Samsung Advanced Institute of Technology
E-mail: {starsu.lee, bj33.lee, jh0278.bang}@samsung.com
**Dept. of Information and Communications Engineering, Hankuk Univ. of Foreign Studies. E-mail: shjeong@hufs.ac.kr.

II. Impact of Mobility on QoS Signaling Protocols

IP-based mobility itself includes topological changes due to the change of network attachment point. Topological changes entail the change of routes for data packets sent to or from a mobile node (MN) and may lead to the change of host IP addresses. These changes of route and IP addresses in mobile environments are typically much faster and more frequent than traditional route changes and may have some significant impact on QoS signaling protocols.

Although the well-known resource reservation protocol, RSVP [4], is able to setup resource reservation for real-time traffic in the wired Internet, it is not adequate to reserve resources in mobile networks. For example, a change in the location of the MN may make the reserved resources on the old path useless and a new reservation on the new path has to be established while maintaining the old reservation (this is called 'double reservation problem'). This results in the inefficient use of network resources and also introduces an additional delay due to end-to-end signaling.

To overcome such drawbacks of RSVP, many solutions have been proposed, which are mostly based on the modification or extension of RSVP [5,6,7]. However, most of RSVP-based solutions do not have yet appropriate QoS mechanisms for preventing service disruption during handover.

To develop a mobility-aware QoS signaling protocol which solves the problems, it is necessary to analyze the key differences between generic route changes and mobility. The generic route changes (hereafter 'route changes') occur due to load sharing, load balancing, or a link (or node) failure, but the mobility is associated with the change of the network attachment point. These will cause divergence (or convergence) between the old path where QoS state has already been installed and the new path where data forwarding will actually happen. Although the mobility can be considered similar to the route changes, the main difference is that the flow identifier may not change after the route changes while the mobility may cause the change of flow identifier by having a new network attachment point. Since the reservation session should remain the same after a mobility event, the flow identifier should not be used to identify any signaling application session [2].

In general, a mobility event results in creating a common/unchanged path, an old path, and a new path. The old and new paths converge or diverge according to the direction of each signaling flow as shown in Fig. 1. Such topological changes make the signaling state established on the old path useless, and thus it should be removed (in the end). In addition, the existing QoS state should be re-established along the new path as fast as possible and updated along the common path.

To minimize the impact of mobility on seamless QoS service and to improve the scalability of signaling, QoS signaling for state re-establishment should be localized within the affected area. This localized signaling procedure is referred to as local repair in this paper. The major issue in this case is to find a node which performs the local repair. One of the most appropriate nodes for the local repair is the crossover node (CRN) where the old and new session paths meet.

III. An IP Mobility-Aware QoS Signaling Protocol

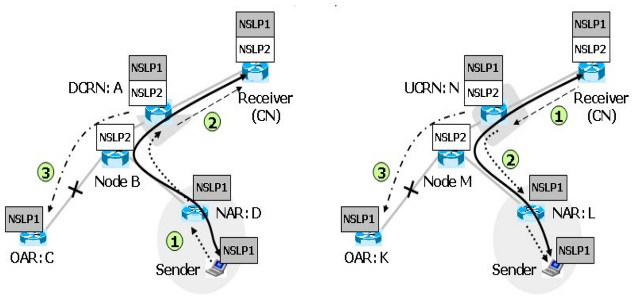
In this section, we propose an efficient QoS signaling protocol (MQSIG) which operates adaptively in IPv6-based mobile networks. The key features of MQSIG include CRN discovery and local repair.

A. CRN Discovery

The CRN discovery can be performed according to whether the discovery is coupled with the transport of signaling application messages (i.e., a coupled approach and an uncoupled approach). Generally, the coupled approach would be preferred to the uncoupled approach to reduce the signaling delay. In this paper, the CRN discovery and local repair are based on the coupled approach. We also assume that the CRN discovery is considered as an extension to the peer discovery at the NTLP layer to reduce overall processing overhead [3].

1) Identifiers for CRN Discovery: To discover the CRN in a fast and efficient manner, the following basic identifiers are required: session identifier (Session_ID), flow identifier (Flow_ID), signaling application identifier (NSLP_ID), and NSLP branch identifier (NSLP_Br_ID).

The Session_ID is contained in the NTLP message and used to easily identify the involved session because it remains the same while the Flow_ID may change after handover. Note that the uniqueness of Session_ID is one of the keys features to solve the double reservation problem. On the other hand, the Flow_ID is used to specify the relationship between the address information and the QoS state re-establishment. In other words, the change of Flow_ID indicates topological changes, and therefore it represents that the state along the common path should be updated after the CRN is discovered.



(a) Downstream flows (b) Upstream flows

Fig. 1. The topology for NSIS signaling caused by mobility

TABLE I
Routing State Table at Node A

Message Routing Information	Session ID	NSLP ID	Upstrea m peer	Downstr eam peer	NSLP Br. ID
Method = Path Coupled; Flow ID= {IP-#X, IP-#V, protocol, ports}	0xABCD	NSLP1		Pointer to C-A	1-D-#1
				Pointer to D-A	1-D-#2
			Z		1-U-#1
Method = Path Coupled; Flow ID= {IP-#X, IP-#V, protocol, ports}	0x1234	NSLP2		B	2-D-#1
			Z		2-U-#1

TABLE II
Routing State Table at Node N

Message Routing Information	Session ID	NSLP ID	Upstrea m peer	Downstr eam peer	NSLP Br. ID
Method = Path Coupled; Flow ID= {IP-#X, IP-#V, protocol, ports}	0xABCD	NSLP1	Pointer to N-K		1-U-#1
			Pointer to N-L		1-U-#2
				O	1-D-#1
Method = Path Coupled; Flow ID= {IP-#X, IP-#V, protocol, ports}	0x1234	NSLP2		Pointer to N-R	2-D-#1
			M		2-U-#1

The NSLP_ID is used to refer to the corresponding NSLP at the NTLP level, and in the context of CRN discovery it helps to discover an appropriate NSLP CRN which has installed the corresponding QoS state using the NTLP peer discovery message.

The NSLP_Br_ID is a virtual branch identifier and used to establish or delete NSIS associations between peer nodes. It is also used as an identifier to determine the CRN at the NTLP layer. The NSLP_Br_ID consists of the location information of peer nodes with the corresponding NSLP_ID by the procedure of NTLP message association [8]. For instance, as shown in Fig. 1 (a) and Table I, for the downstream direction (i.e., the direction from a data sender towards the destination) NSLP1 (state) of node A requires a messaging association for sending its messages towards node D after a route changes. In this case, NSIS entity (NE) A creates an NSLP_Br_ID for NSLP 1 towards node D and increases the NSLP_Br_ID counter to locally distinguish each virtual interface identifier between adjacent NSLP peers. That is, the corresponding NSLP_Br_ID is 1-D-#2: 1, D, and #2 indicate NSLP_ID-flow, flow directions

(Downstream (D) or Upstream (U)), and the counter number of branch, respectively. Note that this identifier would be more useful when the physical merging point of the old and new paths is not an NSLP CRN after the route changes as shown in Fig. 1 [3].

Optionally, the Mobility identifier as an object form can be used to inform of the handover status or a route change and therefore to expedite the CRN discovery. The Mobility object is defined in the NTLP (e.g., in GIMPS payload) [8] or NSLP messages to notify of any mobility event explicitly, and it contains various mobility-related fields such as mobility_event _counter (MEC) and handover_init (HI) fields. The 'MEC' field is used to detect the latest handover event to avoid any confusion about where to send a reservation confirmation message and to handle the ping-pong type of movement. The 'HI' field is used to explicitly inform that a handover is now initiated for fast state re-establishment [9].

2) The Procedures for CRN discovery: When a mobility event occurs, the CRN can be recognized by comparing the existing stored identifiers with the identifiers included in the NSIS peer discovery message initiated by an NSIS initiator (NI) (e.g., an MN or a CN). If an NSIS message is routed to an NSIS peer node, the node should check the following information (shown in Tables I and II) to determine whether it is a CRN:

- Whether the same NSLP ID exists
- Whether the corresponding CRN has already been discovered
- Whether the same Session_ID and Flow_ID exist
- Whether the NSLP_Br_ID has been changed; for example, as shown in Table I, for NSLP 1 it is changed to 1-D-#2 from 1-D-#1
- Optionally, whether any Mobility identifier exists; for example, the Mobility object may be examined to find out which message is sent due to the latest handover by checking the MEC field.

The CRN discovery can be further divided into downstream CRN (DCRN) discovery and upstream CRN (UCRN) discovery (owing to asymmetric routing) according to which node is a signaling initiator (by upstream or downstream), or whether the MN is a data sender. The procedure of DCRN discovery is similar to the creation procedure of the routing table of node N as shown in Table I and Fig. 1 (a), and the procedure of UCRN discovery is similar to Table II and Fig. 1 (b). Note that since the UCRN is determined by examining whether the outgoing path diverges or not, the UCRN discovery is more complex than the DCRN discovery [3].

B. Local repair Procedures

The CRN discovery procedures are different according to the direction of signaling flows in mobility scenarios, and therefore the procedures for local repair also differ depending on the direction of signaling flows: downstream local repair and upstream local repair.

For either type of local repair, the NSIS protocol needs to interact with mobility signaling protocols (e.g., Mobile IPv6), if any (during or posterior handover), to achieve fast re-establishment of the NSIS states along the new path. For example, the signaling protocol should interact with the binding process of Mobile IPv6 through several methods to immediately perform CRN discovery and the local repair [3].

In the downstream local repair, if resource availability is assured (after detection of mobility by Binding Update (BU) message), the MN initiates NSIS signaling for state setup towards a CN along the new path, and the DCRN discovery is implicitly done by this signaling (Fig. 1 (a)-①). The node where the old and new logical session paths converge realizes that it is a DCRN (e.g., node A) through the CRN discovery procedures described above, and afterward it sends a response message toward the MN to notify of the installed NSLP state. The DCRN then sends a refresh message towards the signaling destination to update the changed Flow_ID on the common path (Fig. 1 (a)-②), and it also sends a teardown message towards the old AR to delete the NSIS states on the obsolete path (Fig. 1 (a)-③).

In case of upstream local repair, the CN (or a HA) sends a refresh message toward the MN to perform local repair (Fig. 1 (b)-①). The UCRN is discovered implicitly by the CN-initiated signaling along the common path, and the node from which the common path begins to diverge into the old and new logical paths realizes that it is a UCRN (e.g., node N). In this case, the CN should be informed of the mobility event by detecting a change in its binding entry (BU message). After the UCRN is determined, it may send a refresh message to the MN along the new path while establishing the NSIS association between the updated peers (Fig. 1 (b)-②), and afterward the UCRN may send a teardown message toward the old AR to delete the NSIS state on the obsolete path (Fig. 1 (b)-③).

One of the goals of local repair is to avoid double reservations on all paths described in Section II. The double reservation made along the common path can be torn down by establishing a signaling association using the unique Session_ID and by updating packet classifier/flow identifier. In this case, the NSLP state should be shared for flows with different flow identifiers. After re-establishment of the NSIS state along the new path, the state on the obsolete path needs to be quickly removed by the local repair mechanism to prevent waste of resources (and resource allocation problem by call blocking). Although the release of the state on the old path can be accomplished by the timeout of soft state, the refresh timer value is quite long (e.g., default value of 30 s in RSVP [4]) and the maintenance of the obsolete state in mobile environments may not be necessary. Therefore, the transmission of a teardown message is particularly preferred to the use of refresh timer to delete the old state in a fast manner.

The release of old state on the obsolete path is also accomplished by comparing the existing and the new NSLP_Br_IDs. This will prevent the teardown message from being forwarded toward along the common path. However, whether the teardown message can be sent toward the opposite direction to the state initiating node is still for further study. This also leads to authorization problem because a node which does not initiate signaling for establishing the NSIS state can delete the state.

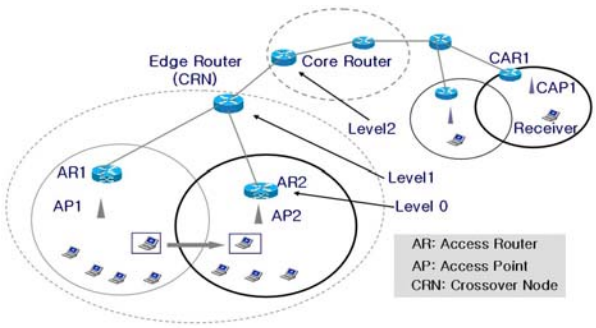


Fig. 2. A simulation topology.

IV. Performance Analysis

In this section, we evaluate the performance of MQSIG in terms of resource re-reservation delay on the new path after handover and blocking probability of requested sessions. Then, we compare the related performance of existing signaling protocols such as RSVP and RSVP-MP [6]. Some experimental results are also provided to demonstrate that the proposed/implemented signaling protocol works well in MIPv6-based mobile environments [10].

A. Resource Re-reservation Delay

We have performed simulation studies to measure the performance of RSVP, RSVP-MP, and MQSIG in terms of resource re-reservation delay. Fig. 2 depicts a simulation topology where there are 8 MNs. The number of hops from the MN and the CN is 7, and every MN may generate UDP traffic. It is assumed that the refresh period of RSVP and RSVP-MP is 30s. Initially, only one MN which communicates with the CN generates UDP traffic. The traffic load increases when MNs other than the current MN begins to generate UDP traffic. For example, the amount of added traffic load is 0.1 when another MN starts to generate traffic. Our simulation model is based on Marc Greis' RSVP model implemented in ns-2.1b3 and Rui Prior's RSVP model implemented in ns-2.26 which is an updated version of Marc Greis' model.

As depicted in Fig. 3, the proposed signaling protocol shows better performance in terms of signaling delay for resource re-reservation, compared to RSVP and RSVP-MP even when the traffic load increases. This is because MQSIG performs CRN discovery for localized signaling after handover and MIPv6 binding process is closely associated with QoS signaling for fast resource re-reservation. Furthermore, only QoS state update (not re-reservation) is performed on the common/unchanged path to minimize the signaling delay through local repair procedures.

B. Blocking Probability

To obtain the blocking probability of requested sessions, we used the analytical model in [6]. Based on the formula, in case of RSVP, the blocking probability is given as

$$P_b = \frac{A_{RSVP}^{c_2} / c_2!}{\sum_{i=0}^{c_2} A_{RSVP}^{i} / i!},$$

where $A_{RSVP} = \frac{N_0 \lambda_0 (1 - P_{b,0})}{\mu}(1 - P_{b,1}) + N_0 \lambda_{h,0} d_{RSVP}$.

In case of MQSIG, the blocking probability is given as

$$P_b = \frac{A_{MQSIG}^{c_2} / c_2!}{\sum_{i=0}^{c_2} A_{MQSIG}^{i} / i!},$$

where $A_{MQSIG} = \frac{N_0 \lambda_0 (1 - P_{b,0})}{\mu}(1 - P_{b,1}) + N_0 \lambda_{h,0} d_{MQSIG}$.

All parameters in the above equations are listed below:

- A_{RSVP} is the aggregate traffic that the edge router passes to the core router.
- A_{MQSIG} is the aggregate traffic that the edge router passes to the core router.
- P_b is the blocking probability of requested sessions.
- c_i is the number of shared channels at each intermediate router of level *i*.
- N_i is the number of clusters at level *i* (The cells are organized in clusters served by access routers as depicted in Fig. 2).
- λ_i is the arrival rate at level *i*.
- $P_{b,i}$ is the blocking probability at the routers of level *i*.
- $\lambda_{h,i}$ is the generation rate of handoffs at level *i*.
- *d* is the RSVP flow reservation maintenance time without refreshes.
- The call durations are modeled as independent random variables, following the exponential distribution with parameter μ.

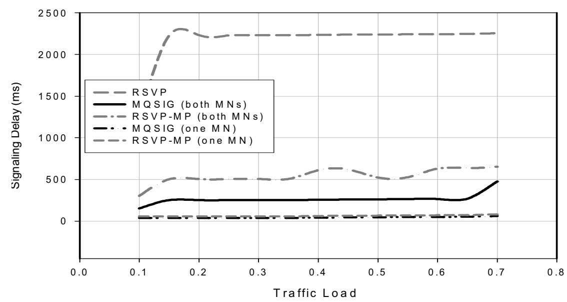


Fig. 3. Signaling delay for re-reservation vs. traffic load

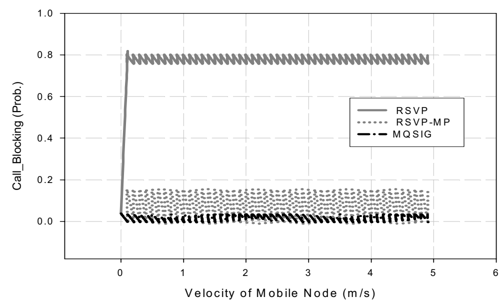


Fig. 4. Blocking probability vs. MN's velocity

In case of RSVP-MP, the blocking probability is similar to the equations above. With MQSIG, d is much shorter compared to RSVP and RSVP-MP. Fig. 4 shows the blocking probability of the requested sessions as the velocity of the MN increases. The values of the key parameters, C_0, C_1, N_0, μ, d_{RSVP}, $d_{RSVP\text{-}MP}$, and $d_{Proposed}$ are 512, 1536, 21, 1/120, 30s, 325ms, 45ms, respectively. The maximum value of i is 2. As shown in the Fig. 4, MQSIG shows better performance compared to RSVP and RSVP-MP. This is because resources on the old path are immediately released after resource re-reservation is completed on the new path and double reservation is avoided through local repair on the common path.

C. *Experimental Results from a Testbed*

To demonstrate that the proposed signaling protocol works well in MIPv6-based mobile environments, we configured a testbed platform which consists of four routers, a mobile node (MN), and a fixed node (CN) as shown in Fig. 5. The proposed signaling protocol is installed at each router and mobile/fixed node. All devices in Fig. 5 are based on Linux OS (Kernel version 2.4.26). Video LAN Client (VLC) and MGEN6 were used to generate video traffic (which needs resource reservation) and best effort (BE) traffic, respectively. For traffic monitoring and measurement, we used Tele Traffic Tapper (TTT) software. We assume that the MN is a data sender, and the CN is not mobile.

In our experiment, the MN is initially attached to AR1 and will move to the new AR, AR2, at a certain time as depicted in Fig. 5. Before the MN moves to AR2, a certain amount of bandwidth is reserved (on the current path) for the VLC flows generated by the MN.

The total amount of bandwidth available to VLC and BE traffic is 4.5 Mbps. As shown in Fig. 6, the VLC flow is using the reserved bandwidth (1.5 Mbps) before handover, while the BE traffic is allowed to use up to 3 Mbps (although MGEN6 generates BE traffic at the rate of more than 3Mbps). At 61.309s, the MN moves to AR2, MIPv6 performs binding update, and the CRN initiates signaling messages to re-setup resource reservation on the new path. Fig. 7 shows that the VLC flow uses the reserved bandwidth immediately after handover while the bandwidth consumption of BE traffic is limited. The low signaling delay was obtained using localized signaling initiated by CRN within the area affected by handover (note that this avoids end-to-end signaling), and the close coupling between MIPv6 binding update and QoS signaling. In this experiment, the measured handover delay is approximately 500ms, and the measured signaling delay for re-reservation is about 100ms on the average.

V. Concluding Remarks

In this paper, we identified some crucial issues including double reservation and end-to-end signaling problems which may occur when QoS signaling interacts with macro-mobility management protocols (e.g., Mobile IPv6). Based the analysis, we proposed a mobility-aware QoS signaling protocol (MQSIG). We also demonstrated that the proposed signaling protocol reduced both the resource re-reservation delay and call blocking probability by fast localized CRN discovery and local repair mechanisms.

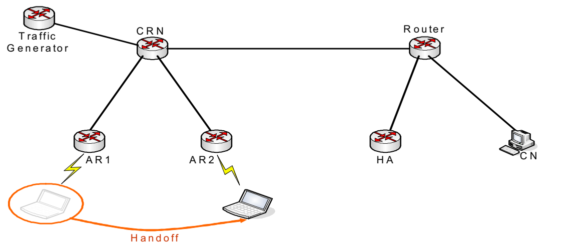


Fig. 5. Experimental testbed configuration

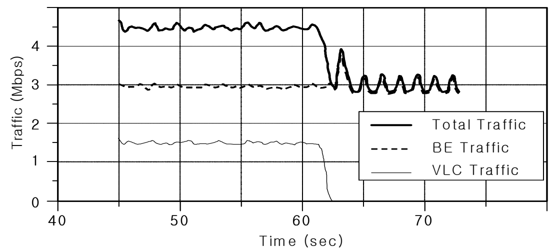


Fig. 6. Reserved bandwidth for VLC traffic before handover

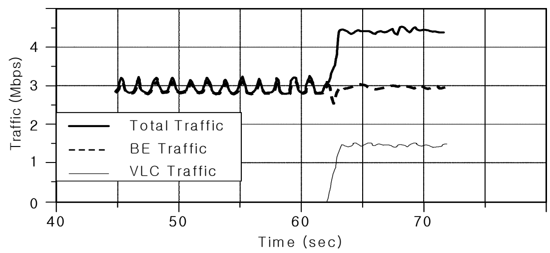


Fig. 7. Reserved bandwidth for VLC traffic after handover

References

[1] B. Busropan et al., "AN Scenarios,Requirements and Draft Concepts," IST-2002-507134-AN/WP1/D02, October 2004.

[2] R. Hancock, "Next Steps in Signaling: Framework," draft-ietf-nsis-fw-04 (work in progress), September 2003.

[3] S. Lee, S. Jeong, H. Tschofenig, X. Fu, J. Manner, "Applicability Statement of NSIS Protocols in Mobile Environments," draft-ietf-nsis-applicability-mobility-signaling-00.txt (work in progress), October 2004.

[4] B. Braden, L. Zhang, S. Berson, S. Herzog, and S. Jamin, "Resource ReSerVation Protocol (RSVP) - Version 1 Functional Specification," RFC 2205, September 1997.

[5] A. Talukdar, B. Badrinath, and A. Acharya, "MRSVP: A Resource Reservation Protocol for an Integrated Services Packet Network with Mobile Hosts," Technical report DCS-TR-337, Rutgers University, 1997.

[6] W. Chen and L. Huang, "RSVP Mobility Support: A Signaling Protocol for Integrated Services Internet with Mobile Hosts," Proc. IEEE Conference on Computer Communications, Vol. 3, pp.1283-1292, 2000.

[7] I. Mahadevan and K. M. Sivalingam, "Architecture and Experimental Results for Quality of Service in Mobile Networks using RSVP and CBQ,"ACM Wireless Networks 6, pp. 221-234, July 2000.

[8] H. Schulzrinne, "GIMPS: General Internet Messaging Protocol for Signaling," draft-ietf-nsis-ntlp-04 (work in progress), October 2004.

[9] S. Van den Bosch, "NSLP for Quality-of-Service Signaling," draft-ietf-nsis-qos-nslp-05 (work in progress), October 2004.

[10] D. Johnson, C. Perkins, J. Arkko, "Mobility Support in IPv6," RFC 3775, June 2004.