Network Working Group
Internet-Draft
Expires: April 16, 2004

X. Fu
Univ. Goettingen
P. Mendes
DoCoMo Euro-Labs
H. Schulzrinne
Columbia Univ.
H. Tschofenig
Siemens

October 17, 2003

Mobility Issues in Next Steps in Signaling (NSIS)
draft-fu-nsis-mobility-01.txt

Status of this Memo

Copyright Notice

Abstract

This document attempts to identify the various problems with signaling in the data path for the mobile node's on-going flows after it moves to a new point of attachment, and analyzes emerging issues with mobility support in the two-layer Next Steps in Signaling (NSIS) architecture.

Table of Contents

1. Introduction

The Next Steps in Signaling (NSIS) working group is chartered with the goal of standardizing a generic IP signaling protocol, having Quality of Service (QoS) signaling as the first use case. A two-layer signaling architecture of NSIS, which consists of a generic transport layer protocol (NTLP) and various application signaling procotols (NSLPs, e.g. QoS signaling [I-D.ietf-nsis-qos-nslp] and NAT/firewall traversal), enables a separation of the transport of the signaling from the application signaling. The NSIS framework [I-D.ietf-nsis-fw] describes the functionality of the individual layers in detail.

The interactions between mobility and signaling protocols have been analyzed in recent years, in the context of RSVP and Mobile IP interaction [I-D.thomas-nsis-rsvp-analysis], [I-D.shen-nsis-rsvp-mobileipv6], [I-D.manner-lrsvp], [paskalis03], [I-D.ietf-nsis-signalling-analysis], in the context of the DiffServ model [heijenk01] and in the context of optimizing QoS signaling for Mobile IP handovers [I-D.westphal-nsis-qos-mobileip].

A previous study about QoS provisioning in mobile IP [manner02] shows that the difficulty with the interaction between mobility and QoS signaling protocols is mainly due to the design of the latter ones. The I-D on requirements for signaling protocols [I-D.ietf-nsis-req] and a recently published RFC on mobile IP QoS requirements [RFC3583] discuss some issues concerning NSIS signaling and QoS signaling in various mobility environment. However, the interaction between mobility protocols and the NSIS protocols has not yet reached a conclusion. The goal of this document is two-fold. First, it aims to identify the problems that mobility causes to NSIS signaling, in particular in the case of QoS signaling. Secondly, it presents an analysis about mobility support in the two-layer NSIS architecture.

In a long term, it may be necessary to study how to use NSIS

signaling in particular network environments, such as 3G and WLAN networks. However, we do not assume a specific mobile access network, since it is important to avoid tightening NSIS signaling with any specific access technology. Moreover, NSIS signaling should be also independent from the used mobility management scheme. Hence, we assume a general scenario where mobile nodes (MNs) can have local mobility (inside the same access network) or global mobility (between different access networks), without making NSIS signaling aware of details specific to different mobility management schemes.

This document attempts to identify problems with signaling in mobility environments, and the implications caused by different design choices for mobility support in NSIS signaling.

2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

In addition, this document frequently uses the following terms or abbreviations most of which are defined in Mobile IP, SEAMOBY and NSIS related documents (e.g., [I-D.ietf-seamoby-mobility-terminology]):

Home Address

Home Agent (HA)

Mobile Node (MN)

Correspondent Node (CN)

Access Router (AR), Previous AR (PAR) and New AR (NAR)

Mobile IP (MIP): Mobile IPv6 (MIPv6) or Mobile IPv4 (MIPv4)

Hierarchical Mobile IPv6 (HMIPv6)

Localized Mobilitiy Management (LMM): Fast/Low Latency Handovers, HMIPv6 or Mobile IPv4 Regional Registration

Seamoby mechanisms: Context Transfer, Candidate Access Router Discovery

Care of Address (CoA)

Regional Care-of-Address (RCoA), On-Link Care-of-Address (LCoA)

Mobility Anchor Point (MAP)

Mobility Agent: e.g., MAP and HA

NSIS Entity (NE)

NSIS Transport-Layer Protocol (NTLP)

NSIS Signaling-Layer Protocol (NSLP)

(Next-)Peer discovery

Authentication, Authorization and Accounting (AAA)

3. Problems with Mobility Support in Signaling

In this section we identify the problems that mobility poses to signaling. The described problems are due to two main characteristics of mobile systems. First, changes caused by re-routing of data packets, essentially posed by topology changes, including the change of host IP addresses and the change of routes for data packets sent to or from the mobile node. Second, the use of IP-in-IP encapsulation in some sections of the end-to-end path.

3.1 Problems Caused by Rerouting of Data Packets

Rerouting of data packets can be due to non-mobility related events inside the network, or due to the movement of end hosts (mobility). The reparation of a data path due to non-mobility issues, such as load balancing and router failure, are not a concern of this I-D.

The movement of end-hosts leads to changes in the data path due to the change of their point of attachment in the network. This results in the original data path between a sender and a receiver to be divided into three paths, all of which intersect at a crossover

router: the unchanged path, the newly-added path and the old ("obsoleted") path.

The attachment of a mobile node (MN) to a new network point normally means a change in the IP address used to build a data path. However, in some situations, the movement of end-hosts can lead to changes in the data path even if the addresses associated with the data flow do not change in the path. This happens with Hierarchical Mobile IPv6 (HMIPv6) [I-D.ietf-mobileip-hmipv6], when a Mobility Anchor Point (MAP) allows an MN to use the Regional Care-of Address (RCoA) as source address without tunneling through the MAP. This means that although the uptream flow have the same source address (RCoA) and destination IP address it can be routed through a different path towards the correspondent node (CN). This section describes a set of problems posed by rerouting due to the movement of end-hosts.

3.1.1 State Management

Due to rerouting of data packets after handovers, signaling-associated states need to be updated or removed. This concerns with which information is needed for indexing states and where a creation, update or removal of these states is required.

In general, a mobile node has a home address as its permanent IP address; after a movement it obtains a new CoA which is the basis for routing its data. If signaling-associated states, which are stored at routers along the path, are indexed based on some temporary data

plane information, such as CoA, the states created by previous CoAs can be inaccessible for the signaling after most handover procedures. Furthermore, updating state information along the entire path might be necessary to reflect the topology change of the MN (as one signaling end).

3.1.2 Local Path Repair

In any mobility approach, a handover causes route changes in some network nodes along the path: for the upstream direction signaling, in the MN and possibly mobility agent(s) (if reverse tunneling is used); for the downstream direction signaling, in the CN, the HA and/or other mobility agents (when non-route-optimization or LMM is used). Therefore, state needs to be installed on the new path and

removed from the old one. We call this "path repair".

The installation of state in the new path must be done as quick as possible so that the mobile node does not experience service interruption or service degradation. To allow a quick set up of state in a new path, the path repair should be done locally, i.e., only between the MN and the crossover router, which is the first signaling-aware router where the old and new path meet. This also avoid state duplication on the unchanged path, i.e., the path from the crossover router to the correspondent node.

Besides the installation of state in the new path, the local path repair mechanism should release state in the old path, as soon as possible, to avoid wasting resources. Typically, the changed path is located inside an access network, where resources are relatively expensive, thus it might be inefficient to wait for typical soft-state timeouts. However, immediately releasing resources along the old path might cause problems. In case of a ping-pong type of movement resources along the old path might need to be reused again after a very short time period. This means that the MN may return to the previous access network shortly after leaving it, which brings some problems about deciding to when to release state in the old path.

The local path repair may be triggered by the mobile node, the mobility agent(s), or by the access router at which the mobile node is attached to. However, the triggering may be constrained by which entities are authorized to carry out what state manipulations, which is then a signaling application and security question.

To install states in the new path and to release states in the old path, at least the following problems have to be solved:

- Know when to trigger the reparation of the local path. This

  triggering requires some interaction with mobility management schemes;

- Know where to end the local repair (i.e., discovery of the crossover router, which requires differentiating between the MN as sender or receiver.

3.1.3 Crossover Router Discovery

The problem of finding the crossover router in a new data path is a generalization of the problem of finding next signaling-aware node. It requires to probe the new data path until the old path is reached. Since we have to assume that the MN can be a sender as well as a receiver, the first difficulty to find a crossover router is posed by the asymmetric characteristic of routing.

Due to routing asymmetry there is no reason for the crossover router to be the same in the upstream direction and in the downstream direction, even for the same correspondent node.

When an MN changes its point of attachment to the network, the discovery of the crossover router is easier for the upstream direction, in which the MN is the sender, than for the downstream direction. In the latter case, although it is the MN that detects data path changes, the discovery of the downstream crossover router has to be done by signaling via the correspondent node.

Hence, the problem of discovering the new crossover router can be divided into the following issues:

- When to trigger the discovery mechanism. This problem is related with the interaction between a signaling protocol and the mobility scheme, to allow the former to be notified from the latter about the movement of mobile devices.

- How to trigger the crossover router discovery mechanism, which depends on the role of the MN as a sender or receiver.

- How to identify a crossover router.

- Since both end-hosts involved in a bi-directional communication can be mobile, the discovery mechanism trigger by each one of them, as sender and receiver, has to be coordinated.

3.1.4 Update the Unchanged Path

As discussed earlier, double reservations in the unchanged path should be avoided. This can only be done by being able to establish a relationship between the old and the new flow. This is essentially the same problem faced to release resource in the old path.

After the identification of the old flow in the unchanged path, the network control state on the unchanged path must be updated to reflect new flow identification. This leads to the problem of requiring end-to-end signaling, which should be avoided to decrease the control load overhead. However, it should be possible to avoid AAA and admission control processing.

3.1.5 Service-aware Signaling

Signaling in mobile environments can be quite dependent on the type of applications. For instance, for QoS signaling, having reservations on old and new paths has a cost and should be avoided. On the other hand, in the presence of ping-pong movement, the immediately release of state in the old path can have a performance impact higher than the cost of keeping that state.

The main problem posed by these issues is the difficult definition of a signaling protocol general to all types of application in mobile environments. Therefore, interaction of signaling and mobility imposes an analysis of signaling responsibilities of each one of the two NSIS layers in mobile environments.

3.2 Problems Caused by IP-in-IP Encapsulation

IP-in-IP encapsulation is one type of IP-in-IP tunnels, which have been mentioned in [RFC2746]. Mobile IP's IP-in-IP encapsulation for data packets introduces a number of problems, described below.

3.2.1 (Re-)Routing of Signaling Messages

One concern is which information can be used for routing of signaling messages. In NSIS, signaling is expected to follow the standard IP routing and mobile IP routing. In a standard IP routing case, messages can be routed according to the destination host's IP address and possibly additional IP header fields. In presence of mobile IP routing protocols, especially when multiple mobility protocols are applied for the same MN-CN communications, possibly nested (e.g., MIPv6+HMIPv6+FMIPv6), it can be difficult to use the MN's address (CoA) as the destination address. Also, as asymmetric routing is common, routing of signaling messages must differentiate between the case of the MN as data sender and the case of the MN as data

receiver.

3.2.2 IP-in-IP Tunnels

If the two end points of a tunnel is not signaling-aware, it might be difficult to provide signaling services for signaling-aware nodes inside the tunnel. In order to, for example, make a QoS reservation for tunnels, it might be necessary to have the tunnel end points to support NSIS. To make things more complicated, mobile IP tunnels co-exist with the change of host addresses (e.g., CoAs) and multiple of them can be possible nested. Therefore, signaling operations over these tunnels, including routing of signaling messages and state management, need to be systematically constructed.

3.2.3 Service-aware Signaling

Essentially, the purpose of signaling is to provide certain services for the MN's flows. Therefore, signaling must be able to install correct packet classifiers in the signaling-aware nodes. Since the tunnels can be physically long, without loss of generality, the ability to signal and install service-aware state inside the (mobile IP) tunnels is required. Such information may need to be delivered to both the tunnel endpoints and internal signaling-aware nodes. This is difficult because signaling messages - same as other data packets - can be encapsulated so as to be invisible of host addresses. In case that nested tunnels are used (e.g., due to the concurrent usage of several tunneling protocols), where an MN can have several level of CoAs at one time together with a home address, special care would be needed.

Another issue with service-aware signaling concerns with selective signaling over tunnels. For example, when an application triggers a per-flow QoS reservation, in some cases it is desired to trigger a QoS reservation also for the tunnel. In some cases it is, however, not desirable to support a QoS reservation for a tunnel. It must therefore be possible for NSIS to decide whether a reservation for a tunnel is desired and at which granularity. An end-to-end QoS reservation need not to be automatically extended to the tunneled region since the tunnel might also serve as a means for aggregation.

3.2.4 Routing Interface

Mobility reflects different route change due to creating, updating or

deleting tunnels, as well as changes in the routing entries, for example in HA, CN, FA/MN, GFA/MAP). It is necessary for the signaling process to obtain such information. RFC 2205 [RFC2205] provides an RSVP/routing interface which can be applicable for mobility/ NSIS-signaling interface. However, it is necessary to coordinate

signaling behaviors based on several mobility route change events in different parts in the network, even when a single mobility routing protocol is used.

3.2.5 Crossover Router Discovery

As previously stated, signaling messages can be hidden for signaling-aware routers inside a tunnel. Therefore, it may be necessary to introduce a separate discovery and signaling message for the tunneled region.

4. Analysis on Mobility in the Two-Layer NSIS Architecture

The problems described in Section 3.1 have several implications to the funtionality of the two-layer NSIS architecture in providing mobility support. Hence, we analyze how mobility can be supported in the two-layer NSIS architecture. Specifically, we focus on the following issues to understand how different design choices can possible address the enumerated problems.

4.1 Identifiers in Data and Control Planes

The main purpose of NSIS signaling is to manage flow-based state in the network nodes. As mentioned in Section 3.1.2, to be able to perform a local path repair, the flows in the new path and the correspondent ones in the old path must be identified. The problem is that although a flow is the same, one or both of the source and destination addresses are different due to the new Care-of Address (CoA) which the MN obtained from the new access network. Hence, A stable identifier within NSIS is required which does not change due to mobility. This identifier helps in the discovery of the crossover router, allows the identification of the same flow in the old and new path, and avoids setting up double state in the unchanged path.

The definition of a stable identifier must consider whether that identifier should be placed in the data plane or in the control plane. The former requires the definition of a flow identification that does not change. This simplifies the problem of session update, but introduces some problems on packet classification and security. No such mobility independent identifier is available although the

IPv6 Flow Label [I-D.ietf-ipv6-flow-label] based on the Home Address was discussed on the mailing list. If the flow identifier is based on the Home Address, for example, once an MN has more than one signaling flow to the same CN, say, one for the purpose of QoS and another for firewall pinhole, it becomes difficult to distinguish between them.

In general, a distinguish between data plane and control plane identifiers is necessary: the former of which reflects real, routable IP addresses (home address, CoA or other possible tunnel endpoints' addresses) in order to reflect packet classification, while the latter reflects a stable identifier for managing state information. In other words, the following two identifiers are useful to support mobility in NSIS:

- A session identifier (session-ID) in the control plane that must not change during mobility, in order to manage the control state. This session-ID can be a random identifier or cryptographically generated. It should be "globally unique". [I-D.tschofenig-nsis-sid] discusses some possible mechanisms for creating such an identifier.

- A flow identifier (flow-ID) in the data plane that describing packets belonging to the flow. The flow-ID should contain fields that are used by NSIS-aware routers to install filter specs (packet classification). Some possible way of composing flow-IDs have been suggested in the NSIS framework document [I-D.ietf-nsis-fw]: 5-tuple, IPv6 flow label or a 3-tuple. Since in mobile IP one or more IP-in-IP tunnels is(are) inserted in the data path (see discussions in Section 3.2), an end-to-end universe flow-ID might be insufficient to route signaling messages, or even to make filter specs for applications' flows. It might be not necessary to have the flow-ID in the NSIS message (and the associated packet classifier/state stored at NSIS nodes) along the entire end-to-end path to be the same; some network nodes (e.g., HA) might need to be able to change it.

Note it can also possible that the NSLP has its own session-ID to identify individual sessions. They may or may not be correlated. In the following text, without special statement, session-ID represents NTLP session-ID.

The use of a flow-ID based on MN's addresses, e.g., the CoA, and possibly other fields in the IP header is meaningful for many signaling applications. However, whether flow-ID would be needed for all NSLPs is still an open issue. Therefore, the flow-ID could be constructed either by NTLP or by the NSLP layer. While constructing flow-ID, if CoAs are used for identifing flows, even on the unchanged path, signaling would be still needed to update a changed flow identification in the unchanged path. In this case, signaling in the unchanged path should be possible to avoid AAA and admission control processing. Note if flow-ID is managed by the NTLP layer, it can be avoided to be constraint to host addresses, thus make it possible to limit the update of flow-ID information locally, while still allowing flow-ID useful/visible for NSLPs. If the flow-ID contains a CoA, it is necessary to update flow-IDs stored at NSIS nodes along the the unchanged path. Furthermore, from a performance point of view it would be highly advisable to avoid AAA and admission control processing.

The management of the session-ID and flow-ID in the two-layer NSIS architecture requires more study, especially after an analysis of the NTLP proposal. In particular, routing of signaling messages at NTLP level can be based on either flow-ID or session-ID. In a flow-ID-based approach, NTLP has to rely on a mapping between certain fields of the flow-ID (e.g., destination IP address and additional IP header information) and local IP (and mobility) routing table. In a session-ID-based approach, NTLP can route the signaling messages

based on a mapping between the session-ID and the local NTLP-level state (for example, next/previous NE's address). If there is no existing state, a next-peer discovery has to be performed to create such a state.

As the association of different flow-IDs to a single session-ID is a problem common to many signaling applications, the association between both identifiers might be done at the NTLP. However, it could be also possible this association is done at the NSLP layer, if the method used to perform such association is specific to each application. In both possibilities, it is assumed that the session-ID should be visible within the NTLP, allowing it to perform an enhanced forwarding control for packets belonging to that session.

Another three related identifiers, namely the message identifier

(message-ID) that is introduced in RFC 2961 [RFC2961], the branch identifier (branch-ID) suggested in CASP [I-D.schulzrinne-nsis-casp][fu03] and the Reservation Sequence Number (RSN) proposed in QoS NSLP [I-D.ietf-nsis-qos-nslp], have been also discussed as potential mechanisms useful for mobility support in NSIS. All of three indicate the order in which corresponding signaling messages are processed by the corresponding signaling entities (RSVP, CASP-NTLP and QoS-NSLP daemons, respectively) and try to address the out-of-order problems of signaling messages.

Message-ID, together with Epoch object in RFC 2961, concerns with signaling messages between peering neighbors, where the out-of-order problem can come from retransmission/refresh. It was not designed for mobility support specifically. As an extension to message-ID concept, the branch ID (together Session ID) can be used for detecting out-of-order signaling messages along different branches each can consist of multiple hops) for route change and mobility scenarios; it can be generally useful for determining end of (explicit) teardown message to forward on along the unchanged path. Different from the branch ID, the RSN is meaningful in a QoS-NSLP node for protecting out-of-order problems in the associated branches, each of which can consist of multiple QoS NEs.

4.2 Crossover Router Discovery

The discovery of the crossover router, due to the mobility of end-hosts, must be based on the session-ID of the MN. Additional information might be used for computing an authorization decision or to verify ownership.

Since the session-ID is processed at the NTLP layer, we can say that the crossover router of a mobile session is an NTLP-aware node that for a session request coming from one of its interfaces, already has

forwarding information about that session stored on one or more interfaces, excluding the incoming one. However, a session may have a different crossover router for the upstream and downstream directions. Hence, we can specify the definition of a crossover router as follows:

- Upstream direction: an NTLP-aware NE that for a session request coming from one of its interfaces, already has information about

that session stored on more than one interface, excluding the incoming one.

- Downstream direction: an NTLP-aware NE that for a session request coming from one of its interfaces, already has information about that session stored on more than one interface, excluding the outgoing one.

The discover of the downstream crossover router must be started by the correspondent node, which means that the MN should signal the correspondent node when it detects some movement. However, this procedure does not necessarily increase the handover latency, since end to end message exchanges will be required at the application layer anyway to update the sender with the new address of the receiver.

In either situation, the discovery of the crossover router is typically done at the NTLP layer. For instance, to discover the crossover router in the upstream direction, all NEs in the new path have to be found out. A natural way is to use NTLP next peer discovery for this purpose and then one can do state repair immediately after finding out the next NE peer, or perform the state repair after finding the new NE-chain in the new path (as discussed in Section 4.3).

Alternatively, it is possible to discover the crossover router at the NSLP layer, since the NSLP is the entity who does the real work for signaling sessions and leaves useful information in NEs. However, this is problematic since the number of NSLPs for a given MN-CN pair may be more than one, which requires discovery of possibly different crossover routers for each NSLP. Furthermore, as NSIS is assumed to follow normal IP routing mechanism, all NSLP sessions between a same MN-CN pair will have the same crossover router. Therefore, definition and discovery of the crossover router at NSLP layer would only increase the complexity.

The discovery of the crossover router due to route changes inside the network caused by non-mobility reasons, such as load balancing and node failure should also be taken care by NTLP. In this case, if end-to-end NTLP addressing is used the discovery of the crossover

router can be done based on the flow-ID, which does not change;

otherwise, session-ID will assist the discovery. The analysis of this situation is beyond the scope of this document, since it is not an issue caused by IP mobility.

4.3 Local Path Repair

There is a tight relationship between the crossover router discovery mechanism and the local path repair mechanism, because the local path is defined as the path between the MN and the crossover router. Local repair can be done either during (coupled approach) or after (separated approach) the crossover router discovery process. This affects whether only NSLP or both NSLP and NTLP is involved in the local repair. In both cases, all NEs in that path have to be discovered by the NTLP peer discovery mechanism. In the coupled approach, NSLP state in a new local path has to be recovered upon a notification of local NTLP (more details are discussed in Section 4.7); the advantage is that it requires less recovery time and the same procedure as normal NSIS signaling. In contrast, the separated approach utilizes an additional NSLP-based procedure which might add the complexity to the whole NSIS signaling; NTLP essentially is used as a transparent underlying mechanism to transport NSLP messages.

The interaction between the crossover router discover mechanism and the local path repair mechanism is different for the upstream and downstream directions:

- Upstream direction: Since the definition of local path depends upon the location of the crossover router, in the separated approach, it may be considered that is the crossover router discovery mechanism that triggers the local path repair, which itself is triggered by the mobile node. In the coupled approach, the crossover router discovery mechanism takes place before the local path repair is performed, until the crossover router is discovered.

- Downstream direction: Since the crossover router of downstream direction can be located in any section in the data path (e.g., between CN and HA, or between HA and CN) a general assumption would be that the crossover router discovery mechanism has to started by the correspondent node (for example, after receiving a binding update message or detecting of a change in its binding entry, see discussions in Section 4.6). However, if LMM mechanisms are used for mobility, it can be also possible that mobility agents such as HA or MAP to start this process. Also, it can be considered that it is the crossover router that starts the local path repair mechanism. A local repair mechanism as the one used by RSVP [RFC2205], where local repair may be triggered by a local

route change, is impossible or at least difficult in this case. The reason is that, the node experiencing a route change can only be the CN, HA, or other mobility agents, which is not necessarily the crossover router and it is the destination address that changed and not the route to the old CoA.

During the local path repair procedure, setting state in a new path may be conditioned by the session ownership and the availability of resources. In the latter case, when the network is overloaded, it is preferable to keep state belonging to previously established flows while blocking new requests. Therefore, the local path repair mechanism for mobiling sessions should have priority over local requests for resources.

4.4 Routing of Signaling Messages

As stated in the NSIS framework document [I-D.ietf-nsis-fw], there are two ways to address a signaling message being transmitted between NEs:

- Peer-to-peer, where the message is addressed to a neighboring NE that is known to be closer to the destination NE.
- End-to-end, where the message is addressed to the flow destination directly, and intercepted by an intervening NE.

Each type of message is necessary for some aspects of NTLP operation: in particular, initial discovery of the next peer will usually require end-to-end addressing, whereas reverse routing will always require peer-peer addressing. For other message types, the choice is a matter of protocol design. The mode used is not visible to the NSLP, and the information needed in each case is available from the flow-ID or locally stored as NTLP state.

With peer-to-peer addressing, an NE uses the payload of the message to determine the address of the next NE. This requires the address of the destination NE to be derivable from the information present in the payload. Peer-peer addressing inherently supports tunneling of messages between NEs.

In the case of end-to-end addressing, the message is addressed to the data flow receiver, and some of the NEs along the data path intercept the messages. The routing of the messages should follow exactly the same path as the associated data flow. To allow signaling packets follow the same route as data packets, network nodes that route packets based on different information from flow-ID must be

NSIS-aware. An example is the forwarding of signaling massages into an IP tunnel. In this case the NTLP may need to force a signaling

packet to use an output interface that it knows data packets are going to use, even if the IP stack would naturally use a different one. The session-ID, which may be visible within the NTLP, as stated in the framework draft, can be used to identify sessions that should be tunneled instead of routed based on the flow-ID. In this situation, the NE at the end of the tunnel should restart routing messages for this session by using the flow-ID.

4.5 IP-in-IP Encapsulation

RFC 2746 [RFC2746] provides an approach for RSVP operation over IP-in-IP tunnels. Basically, the same considerations can be also applicable to NSIS operation over mobile IP IP-in-IP encapsulations.

Two methods of dealing with this issue are conceivable. One is to adapt signaling payloads which refer to header fields to allow signaling inside the tunnel. Another is to use an adaptable flow-ID to indicate the changed situation of the signaling message. Both can be used together, with certain extensions to RFC 2746 [RFC2746]. However, since NTLP itself is not yet fully determined, the issue of operations over mobile IP's IP-in-IP encapsulated paths needs further study.

4.6 Interaction between Mobility Routing and NSIS Signaling

One issue related to routing NSIS messages is the interface between the routing protocol and NTLP/NSLP. In normal situations, end-to-end and peer-to-peer addressing can be handle by the NTLP. In the presence of tunnels, the use of the session-ID by the NTLP allows the forwarding of packet through a tunnel. In this situation, only the NTLP need to have an interface to the routing protocol.

Another issue related with the interaction between mobility routing and NSIS signaling is whether to use the receipt of binding messages (e.g., in CN, MN or HA) (active way) or to use routing/NSIS interface (passive way), to trigger NSIS signaling. The active way allows faster state recovery and removal, but its disadvantage is that fast or ping-pong movements may result in considerable signaling overhead and possible errors, and moreover, by mobility protocol binding

updates can take place periodically even for the MN with the same point of attachment. The passive way, typically after seconds of routing change detection (the MN and the CN in typical cases) of a routing/NSIS interface mechanism to obtain route change and tunnel change information, can be less processing-intensive and thus more promising.

If the session-ID is not visible by the NTLP, two alternatives are conceivable: either other methods need to be used at the NTLP level

to identify sessions that need to be tunneled, or the forwarding of messages may be done by the NSLP. The latter case required an interface between the NSLP and the routing protocol, which may break/ complicate the NSIS layering.

4.7 Interaction between NTLP and NSLP Signaling

In the two-layer architecture, there is a separation between NTLP and NSLPs. In general, NTLP is needed to be involved with mobility anyway, for example to transport signaling messages along the new path in order to create new state, or to transport explicit release signaling messages along old path to release old state. In such cases, NTLP should be able to notify NSLP to update state (by initializing NSLP refresh/teardown messages appropriately). An open issue is, however, how and what information the NSLP can expect from NTLP, or directly from the routing interface.

5. Open Issues

This section discusses some open issues for mobility support in NSIS.

5.1 Both End-Hosts are Mobile

Considerations about signaling between two MNs. Until now, we are assuming a non-mobile corresponding node. Problems can show up if both devices start to signal at the same time, namely in the local path repair, and signaling of NSIS messages.

5.2 Interact with Seamoby Protocols

A term "seamless mobility" is often referred to mean that the MN is able to keep an ongoing session seamlessly (without experiencing perceivable service interruption or performance penalty) during and after moving from one access network to another. Measures to achieve seamless mobility include, but not limited to, various LMM mechanisms, seamoby context transfer [I-D.ietf-seamoby-ctp] and candidate access router discovery [I-D.ietf-seamoby-card-protocol] protocols, as well as predictive/anticipated handling mechanism. Issues related to signaling in LMM scenarios have been discussed in previous sections, while interaction of NSIS with seamoby protocols are different because of their effects on parts that are out of the signaling path. [I-D.westphal-nsis-qos-mobileip] studied the issue

of QoS signaling for mobile IP with context transfer, but the interaction between NSIS and context transfer still needs further investigation.

In the context of mobility between different access routers, it is common to consider performance optimizations in two areas: selection of the optimal access router to handover to, and transfer of state information between the access routers to avoid having to regenerate it in the new access router after handover. Since these solutions are still under development as well as the NTLP, it is premature to specify details on the interaction.

5.3 Fast State Installation/Advanced Reservations

There are some other issues concerned with performance optimization, for example, the capability to (pre-)discover the required level of end-to-end resources, and to fast (predictive/anticipated/in-advance) state installation. This section discusses fast state installation.

Since the time required to establish the session in the new path can induce packet losses and delays, which do not contribute to achieve a seamless handover. Therefore, it is important that the NSIS signaling in a new path can be carried out very quickly.

To accomplish a fast state installation, it might be desirable for NSIS to be performed in advance to the movement of MNs. This approach may involve some kind of NSIS proxy, on the new access network, which can signal in the new path on behalf of the MN. If we assume that the end-to-edge communication is done between the MN and its access router, some study is required to determine how to signal between the mobile device currently access router and the NSIS proxy in the new access network - e.g., how to discover the most suitable NSIS proxy, and to establish a communication between access networks. The latter issue is beyond the charter of NSIS, since it involves out-of-path signaling.

Moreover, in some anticipated signaling scenarios, NSIS signaling cannot be triggered by the mobility signaling, which required some study about other possible triggers, such as:

- Cross-layer triggering. For instance, the layer-2 mechanism can give some information about a possible movement.

- Context-awareness triggering. For instance, information about a lower traffic load in some neighbor access networks can trigger the establishment of state in a new path.

5.4 Resource Discovery in an End-to-End Path

The anticipated signaling for a new path and the selection of a suitable access network may not be enough to ensure seamless mobility. Resource availability in a new end-to-end path may be considered as well. This means that a multimedia session can be disrupted if NSIS cannot signal the required resources in the end-to-end path supplied by the routing protocols.

To avoid quality degradation due to mobility, it might be desirable to interact with QoS routing [RFC2386] to discover the optimal (or sub-optimal) end-to-end path. As stated in the framework document [I-D.ietf-nsis-fw], NTLP should work with standard layer 3 routing, thus QoS NSLP needs to be able to handle this type of enhanced routing. Also, in this case, it might be NSLP's responsibility to discover crossover router.

However, NSIS performance should not be deteriorated in the presence of such protocols. For instance, state oscillation can occur if flows are frequently re-routed due to changes in the traffic load of alternative end-to-end paths. Such state oscillation must be avoided.

5.5 Security and AAA Issues

Since a network access authentication protocol can be executed when a host arrives at a new network, AAA procedures are likely to create the necessary financial settlement. In this sense it is helpful to use an identify for the user session which can be mapped to the identify used during the network access procedures to make authorization and charging easier. This is particularly of relevance if the NSLP carries QoS information. For other NSLPs the authorization procedure may be different, but the identity used in the authentication and key exchange procedure (e.g., IKE, IKEv2 or KINK) has to be accessible especially for an entity in the network at the NSLP layer. (Note that this is simpler in case of TLS where the

authenticated identity of the user is available to the NSLP via an API).

The authorization procedure in a mobile environment still needs more investigation, particularly since there is a strong relationship to network access procedures.

5.6 Other Issues

Aggregation (as well as multicast) is not a mobility-specific problem in general. In mobility scenarios, one possibly related issue is that whether the messages for creating, updating and removing existing states in different sections in the network need to be grouped according to the new CoAs, or the home address, when the MN is the data receiver.

Another issue is signaling in Network Mobility (NEMO) or ad-hoc network scenarios, which has different implications from IP mobility scenarios. In NEMO or ad-hoc networks an end-host's address can remain the same while the path is changed, so state management, crossover router discovery and local repair would rely on the change of path not on the change of host addresses. However, as NSIS signaling is assumed to follow normal IP routing (including IP mobility routing), further study of these issues is beyond the scope of this document.

6. Security Considerations

It is obvious that mobility support within NSIS raises security issues. A number of mobility scenarios with impacts on security are

discussed in Section 7 of [I-D.tschofenig-nsis-aaa-issues]. Even if the signaling message exchange is restarted from scratch (i.e. using a new flow-ID), security handling within NSIS is affected. This type of processing is, however, mostly not a topic for this draft.

The introduction of a flow-ID independent identifier referred as session identifier creates security hazards which are discussed in [I-D.tschofenig-nsis-sid]. Once the expected signaling messaging behavior is precisely defined then it is possible to address the raised security concerns. Then it should be possible to define which entities are authorized to perform certain types of actions. In the following subsections we discuss some additional security implications.

6.1 Missing Cost Control

A large number of service providers (e.g. wireless LAN hotspots) with complex roaming agreements create a non-transparent cost structure. In a traditional subscription-based scenario, users are subscribed to their home network and use this trust relationship to dynamically establish a financial settlement between the visited network and the home network. Additionally, security associations are dynamically established as part of this procedure. This is the typical AAA deployment scenario. In these scenarios users do not learn the identity of the access network as part of a regular authentication and key exchange protocol message exchange. The identity of the access network is possibly never revealed (in a secure fashion). The user is therefore only authenticated to the home network (and hopefully vice versa). While issuing a QoS reservation request to the next NSIS peer (for example in the access network), the end host is typically unaware of the cost of such an end-to-end QoS reservation. Without knowing the costs it is not possible to reject a too expensive QoS reservation.

Currently there is no standarized protocol available which allows users to communicate cost limits, to request cost information for network resources or to learn already accumulated costs for a particular reservation.

Especially in mobility environments - where an end host is likely to have access to different networks within a short time period - cost control is even more complicated.

Some mobility/QoS protocol proposals try to merge existing mobility

protocols with QoS signaling (i.e. to apply in-band signaling). Thereby the access network is queried (towards the crossover router or the MAP) for the possibility of making a QoS reservation (without actually making the reservation itself). Without a query mechanism, a user cannot take reservation costs into account when choosing between different access networks (or different access routers). Hence, the user might be able to refuse a more expensive service provider. The ability to allow a user to choose between different providers might be required - not only because of the availability of different access technologies (e.g. IEEE 802.1x, Bluetooth, UTRAN) and different service quality offered, but also for cost reasons.

Although real-time notifications of QoS reservation costs (cost control) to the user are out of the scope of NSIS, some interaction might be required.

6.2 Implications for Price Determination

The problem of determining the price of a QoS reservation has been mentioned in [I-D.tschofenig-nsis-aaa-issues] and closely relates to integrating the end host into the process of authorization. Even if the end host is aware of the price of a QoS reservation during reservation setup the price might change for a number of reasons:

- First, mobility in general causes a different path to be chosen and might therefore require a new price determination. End host mobility is visible to the end host itself, therefore the appropriate actions can be triggered by the end host to always determine the correct price.

- Route changes somewhere along the path, e.g., mobility in NEMO networks or even mobility in ad-hoc networks, create more problems, since the route change might not be visible for the end host. If price determination is based on the number of networks traversed and intermediate nodes or network contribute to the total price of a QoS reservation, then a periodic price query is necessary. Discussions at the NEMO mailing list already point to this problem [Nemo-ML]. If the price of QoS reservation is associated with the authorization itself, then a periodic re-authorization based on the change of prices or on the accumulated costs is necessary.

6.3 Intra-domain Mobility

Intra-domain mobility with the help of the context transfer protocol can help to move established state information between different

access nodes within the same administrative domain including security

associations, QoS parameters (QoS NSLP state), NTLP state and even authorization information. An authorization for a QoS reservation granted along one path through the access network might also valid at a different access router or even at a different path within the same administrative domain. Discussions in the EAP working group, however, reveal that this might not always be the case. However, if we extend the scheme from intra-domain context transfer to inter-domain context transfer then we might encounter some interesting authorization problems. Note that these issues do not only address authorization of QoS resources, but are more applicable to network access authentication and authorization in general. Network access authentication and authorization would not necessarily be executed again after attaching to the new domain. Instead, a trust relationship is established between the new and the old administrative domain.

In addition to the above issues, performance is important in mobility environments. Proper security handling accounts for a high percentage of the total performance.

7. Acknowledgment

The authors would like to acknowledge discussions with Bob Braden, Marcus Brunner, Ruediger Geib, Seong Jeong, Zhigang Kan, Cornelia Kappler, Holger Karl, Gary Kenward, Sung-Hyuck Lee, Jukka Manner, Andrew McDonald, Sarantis Paskali, Cedric Westphal, and Hairong Zheng in the NSIS working group. In particular, the document benefits from Hemant Chaskar, Robert Hancock, Charles Shen and Michael Thomas's insightful comments on various aspects of this topic.

References

[I-D.ietf-ipv6-flow-label]
Rajahalme, J., Conta, A., Carpenter, B. and S. Deering, "IPv6 Flow Label Specification", draft-ietf-ipv6-flow-label-07 (work in progress), April 2003.

[I-D.ietf-mobileip-fast-mipv6]
Koodli, R., "Fast Handovers for Mobile IPv6", draft-ietf-mobileip-fast-mipv6-08 (work in progress), October 2003.

[I-D.ietf-mobileip-hmipv6]
Soliman, H., Castelluccia, C., Malki, K. and L. Bellier, "Hierarchical Mobile IPv6 mobility management (HMIPv6)", draft-ietf-mobileip-hmipv6-08 (work in progress), July 2003.

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

[I-D.ietf-nsis-qos-nslp]
Van den Bosch, S. and et al., "NSLP for Quality-of-Service

signaling", draft-ietf-nsis-qos-nslp-00 (work in progress), September 2003.

[I-D.ietf-nsis-req]
Brunner, M., "Requirements for Signaling Protocols", draft-ietf-nsis-req-09 (work in progress), August 2003.

[I-D.ietf-nsis-signalling-analysis]
Manner, J., Fu, X. and P. Pan, "Analysis of Existing Quality of Service Signaling Protocols", draft-ietf-nsis-signalling-analysis-02 (work in progress), July 2003.

[I-D.ietf-seamoby-card-protocol]
Liebsch, M. and et al., "Candidate Access Router Discovery", draft-ietf-seamoby-card-protocol-04 (work in progress), September 2003.

[I-D.ietf-seamoby-ctp]
Loughney, J. and et al., "Context Transfer Protocol", draft-ietf-seamoby-ctp-04 (work in progress), October 2003.

[I-D.ietf-seamoby-mobility-terminology]
Manner, J. and M. Kojo, "Mobility Related Terminology", draft-ietf-seamoby-mobility-terminology-04 (work in progress), April 2003.

[I-D.manner-lrsvp]
Manner, J. and et al., "Localized RSVP", draft-manner-lrsvp-02 (work in progress), July 2003.

[I-D.schulzrinne-nsis-casp]
Schulzrinne, H. and et al., "CASP - Cross-Application Signaling Protocol", draft-schulzrinne-nsis-casp-01 (work in progress), March 2003.

[I-D.shen-nsis-mobility-fw]
Shen, C., "Several Framework Issues Regarding NSIS and Mobility", draft-shen-nsis-mobility-fw-00 (work in progress), July 2002.

[I-D.shen-nsis-rsvp-mobileipv6]
Shen, C. and et al., "Mobility Extensions to RSVP in an RSVP-Mobile IPv6 Framework", draft-shen-nsis-rsvp-mobileipv6-00 (work in progress), July 2002.

[I-D.thomas-nsis-rsvp-analysis]
Thomas, M., "Analysis of Mobile IP and RSVP Interactions", draft-thomas-nsis-rsvp-analysis-00 (work in progress), November 2002.

[I-D.tschofenig-nsis-aaa-issues]
Tschofenig, H., "NSIS Authentication, Authorization and Accounting Issues", draft-tschofenig-nsis-aaa-issues-01 (work in progress), March 2003.

[I-D.tschofenig-nsis-sid]
Tschofenig, H. and et al., "Security Implications of the Session Identifier", draft-tschofenig-nsis-sid-00 (work in progress), June 2003.

[I-D.westphal-nsis-qos-mobileip]
Westphal, C. and H. Chaskar, "QoS Signaling Framework for Mobile IP", draft-westphal-nsis-qos-mobileip-00 (work in progress), June 2002.

[Nemo-ML] Alper, Y., "[nemo] AAA and NEMO", discussion in the IETF Nemo Mailing List (available at: http://www.nal.motlabs.com/pipermail/nemo/2003-February/

000271.html), February 2003.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC2205] Braden, R., Zhang, L., Berson, S., Herzog, S. and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, Sep 1997.

[RFC2386] Crawley, E., Nair, R., Rajagopalan, B. and H. Sandick, "A Framework for QoS-based Routing in the Internet", RFC 2386, August 1998.

[RFC2746] Terzis, A., Wroclawski, J. and L. Zhang, "RSVP Operation Over IP Tunnels", RFC 2746, January 2000.

[RFC2961] Berger, L., Gan, D., Swallow, G., Pan, P., Tommasi, F. and S. Molendini, "RSVP Refresh Overhead Reduction Extensions", RFC 2961, April 2001.

[RFC3583] Chaskar, H., "Requirements of a Quality of Service (QoS) Solution for Mobile IP", RFC 3583, September 2003.

[fu03] Fu, X., Schulzrinne, H. and H. Tschofenig, "Mobility Support for Next-Generation Internet Signaling Protocols", Proceedings of IEEE Vehicular Technology Conference 2003-Fall, October 2003.

[heijenk01] Heijenk, G., Karagiannis, G., Rexhepi, V. and L. Westberg, "DiffServ Resource Management in IP-based Radio Access Networks", Proceedings of 4th International Symposium on Wireless Personal Multimedia Communications (WPMC'01), Aalborg, Denmark, September 2001.

[manner02] Manner, J., Lopez, A., Mihailovic, A., Velayos, H., Hepworth, E. and Y. Khouaja, "Evaluation of mobility and QoS interaction", Computer Networks vol.38, no.2, pp.137-163, February 2002.

[paskalis03] Paskalis, S., Kaloxylos, A., Zervas, E. and L. Merakos, "An efficient RSVP-mobile IP interworking scheme", Mobile Networks and Applications vol.8, no.3, pp.197-207, June 2003.

Authors' Addresses

Xiaoming Fu
University of Goettingen
Telematics Group

Lotzestr. 16-18
Goettingen 37083
Germany

EMail: fu@cs.uni-goettingen.de

Paulo Mendes
DoCoMo Communications Laboratories Europe GmbH
Landsberger Str. 312
Munich 80687
Germany

EMail: mendes@docomolab-euro.com

Henning Schulzrinne
Columbia University
Dept. of Computer Science
1214 Amsterdam Avenue
New York, NY 10027
USA

EMail: hgs@cs.columbia.edu

Hannes Tschofenig
Siemens AG
Otto-Hahn-Ring 6
Munich 81739
Germany

EMail: Hannes.Tschofenig@siemens.com

Intellectual Property Statement

Full Copyright Statement

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.