Securing the Next Steps In Signalling (NSIS) protocol suite

Hannes Tschofenig*
Siemens AG, Corporate Technology, Otto-Hahn-Ring 6, Munich 81739, Germany
Fax: +49 89 636 48000 E-mail: hannes.tschofenig@siemens.com
*Corresponding author

Xiaoming Fu
Institute for Informatics, University of Goettingen, Lotzestr. 16-18, Goettingen 37083, Germany
Fax: +49 551 39 14403 E-mail: fu@cs.uni-goettingen.de

Abstract: The Next Steps In Signalling (NSIS) protocol suite represents an extensible framework for enabling various signalling applications over IP-based networks. The framework consists of two layers that need different types of security protection; the lower layer mainly deals with the discovery of adjacent peers and establishment of channel security to protect the delivery of signalling messages between two peers, while the upper layer provides the signalling application specific functionalities. Different security properties are required at the two layers with stronger authorisation functionality at the signalling application layer. In this paper we examine how various security vulnerabilities can be utilised by an adversary, including eavesdropping, Man-In-The-Middle (MITM) attacks, fraud and Denial of Service (DoS) attacks. Moreover, we describe how to protect against a number of selected security threats and highlight some security challenges that require further research.

Keywords: QoS signalling; RSVP; Next Steps In Signalling (NSIS); General Internet Signalling Transport (GIST); security; AAA.


Biographical notes: Hannes Tschofenig received his Diploma in Computer Science from the University of Klagenfurt, Austria in 2001. Since then he has been a Research Scientist at Corporate Technology, Siemens AG, Germany. His primary research interests lie in network security, with a focus on mobile communications and QoS contexts. He is Secretary of the NSIS working group and Chair of ECRIT working group of the IETF. He is a coauthor of a number of RFCs including RFC 3726 ‘Requirements for Signalling Protocols’, RFC 4081 ‘Security Threats for Next Steps In Signalling (NSIS)’ and an Editor of ‘RSVP Security Properties’, ‘NAT/Firewall NSIS Signalling Layer Protocol (NSLP)’, ‘Pre-Shared Key Ciphersuites for Transport Layer Security (TLS)’, ‘Design of the MOBIKE Protocol’, ‘Protocol for Carrying Authentication for Network Access (PANA)’, and ‘Carrying Location Objects in RADIUS’, among others.

Xiaoming Fu received his PhD in Computer Science from Tsinghua University, China in 2000. He joined the Institute for Informatics, University of Goettingen, Germany as an Assistant Professor (‘wissenschaftlicher Assistent’) in 2002. Prior to that he was a member of the research staff at the Technical University Berlin, Germany. He is a co-author of RFC 4094 ‘Analysis of Existing Quality-of-Service Signalling Protocols’ and about 40 refereed papers. His research interests encompass network architectures, protocols and distributed systems design, implementation, verification and performance analysis, particularly on signalling and QoS, security, overlay networks, wireless and mobile networking. In these areas he currently leads several EU-sponsored and other research collaboration projects and actively contributes to the internet community.
1 Introduction

The internet is continuously evolving from a traditional packet-switching network to a network with a number of architectural changes; for example, due to the rise of QoS services (Braden et al., 1994; Blake et al., 1998) and stateful middleboxes (Kempf and Austein, 2004). Subsequently, there has been a general need to employ signalling protocols to install, maintain and remove control states in network nodes which are related to end-to-end communications. Several proposals have been submitted to the IETF to provide such functionality, one of the pioneering works of which is the Resource Reservation Protocol (RSVP) (Braden et al., 1997).

The strengths of RSVP are that it provides a soft-state management mechanism, a modular design to accommodate QoS reservation state setup among multicast and unicast networks, and a separation between signalling and routing (Zhang et al., 1993). However, RSVP has several key weaknesses; for example, in its message delivery mechanism, protocol complexity, inadequate support in extensibility (to support other signalling applications than just QoS signalling), and inability to support mobility (Manner and Fu, 2005). Furthermore, as pointed out in Wu et al. (1999) and in Section 3 of this paper, when QoS (or any other stateful middlebox) service provisioning is concerned, the control plane (signalling) and data plane functions can suffer from various attacks, e.g., DoS attacks on reservation setup or consuming legitimately reserved resources by injecting false packets. Here, security of control plane state installation, maintenance and removal, namely the operations of a signalling protocol, is critical for the success deployment of such new services and applications. However, security was only patched to RSVP later (Baker et al., 2000; Herzog, 2000; Yadav et al., 2001) and RSVP lacks a solid security framework especially for end-to-end addressed signalling messages (e.g., PATH) (Wu et al., 1999; Talwar and Nahrstedt, 2000; Tschofenig and Graveman, 2004).

In particular, discovery and signalling message delivery are combined into a single protocol step in RSVP. This design decision makes it difficult to provide proper security protection using existing security protocols, as well as to extend the protocol into other deployment environments. In addition, authentication and key management are not adequately addressed. For example, only manual configuration is supported for the Integrity object (Baker et al., 2000), which protects the entire RSVP signalling message. Authorisation aspects are provided to some degree (see discussions in Section 3.6) but do not interwork with today’s AAA infrastructure (such as Diameter (Calhoun et al., 2003) or RADIUS (Rigney et al., 2000)) and roaming environments.

The NSIS working group of the IETF is working on a signalling protocol suite aiming to overcome the shortcomings of RSVP while learning as much as possible from its experiences. Realising today’s internet has become a hostile environment, and security functions in general contribute a considerable amount of complexity and performance impacts on the design of a protocol; security aspects have been analysed in the NSIS design from the early stages of the protocol work (see, for example, Tschofenig and Kroeselberg, 2005). We believe that addressing security will manifest as an important prerequisite for NSIS’s success. However, to the best of our knowledge, there is no single paper systematically summarising these aspects and investigating the possible overall security design space for the NSIS protocol suite.

In this paper we describe NSIS communication models and analyse various threats for the NSIS protocol suite. In addition, we describe how to deal with a selected set of security threats for NSIS protocols, especially in some specific application domains.

2 The NSIS protocol suite: preliminaries

2.1 NSIS overview

The NSIS protocol suite is designed to comprise two functional layers (Hancock et al., 2005). The lower layer (the NSIS Signalling Transport Layer Protocol, or NTLP) provides a generic delivery service for different signalling applications, based on the General Internet Signalling Transport (GIST) (Schulzrinne and Hancock, 2006) protocol. Signalling applications reside in the upper layer (the NSIS Signalling Layer Protocols, or NSLPs). Current NSLPs include signalling applications for QoS resource reservation (QoS-NSLP) (Manner et al., 2006) and NAT/Firewall traversal (NAT/FW-NSLP) (Stiemerling et al., 2006). An introduction of the NSIS protocol suite is given in Fu et al. (2005). In addition to a separation between signalling applications and generic delivery services, NSIS differs from RSVP in several other key design choices:

- reuse of existing transport and security protocols (instead of designing a new transport protocol and later adding protocol properties such as reliability and security)
- (whenever possible) decoupling discovery of the next signalling node from delivery of signalling messages between neighbouring signalling nodes
- introduction of a session identifier (carried in signalling messages to uniquely identify installed state, instead of reusing the flow identifier as done in RSVP).

Through GIST, NTLP functions as a universal ‘messenger’ in routing and delivering signalling messages for all signalling applications. First, GIST installs and maintains a ‘routing’ state used to direct signalling messages forwards or backwards along the flow path. In addition, a ‘messaging’ state is maintained at the NTLP level for multiple signalling sessions to reuse existing transport associations between neighbouring NSIS peers. These two types of state are different from signalling application (NSLP)-specific states, such as QoS reservation states or NAT bindings.
Securing the Next Steps In Signalling (NSIS) protocol suite

GIST defines two modes of delivering signalling messages. The first mode, known as the Datagram mode (D-mode), follows an RSVP signalling style by using end-to-end addressed messages. The end-to-end addressed message contains the source and the destination IP addresses of the data flow. The messages are intercepted along the path by NSIS nodes interested in these messages (by using the Router Alert Option (Katz, 1997)). The second mode, so-called Connection mode (C-mode), is used when NSIS nodes are directly addressed. This mode assumes that a discovery procedure should have been performed (or the address of the receiving node is known via other means, e.g., by manual configuration or a new discovery mechanism). The default discovery mechanism is based on a Query-Response message exchange using D-mode encapsulation, which afterwards allows establishing a C-mode messaging association (upon which NSIS signalling messages can be delivered). Any NSIS node that implements the desired NSLP functionality, upon receipt of a Query D-mode message, will respond with a Response message to the D-mode message sender. Therefore, GIST in D-mode for NSIS message delivery essentially does not differ much from RSVP; both provide unreliable transport for delivering signalling messages, maintaining soft state for routing of the signalling messages. Therefore, the focus here is on securing GIST C-mode message delivery. However, unlike in RSVP, GIST does not employ two-way signalling message exchange, nor does it install a signalling application specific state (QoS reservation in particular), since they are related to specific signalling applications, such as QoS reservation, firewall pinhole or NAT binding configuration. These functions are now part of the NSLP layer functionality, e.g., Resv, possibly followed by a Response. Figure 1 compares these two signalling approaches (RSVP and NSIS) according to the necessary function components. More discussions about the difference are given in Fu et al. (2005).

![Figure 1](image)

We also discuss to some extent how to secure D-mode discovery process as it forms a basis of initialising the C-mode operations.

2.2 NSIS signalling scenarios

An example NSIS signalling scenario is shown in Figure 2. Each NSIS node may store messaging association state information about its peers. A node, the NSIS Initiator (NI), initiates the signalling exchange, while some nodes along the signalling path, called NSIS Forwarders (NFs), intercept and then forward signalling messages, and the NSIS Responder (NR) terminates the signalling.

![Figure 2](image)

Figure 2 also shows that not all routers along the data path need to be NSIS aware, nor do all NSIS nodes necessarily support all signalling applications. For a particular NSIS session, nodes not supporting the desired signalling application are bypassed. In this example, messages of signalling application type A will be delivered between Router 2 and the edge node, without being processed in Router 3.

The NSIS protocol suite is envisioned to support various signalling applications that need to install and manipulate these application-specific states as well as message routing and transport states at signalling nodes along the data flow path through the network. Unlike many other protocols, which work in an end-to-end fashion and involve no intermediate nodes, NSIS is a protocol suite for distributed state establishment with the ability to interact with
intermediaries. The set of nodes participating in the chain in the signalling communication can change over time, e.g., due to re-routing events and possibly, node mobility. As such, the communication between an NI and an NR (i.e., the two signalling ends) is more complex. Furthermore, for deployment reasons it is likely that the proxy signalling functionality (as developed with RSVP proxy (Gai et al., 2002)) will be desirable, whereby the data receiver and or the data sender are not NSIS aware.

As illustrated in Figure 3, the entire NSIS end-to-end communication path can be split into different parts, namely first-peer, last-peer, intra-domain, or inter-domain.

**Figure 3** An end-to-end signalling environment

For example, in some cases it may be necessary to allow an NSIS signalling node to explicitly authorise a non-adjacent NSIS node. Each of these models can represent some different impacts on security requirements.

- **First-peer and last-peer communication.** First-peer (and last-peer as a variant) communication can vary from one access scenario to another. For example, in an enterprise network scenario, there can be a pre-established security association between the NSIS entities for a first- or last-peer communication. In a roaming scenario, it is difficult to assume a pre-established key between a mobile node and the attached network (referred to as visited network) since the mobile node will most likely be unknown to the visited network. Generally, in a hostile environment, it would be desirable to perform a proper authentication and key exchange to establish the necessary pre-condition for providing channel security.

- **Intra-domain communication.** When an NSIS signalling message travels between NSIS nodes belonging to the same administrative domain, authorisation and key management can be simpler. However, security protection is still desired to prevent non-NSIS nodes from interfering with the NSIS-aware signalling nodes.

- **Inter-domain communication.** Inter-domain communication will be necessary, e.g., when a QoS signalling request needs to traverse multiple domains, or when one network wants to signal a QoS reservation towards a neighbouring domain. Note that the difference between these two cases is mainly related to the granularity of the QoS reservation request. In the former case a per-flow reservation needs to be made, whereas in the latter case most likely a reservation for an aggregated flow is desired. Neighbouring domains can establish the necessary infrastructure, such as key distribution and contractual aspects, to subsequently secure signalling messages.

In short, the NSIS protocol suite needs to be able to protect signalling messages and payloads between adjacent NSIS nodes, between non-adjacent NSIS nodes over multiple NSIS hops (such as required by middle-to-middle protection of payloads), end-to-middle or middle-to-end protection and end-to-end protection. End-to-middle (or middle-to-end) protection is necessary when intermediate NSIS nodes are not allowed to eavesdrop or to modify certain objects, or are required to cryptographically verify a certain payload. As an example, the protected exchange of signalling messages (or selected payloads) from ingress towards egress node of an administrative domain might be desired as described in RMD (Bader et al., 2006). End-to-end protection of the entire signalling message might not be possible (or useful) since intermediate NSIS nodes need to add, inspect, modify or delete objects in this message. However, some objects may need to be protected between the signalling initiator and the signalling responder (i.e., end-to-end) or between non-neighbouring NSIS nodes.

### 3 Security threats

Based on the discussions in Section 2, we analyse the various threats to the NSIS protocol suite in more detail. For the convenience of discussion we follow the definition in Wu et al. (1999) and differentiate adversaries into three types: **InsiderNSIS**, **OutsiderOnPath**, and **OutsiderOffPath**. These terms refer to participating NSIS nodes, traversed non-NSIS nodes along the signalling path, and nodes not along the signalling path, respectively. Note that this section does not assume any particular security solutions for NSIS.

#### 3.1 Eavesdropping and traffic analysis

An **OutsiderOnPath** and an **InsiderNSIS** adversary can eavesdrop NSIS signalling messages, and use the collected signalling packets to perform traffic analysis; for example, learning QoS parameters, communication patterns, policy
rules for firewall traversal, policy information, application identifiers, user identities, NAT bindings, authorisation objects, network configuration and performance information, etc. Based on collected information, an eavesdropper can also perform replay attacks (see Section 3.2).

An adversary’s capability to eavesdrop on signalling messages may violate a user’s preference for privacy, particularly if unprotected authentication or authorisation information (including policies and profile information) is exchanged. Note, for certain objects or messages it is highly desirable to permit the active participation of intermediate NSIS nodes to inspect either all or most of the signalling message payloads, in order to allow these nodes to perform the desired protocol processing. Hence, InsiderNSIS nodes are also considered as a potential threat source for the NSIS protocol suite.

3.2 Replay attacks

An Outsider OnPath and an InsiderNSIS adversary can replay (or reorder) eavesdropped signalling messages. Making the assumption of InsiderNSIS adversaries makes the protocol interaction very difficult and is likely not to be useful for many signalling applications due to the nature of the signalling interaction. With a QoS signalling, for example, a mobile host might request a certain QoS treatment from QoS routers in the access network. The access network might need to forward the QoS signalling message towards the destination address and will therefore be charged by its neighbouring network accordingly, too. The existence of such an authorisation relationship between neighbouring entities and the utilised charging model does not make it beneficial for the access network node to replay any QoS reservation requests for end hosts.

Without data origin authentication, integrity, replay and optional confidentiality protection of signalling messages, an adversary can impact the functionality of a network considerably. For example, a DoS attack can be launched by replaying past QoS signalling requests to install faked QoS reservations. Replaying refresh messages can also be exploited to prevent soft state timeouts. This threat may be particularly severe in mobile environments.

3.3 Man-In-The-Middle (MITM) attacks

The discovery phase and the establishment of a messaging association are particularly vulnerable to MITM attacks. In the discovery phase, an OutsiderOnPath adversary can inject a bogus response message, forcing the querying node to start a messaging association establishment procedure with either an adversary or with another NSIS node (which is not on the signalling path). Note that an adversary located in a broadcast medium, such as Ethernet or Wireless LAN, where ARP or Wireless LAN, where ARP spoofing is possible is referred as an Outsider OnPath entity.

Clearly, for end-to-end addressed messages such attacks are possible, particularly if the adversary is located along the path and able to intercept the discovery message, which traverses the adversary. The MITM adversary can redirect signalling messages to another legitimate NSIS node that might not even be located at the signalling path.

Besides, without a proper authorisation procedure, an NSIS node implementing the functionality of NSLP X can pretend to implement the functionality of NSLP Y. Therefore, it is necessary to have a trusted third party to warrant for the capabilities of individual routers. Furthermore, a malicious non-NSIS node can be detected with the corresponding security mechanisms. A legitimate NSIS node, which is not the next NSIS node along the path, cannot be detected without using topology knowledge as part of the authorisation decision. A more detailed discussions of the problem associated with this type of authorisation can be found in Aoun et al. (2005).

In the example shown in Figure 4 we investigate the NSLP peer discovery exchange used in the context of C-Mode messaging. The subsequent paragraph provides a high-level and abstract description of the exchange. The details of the discussed proposal can be found in Schulzrinne and Hancock (2006). Hence, this message which is addressed towards the data receiver’s IP address will be intercepted by an intermediate NSIS aware node.

Figure 4  MITM attack during the discovery exchange

An OutsiderOnPath, OutsiderOffPath or a malicious InsiderNSIS may respond to the discovery with a Query-Response with its own IP address as the address of the next NSIS aware node along the path. When the adversary is located in a same physical network as the legitimate NSIS node, the attacker may succeed if its ‘response’ reaches the querying node faster than the legitimate response. Without any additional information the querying node has to rely on the data returned by the Query-Response message. Then a messaging association is established with an entity at a given IP address (IPx) in step (Kempf and Austein, 2004). The adversary may then again establish a messaging association with the next NSIS node to forward the signalling message. Note that the adversary can just modify the Query-Response message, forcing the querying node to establish a messaging association with any other NSIS node that is not even along the data path.

As a variant of this attack, an OutsiderOffPath adversary can also flood a node with bogus discovery reply messages, even if it is not able to eavesdrop on transmitted discovery requests. If the discovery message sender accidentally
accepts one of those bogus messages then a MITM-attack as described in Figure 4 is possible. It should be noted that the process is self-healing since the discovery process is periodically performed. If an adversary is unable to mount this attack with every discovery message, the correct next NSIS node along the path will be discovered again. However, this could result in a ping-pong effect on the underlying NSIS state (and also the message routing and delivery behaviours).

During the process of establishing a security association based on the previous discovery, an adversary can fool the signalling message sender with respect to the entity with whom it has to establish the security association. Without proper authorisation the sender may successfully finalise the security protocol run with an adversary. If successful it is able to modify signalling messages to mount DoS attacks or steal services (depending on the used signalling application). In addition, an adversary can terminate the sender's messages and inject messages to another NSIS peer.

### 3.4 Denial of Service (DoS) attacks

An **Outsider OnPath**, **Outsider Off Path** or an **InsiderNSIS** adversary can flood an NSIS node with bogus messages to cause a DoS attack. These bogus messages can be, for example, discovery query requests, idempotent refresh messages, or NSLP-level trigger messages requiring the involvement of a third-party node (such as policy or AAA servers), faked GIST or NSLP error or response messages, or vulnerabilities of the used transport layer protocols (see for example Gont, 2006) or even causing a node to perform heavy cryptographic operations. Sometimes even the usage of a Router Alert Option is seen as an opportunity for a DoS attack. If the adversary transmits a large number of such messages, the responding node(s) may be overloaded with high amount of computation and transmission requirements, and thus be unable to process other messages originating from legitimate entities/users. This can be even more challenging if an NSIS node needs to issue signalling messages on behalf of someone else (by acting as a proxy).

As an example of DoS attacks, an adversary can flood a QoS aware network with receiver-initiated reservation requests, such as RSVP PATH-like messages. If these messages are unauthenticated and contain no authorisation information, then state at QoS aware routers along the path will be created until the data responder authorises the reservation in the reverse direction. In such a scenario an adversary can force the QoS aware routers to store state information. Furthermore, since the state expiration might take some time, the routers might reject legitimate QoS reservation requests due to lack of resources.

Among all participating NSIS nodes, NSIS nodes located at the edges of administrative domains are the most critical entities of NSIS signalling, and they may also act as a security gateway or firewall for incoming and outgoing signalling messages (as well as data traffic). For outgoing traffic these devices have to implement the security policy of the local domain and apply the appropriate security protection.

### 3.5 Identity spoofing

Identity spoofing in NSIS operations can occur in three forms: first, spoofing the identifier that is used for computing the authorisation decision; for example, caused by the usage of weak authentication mechanisms; second, an adversary can modify the flow identifier carried within a signalling message; third, data traffic can be spoofed. The latter two issues refer to aspect that the flow identifier can be seen as an identifier of identifying the QoS reservation on the data plane.

The first type of attack can be mounted by an **Outsider OnPath**, **InsiderNSIS** and an **Outsider Off Path** adversary. The latter two attacks require the capability of the **Outsider OnPath** and the **InsiderNSIS** adversary.

In the first case, Eve, acting as an adversary, may claim to be the registered user Alice by spoofing Alice's identity. Eve thereby causes the network to charge Alice for the network resources, which are consumed by Eve. This attack could also be classified as theft of service. The choice of selecting an identifier that can be associated to a chargeable entity depends on the deployment environment. Making inappropriate assumptions can easily lead to vulnerabilities, such as using a plaintext identifier in a QoS signalling protocol as the basis for authorisation as, for example, provided by the RSVP Policy Object (Yadav et al., 2001) allows a malicious end host or a MITM to inject a modified identifier.

In the second case, an adversary may be able to exploit the established flow identifiers (required for QoS and NAT/FW NSLP). These identifiers are, among others, IP addresses, transport protocol type (e.g., UDP, TCP), port numbers, IPv6 flow labels, etc. Modification of these flow identifiers allows adversaries to make a QoS reservation ineffective for the QoS resource requesting entity, to reuse it for its own traffic, or to install arbitrary policy rules at middleboxes for further attacks against end hosts (or the network infrastructure).

In the third case, an adversary may spoof data traffic, i.e., inject data traffic with a flow identifier matching the installed identifier. As described in the previous section, NSIS signalling messages carry a flow identifier, which is associated with a specified behaviour (e.g., allow a particular flow to receive certain QoS treatment or for packets to traverse a firewall) based on the type of signalling application. Thus, an adversary may use IP spoofing in order to inject data packets to benefit from previously installed flow identifiers (which is also a theft of service). For example, an adversary might observe a NAT/Firewall NSLP message towards a corporate network firewall. After the signalling message exchange was successful, user Alice is allowed to traverse the company firewall based on the established packet filter to contact her internal mail server. Now, adversary Eve, who monitored the signalling exchange, is able to craft data packets by spoofing some header fields that will allow the packets to
pass the corporate firewall. Depending on the exact location of the adversary and the degree of routing asymmetry the adversary might even see the response messages. Here, to actualise this attack, Alice does not need to participate in the exchange of signalling messages. To address this attack one can rely on data origin authentication (e.g., by using IPsec (Kent and Atkinson, 1998)).

### 3.6 Stealing authorisation information

Different information is used for computing an authorisation decision, such as the authenticated identity, availability of sufficient funds, roles and traits, number of concurrent sessions, amount of requested bandwidth or other QoS parameters, etc. The specific authorisation policies might heavily depend on the usage of the specific signalling application and the specific request, such as a request for making a QoS reservation, the request to change a NAT binding or to allocate a firewall pinhole.

One approach to verify an entity’s rights to access these resources is using the authenticated identity. Another approach is to use an authorisation token, e.g., as described in RFC 3182 (Yadav et al., 2001). The functionality and the structure of such an authorisation token for RSVP are described in Hamer et al. (2003a, 2003b). By using such an authorisation token, it is possible to bind the authorisation decision provided by one protocol (e.g., SIP) to the QoS signalling protocol.

However, if an authorisation token is returned to the end host without confidentiality protection, then it might allow an eavesdropper to reuse it (depending on the constraints carried inside the token) to gain the same authorisation rights as the legitimate owner of the token. An adversary might, for example, use such a token to make an expensive QoS reservation. Without appropriate protection of the token an end host may want to modify the token content and thereby grant itself more rights. A more detailed discussion of authorisation tokens used in the context of RSVP and SIP can be found in Hamer et al. (2005a, 2005b). A more recent example of authorisation tokens is available with the Security Assertion Markup Language (SAML) (Maler et al., 2003a, 2003b).

### 3.7 Fraud

Signalling applications (such as QoS NSLP or NAT/FW NSLP) often involve three parties: the user, a network that offers NSLP services such as QoS service support and a third party which authenticates and authorises the user (AAA server). For a QoS reservation the trusted third party, in most cases, needs to ensure that the network providing the QoS service actually receives a financial compensation.

The QoS service-providing network might intentionally deliver incorrect resource reports about the resource consumption to the user’s home network. In a typical AAA setting it would be quite difficult to detect this fact since the user is not involved in the exchange between the AAA client and the AAA server. Furthermore, the price of the QoS reservation might not even be known to the user, or might change due to mobility events or the end host.

Another problem may appear when the service provider (or the user) later denies the existence, or some parameters (e.g., volume or price), of a QoS reservation (or other NSLP services) towards the third party. This is often regarded as a problem if non-repudiation is not provided, which may appear in two forms:

- **Service provider’s point-of-view.** A user may deny having issued a reservation request, which was actually charged for. The service provider may then want to be able to show that a particular user issued the reservation request in question.
- **User’s point-of-view.** A service provider may claim to have received a number of reservation requests from a particular user. The user in question may want to show that such reservation requests have never been issued and may want to see correct service usage records for a given set of QoS parameters.

The ability to provide non-repudiation places additional requirements on security mechanisms and is often associated with a more complex protocol interaction and additional security mechanisms. Therefore, non-repudiation is not provided with network access authentication in today’s networks; the user has to trust the participating network operators to correctly meter the traffic, to collect accounting data, and to ensure that no unforeseen problems occur.

Please note that this issue is also relevant for accessing services in general. The large deployment of wireless LAN hotspots, as one example of a service, also faces the same problems. It is anticipated that a solution applicable for such an environment may be reused for providing non-repudiation protection for NSLPs.

### 3.8 Disclosing information about a network

Some organisations or enterprises may desire not to reveal their internal network structure (or other related information) outside of a closed community. However, an adversary may be able to use NSIS messages for disclosing network topology (e.g., discovering which nodes exist, which use NSIS, what version, what resources are allocated, what capabilities nodes along a path have, etc.). Discovery messages, traceroute, diagnostic messages (see Terzis et al. (2000) for a description of diagnostic message functionality for RSVP and a Ping tool for NSIS (Dickmann et al., 2005)), and query messages, in addition to record route and route objects and provide potential assistance to an adversary. Hence, the requirement of not disclosing a network topology may conflict with other requirements (e.g., to provide diagnostic facilities for network monitoring and administration).
3.9 Session manipulation

A Session Identifier is included in NSIS signalling messages as a reference to the established state. This helps to simplify mobility and general state manipulation. However, this has a few security implications. The session manipulation issue is a key issue among them (see Tschofenig et al., 2003a for a detailed discussion).

Figure 5 shows an NI that has established state information at NSIS nodes along a path as part of the signalling procedure. As a result, AccessRouter1, Router3, and Router4 (and other nodes) have stored session state information including the Session Identifier SIDx.

If an adversary were able to obtain the Session Identifier (e.g., by eavesdropping on signalling messages), it would be able to add the same Session Identifier SIDx to a new signalling message. When the new signalling message hits Router3, existing state information can be modified. The adversary can then modify or delete the established reservation and cause unexpected behaviour for the legitimate user. The problem appears at Router3 (i.e., the cross-over router) that is unable to decide whether the newly received signalling message was initiated from the owner of the session or not.

In addition, nodes other than the initial signalling message originator may be allowed to signal state information during the lifetime of an established session. The flexible design of NSIS with regard to the supported deployment scenarios allows any NSIS-aware node along the path to trigger or terminate a signalling message exchange (e.g., a local repair procedure, or an asynchronous notification). If only the initial signalling message originator were allowed to trigger signalling message exchanges, some enhanced protocol functionality might not be possible. As an example, idempotent refresh messages need to be sent in an end-to-end fashion rather than allowing intermediate nodes to inject them. This tradeoff between signalling protocol flexibility and security become more outstanding for NSIS mobility support (Lee et al., 2006). Furthermore, the number of NSIS nodes in the flow path may change over the lifetime of a signalling session, e.g., due to node mobility or route changes. Without properly addressing the session ownership problem, an adversary can launch DoS, theft of service and other attacks.

This particular threat reflects an important issue that needs to be considered in many aspects of the protocol design; namely, who is authorised to modify established state at various nodes in the network. Even though authorisation for the establishment of resources, such as a QoS resource, might need to be provided only between adjacent nodes the aspect of session ownership appears in various places of the NSIS protocol suite. It is also one of the critical issues that need to be addressed in mobile environments (Lee et al., 2006), in the context of QoS signalling and reservation collisions (McDonald et al., 2003) and a concern needs to be resolved for the data sender behind the NAT signalling scenario in the NAT/Firewall NSLP (Stiemerling et al., 2006).

A tradeoff analysis is required for each NSLP, to evaluate whether the complexity of a solution for the session ownership problem is justified compared to the mitigated threats.

4 Towards a secure NSIS protocol suite

The design of the NSIS protocol suite is not yet complete and a number of open issues, especially with security, aspects still exist. In this section we present some considerations towards a secure NSIS protocol suite.

Since the NSIS protocol suite is split into two layers, the proposed security solution needs to offer security protection for both NTLP and existing NSLPs. With extensibility in mind, security building blocks within NTLP and some general aspects of NSLPs should be designed to serve as a useful facility for designing future NSLPs. Based on the analysis presented in previous sections, we believe that the following issues are vital for securing the NSIS protocol suite:

- ability to run an authentication and key exchange protocol between neighbouring NSIS peers (supporting either symmetric or asymmetric cryptography or even a hybrid version)
- security association establishment to provide integrity, confidentiality and replay protection for signalling messages exchanged between neighbouring peers
- Denial of Service protection
- lightweight protection for the discovery mechanism
- authorisation of the NTLP signalling peers
- flexible authorisation at the NSLP layer including the ability to interwork with the existing AAA infrastructure.

Notably, designing new, custom security protocols is a highly complex and time-consuming task. Luckily, many of today’s security protocols offer several desired features already: flexible authentication methods, formally verified correctness, offering of DoS protection, ciphersuite negotiation, low number of roundtrips, etc.. Based
on this fact, the use of standard security protocols is highly encouraged in securing the NSIS protocol suite. The following subsections discuss some usage scenarios of existing security protocols and some extensions in order to address the above seven issues.

4.1 Effective authentication and key exchange at NTLP layer

A considerable amount of eavesdropping, replay and MITM attacks can be resolved by the same mechanism; namely, an effective authentication and key exchange protocol. Usually this is a high cost operation. To avoid expensive cryptographic computations (such as those required by a digital signature) for each individual object, message or signalling session, the reuse of an established security association between two NTLP peers will be desirable. As a consequence, the high computational effort can be amortised with the usage of faster symmetric cryptographic protection for multiple signalling sessions. Fortunately, this can be supported at the NTLP layer, and we discuss several approaches to address it.

One approach is based on the Transport Layer Security (TLS) (Dierks and Allen, 1999; Dierks and Rescorla, 2006), which provides both a flexible authentication and key exchange protocol framework (i.e., TLS Handshake Layer) and provides protection of the subsequently exchanged application data via the TLS Record Layer. TLS provides session key establishment based on unilateral or mutual authentication. The mutual authentication function provided by TLS can be used for the establishment of a bidirectional NTLP secure messaging association between any two neighbouring NSIS peers of the same NSLP type. Additionally, TLS supports additional ciphersuites and enhancements to the protocols, such as Kerberos-based authentication (Medvinsky and Hur, 1999), strong password-based authentication (Taylor et al., 2005), pre-shared secret authentication, and a mixture of shared secret and public key based authentication (Eronen and Tschofenig, 2005) and Extensible Authentication Protocol (EAP) (Aboba et al., 2004) support within the TLS Handshake Protocol (Funk et al., 2006). Note that some enhancements are work in progress and are therefore subject to change.

Another approach is to apply the Internet Key Exchange protocol (IKE) (Harkins and Carrel, 1998), IKEv2 (Kauffman, 2005) or KINK (Thomas and Vilhuber, 2006) to support establishment of IPsec security associations, when IPsec is available and chosen for securing signalling messages between two neighbouring peers. In addition, the usage of the Host Identity Protocol (HIP) (Moskowitz and Nikander, 2006; Moskowitz et al., 2006; Jokela et al., 2006) can be used to secure NSIS signalling messages. The rich availability of the mechanisms for IPsec SA establishment allows flexibility in the use of SAs in different deployment scenarios.

Both approaches require minimal changes to existing security protocols. They can be integrated into an NTLP engine without the need to change the protocol itself, and do not necessarily require changes to the key exchange protocol itself.

For example, after a QoS NSLP node discovers the next QoS NSLP peer along the path, GIST establishes a messaging association with the discovered QoS NSLP node when the C-Mode should be used. As a result the QoS NSLP engine receives the payload provided by the NTLP and additionally, security related information via the GIST-NSLP API (such as the authenticated identity and authorisation information). This information can then be used by the NSLP as input to the policy engine for the authorisation decision (which is detailed in Section 4.5).

4.2 Lightweight security in discovery mechanism

The discovery message exchange is a security sensitive process and additionally, very difficult to secure. To prevent adversaries from redirecting messages, a cookie-based mechanism can be used in the discovery procedure. This countermeasure refers to the MITM attacks described in Section 3.3 and in Figure 4.

The cookie-based mechanism (see Figure 6) can be illustrated as follows.

![Figure 6 Protection of the discovery procedure in GIST](Image 339x336 to 529x476)

A Cookie(I) is included in the Discovery-Response message to prevent OutsiderOffPath adversaries from flooding the querying node with bogus responses since the initiator can use Cookie(I) to match the response with the request. Since Cookie(I) is a randomly chosen value, an Out-siderOffPath adversary cannot compute a valid response.

The cookie provided by the responding node (Cookie(R)) is used to prevent DoS attacks in the classical sense as used by other protocols, such as SCTP (Stewart et al., 2000) or IKEv2 (Kauffman, 2005). Note that the responder must not create per-session state with responding to the Discovery-Query, otherwise DoS vulnerabilities will be introduced. The Responder returns the received Cookie(I) and its own Cookie(R).

Finally, when the initiator receives the Discovery-Response it compares the Cookie(I) value and runs an authentication and key exchange protocol (such as TLS) with the discovered node before establishing a messaging association. To prevent an OutsiderOnPath adversary from modifying the Discovery-Response message and adding wrong information about the next NSIS node along the path,
Cookie(R) is repeated once channel security is in place. This allows the responder to verify that it has actually participated in the discovery exchange. Thereby the discovery procedure is bound to the subsequent exchange.

A more detailed treatment of the path-coupled discovery procedure security aspects is provided in Schulzrinne and Hancock (2006).

### 4.3 Basic authorisation at NTLP layer

The goal of computing an authorisation decision at the NTLP layer is to ensure that only legitimate NSIS nodes initiate a signalling communication. In a generic signalling environment it is quite difficult for the GIST engine to make a meaningful decision without consulting the NSLP with respect to signalling application specific functionality. In some deployment environments it is, however, possible for the NTLP layer to perform basic access control operations and to allow only certain nodes from a particular domain to establish a messaging association. The authenticated identity might be used for computing this authorisation decision but it is feasible to utilise authorisation certificates (if available). Such authorisation decisions at the NTLP layer are particularly useful in intra-domain scenarios as well as in environments, such as enterprise networks, where the communicating peers are known in advance based on pre-configuration.

In order to address one of the MITM attack variants, which is described in Section 3.3, where an NSIS node claims to support a certain signalling application, it is necessary to provide information along with the credentials, such as authorisation certificates. Alternatively, it may be possible to tie the protocol operation within a previous protocol exchange, e.g., the network access authentication procedure. A more detailed treatment of these issues is provided in Aoun et al. (2005).

### 4.4 Flexible authorisation at NSLP layer

As described in Section 3.6, authorisation aspects deserve special attention. The work on the NAT/Firewall and the QoS NSLP showed the difficulty in properly tackling authorisation issues in a generic way for all NSLPs (see Stiemerling et al. (2006) for a discussions about NAT/Firewall specific authorisation issues and Alfano et al. (2006) and Tschofenig et al. (2003b, 2003c) and for a discussion of the QoS specific authorisation aspects. It is quite likely that individual NSLPs are able to reuse some building blocks but the authorisation handling is very likely different for each NSLP.

The interworking of NSIS protocols with today’s AAA protocols such as Diameter (Calhoun et al., 2003) and RADIUS (Rigney et al., 2000) is being regarded as an important building block of the network infrastructure.

Although the NTLP layer needs to provide certain authorisation functionality, most authorisation decisions will be made at the NSLP layer. The decision to successfully authorise QoS reservation requests might be related to the ability of the user (or other another entity requesting a resource reservation) to pay for the preferential treatment. Making an authorisation decision to create a NAT binding might likely depend on the traffic direction (network internal traffic towards the internet vs. traffic from the internet towards a private network). To create packet filters at a firewall, the security policy of the administrative domain will certainly play an important role. This policy will be different in a corporate network, home network and in a 3G network. Modifying the established state should only be possible for the entity that created the state. This authorisation decision is also known as sender invariance.

An individual NSIS router will, in many cases, be unable to make an authorisation decision by itself without consulting third parties. This is particularly the case for an environment where hosts roam from one network to another. Hence, a QoS aware router that receives a QoS reservation request might want to contact the AAA infrastructure to off-load the authorisation decision.

In a recent work (Tsenov et al., 2005) we describe the integration of the EAP (Aboba et al., 2004) into NSIS, which attempts to provide a flexible advanced authorisation mechanism.

### 4.5 An example usage of securing NSIS

Based on the above considerations with some individual examples, Figure 7 shows an example of possible overall protocol exchange for the interactions between NSIS, SIP (Rosenberg et al., 2002) and Diameter. Note that a user will be authorised by its neighbouring NSIS node rather than some arbitrary networks along the path towards the end host.

![Figure 7: QoS authorisation with SIP and Diameter interaction](image-url)
First, a service request is sent from the end host Alice to the SIP proxy. In this example, the SIP proxy creates and returns an authorisation token to bind the application layer signalling exchange to the subsequent NSIS signalling session (e.g., a sender-initiated reservation). The authorisation token is attached to the NSIS signalling message and the message itself is intercepted by the first NSIS NSLP node, ‘Bob’. Upon the receipt of this token, Bob needs to authorise the QoS request and delegates this responsibility to the Diameter QoS application (Alfano et al., 2006). A Diameter QoS authorisation request, which includes authorisation information and QoS information, is forwarded to the SIP proxy that created the authorisation token for verification. As a response, the authorisation decision is returned with a corresponding Diameter message. If the Diameter response is positive and the admission control process allows, then the Traffic Control engine is contacted to install the QoS reservation and Bob forwards the NSIS QoS message further along the path. Finally, accounting, and possibly credit control, procedures are initiated to keep track of the consumed resources, to control the re-authorisation policy and to ensure that the user’s credit limit is not exhausted.

5 Summary and outlook

This paper describes security threats and lists the most important aspects of security protection for the NSIS protocol suite. Through identifying open issues in previous QoS signalling protocols, some critical design decisions have been made for NSIS, especially with respect to security.

We believe that the integration of security in the design of NSIS has helped to develop a sound and reasonable framework. Many protocols developed so far have experienced critical shortcomings due to a lack of security considerations (see an analysis of QoS signalling protocols in Manner and Fu (2005).

The main specifications of the NSIS protocol suite are getting closer to their final completion and implementation work is ongoing (Fu et al., 2005). We expect that these activities will reveal further details about performance and the practical usage of security mechanisms. Security will therefore continue to play an important role in further NSIS protocol development since a large fraction of the total performance will go into security processing. Particularly the usage of NSIS in mobile environments demands further investigations and will remain a hot research topic due to performance constraints and the need for optimisations. Lee et al. (2006) provides an overview of ongoing work in this field.

We also plan to assess the deployment aspects of the NSIS protocol suite and their impacts on existing and future network architectures (e.g., 3GPP, 3GPP2, WiMax, ETSI-TISPAN, ITU-T NGN).

Acknowledgements

We would like to thank members of the IETF NSIS working group for their fruitful discussions. In particular, we would like to thank Dirk Kroeselberg, Tseno Tsenov, Henning Schulzrinne, Robert Hancock, Richard Graveman, John Loughney and Allison Mankin.

References


